Understanding Complex Systems

Springer : COMPLEXITY

Bruce Edmonds
Ruth Meyer *Editors*

# Simulating Social Complexity

## A Handbook

Springer

# Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems — cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The three major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "SpringerBriefs in Complexity" which are concise and topical working reports, case-studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

# Understanding Complex Systems

**Founding Editor: S. Kelso**

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition - typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels – and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals:

First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuro- and cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

For further volumes:
http://www.springer.com/series/5394

Bruce Edmonds • Ruth Meyer
Editors

# Simulating Social Complexity

## A Handbook

Springer

*Editors*
Bruce Edmonds
Ruth Meyer
CPM - Centre for Policy Modelling
Manchester Metropolitan University
Manchester
United Kingdom

ISSN 1860-0832 ISSN 1860-0840 (electronic)
ISBN 978-3-540-93812-5 ISBN 978-3-540-93813-2 (eBook)
DOI 10.1007/978-3-540-93813-2
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013934710



Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Contents

# Part I
# Introductory Material

# Chapter 1
# Introduction to the Handbook

**Bruce Edmonds and Ruth Meyer**

**Why Read This Chapter?** To understand some of the background and motivation for the handbook and how it is structured.

## 1.1 Simulating Social Complexity

As the title indicates, this book is about *Simulating Social Complexity*. Each of these three words is important.

*Simulating* – the focus here is on individual- or agent-based computational simulation rather than analytic or natural language approaches (although these can be involved). In other words, this book deals with computer simulations where the individual elements of the social system are represented as separate elements of the simulation model. It does not cover models where the whole population of interacting individuals is collapsed into a single set of variables. Also, it does not deal with purely qualitative approaches of discussing and understanding social phenomena, but just those that try to increase their understanding via the construction and testing of simulation models.

*Social* – the elements under study have to be usefully interpretable as interacting elements of a society. The focus will be on human society, but can be extended to include social animals or artificial agents where such work enhances our understanding of human society. Thus this book does not deal with models of single individuals or where the target system is dealt with *as if* it were a single entity. Rather it is the differing states of the individuals and their interactions that are the focus here.

B. Edmonds (✉) • R. Meyer
Manchester Metropolitan University, Centre for Policy Modelling, Manchester Metropolitan University Business School, All Saints Campus, Oxford Road, M15 6BH Manchester, UK
e-mail: bruce@edmonds.name; meyer.ruth@gmail.com

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_1,

*Complexity* – the phenomena of interest result from the interaction of social actors in an essential way and are not reducible to considering single actors or a representative actor and a representative environment. It is this complexity that (typically) makes analytic approaches infeasible and natural language approaches inadequate for relating intricate cause and effect. This complexity is expressed in many different ways, for example: as a macro/micro link; as the social embedding of actors within their society; as emergence. It is with these kinds of complexity that a simulation model (of the kind we are focussing on) helps, since the web of interactions is too intricate and tedious to be reliably followed by the human mind. The simulation allows *emergence* to be captured in a formal model and experimented upon.

Since this area is relatively new it involves researchers from a wide variety of backgrounds, including: computer scientists, sociologists, anthropologists, geographers, engineers, physicists, philosophers, biologists and even economists. It is only in the past few years that the elements and shape of the field are starting to be discernable. It is for this reason that a handbook is particularly timely. We hope that it will help to introduce and guide newcomers into the field so as to involve more minds and effort in this endeavour, as well as inform those who enter it from one perspective to learn about other sides and techniques.

## 1.2 The Context, Going Back to Herbert Simon

This handbook is in memory of Herbert Simon, since he initiated several key strands that can be found in the work described here.

He observed how people behave in a social system instead of following some existing framework of assumptions as to how they behave (Simon 1947). That is, he tried to change the emphasis of study from a normative to a descriptive approach – from how academics think people should be behaving to how people are observed to behave. Famously he criticised "arm chair" theorising, the attempt to make theories about social phenomena without confronting the theory with observation. There is still a lot of "arm chair" theorising in the field of simulating social complexity, with a "Cambrian Explosion" of simulation models, which are relatively unconstrained by evidence from social systems. If the development of this work is seen as a sort of evolutionary process then the forces of variation are there in abundance but the forces of selection are weak or non-existent (Edmonds 2010).

Importantly for the simulation of complex social systems, Simon observed that people act with a procedural rather than substantive rationality – they have a procedure in the form of a sequence of actions that they tend to use to deal with tasks and choices rather than try to find the best or ideal sequence of actions (Simon 1947, 1976). With the advent of computational simulation it is now fairly common to represent the cognition of agents in a model with a series of rules or procedures. This is partly because implementing substantive rationality is often infeasible due to the computational expense of doing do, but more importantly it seems to produce results with a greater "surface validity" (i.e. it *looks* right). It turns out that adding some adaptive or learning ability to individuals and allowing the individuals to interact can often lead to effective "solutions" for collective problems (e.g. the entities in Chap. 21). It is not necessary to postulate complex problem solving and planning by individuals for this to occur.

Herbert Simon observed further that people tend to change their procedure only if it becomes unsatisfactory; they have some criteria of sufficient satisfaction for judging a procedure and if the results meet this they do not usually change what they do. Later Simon (1956) and others (e.g. Sargent 1993) focused on the contrast between optimisers and satisficers, since the prevailing idea of decision making was that many possible actions are considered and compared (using the expected utility of the respective outcomes) and the optimal action was the one that was chosen. Unfortunately it is this later distinction that many remember from Simon, and not the more important distinction between procedural and substantive rationality. Simon's point was that he observed that people use a procedural approach to tasks; the introduction of satisficing was merely a way of modelling this (Sent 1997). However, the idea of thresholds, that people only respond to a stimulus when it become sufficiently intense, is often credible and is seen in many simulations (for some examples of this, see Chaps. 22 and 25).

Along with Alan Newell, Simon made a contribution of a different kind to the modelling of humans. He produced a computational model of problem solving in the form of a computer program, which would take complex goals and split them into sub-goals until the sub-goals were achievable (Newell and Simon 1972). The importance of this, from the point of view of this book, is that it was a computational model of an aspect of cognition, rather than one expressed in numerical and analytic form. Not being restricted to models that can be expressed in tractable analytic forms allows a much greater range of possibilities for the representation of human individual and social behaviour. Computational models of aspects of cognition are now often introduced to capture behaviours that are difficult to represent in more traditional analytic models. Computational power is now sufficiently available to enable each represented individual to effectively have its own computational process, allowing a model to be distributed in a similar way to that of the social systems we observe. Thus the move to a distributed and computational approach to modelling social phenomena can be seen as part of a move away from abstract models divorced from what they model towards a more descriptive type of representation.

This shift towards a more straightforward (even 'natural') approach to modelling also allows for more evidence to be applied. In the past anecdotal evidence, in the form of narrative accounts by those being modelled was deemed as "unscientific". One of the reasons that such evidence was rejected is that it could not be used to help specify or evaluate formal models; such narrative evidence could only be used within the sphere of rich human understanding and not at the level of a precise model. Computational simulation allows some aspects of individual's narratives to be used to specify or check the behaviour of agents in a model, as well as the results being more readily interpretable by non-experts. This has let such computational simulations to be used in conjunction with stakeholders in a far more direct way than was previously possible. Chapter 10 looks at this approach.

Herbert Simon did not himself firmly connect the two broad strands of his work: the observation of people's procedures in their social context and their algorithmic modelling in computer models. This is not very surprising as the computational power to run distributed AI models (which are essentially what agent-based simulations are) was not available to him. Indeed these two strands of his work are somewhat in opposition to each other, the one attempting to construct a *general* model of an aspect of cognition (e.g. problem solving) and the other identifying quite specific and limited cognitive procedures. I think it is fair to say that whereas Simon did reject the general economic model of rationality, he did not lose hope of a general model of cognitive processes, which he hoped would be achieved starting from good observation of people. There are still many in the social simulation community who hope for (or assume) the existence of an "off-the-shelf" model of the individuals' cognition which could be plugged into a wider simulation model and get reasonable results. Against any evidence, it is often simply hoped that the details of the individuals' cognitive model will not matter once embedded within a network of interaction. This is an understandable hope, since having to deal with both individual cognitive complexity and social complexity makes the job of modelling social complexity much harder – it is far easier to assume that one or the other does not matter much. Granovetter (1985) addressed precisely this question arguing against both the under-socialised model of behaviour (that it is the individual cognition that matters and the social effects can be ignored) and the over-socialised model (that it is the society that determines behaviour regardless of the individual cognition).

Herbert Simon did not have at his disposal the techniques of individual- and agent-based simulation discussed in this handbook. These allow the formal modelling of socially complex phenomena without requiring the strong assumptions necessary to make an equation-based approach (which is the alternative formal technique) analytically tractable. Without such simulation techniques modellers are faced with a dilemma: either to "shoe-horn" their model into an analytically tractable form, which usually requires them to make some drastic simplifications of what they are representing, or to abandon any direct formal modelling of what they observe. In the later case, without agent-based techniques, they then would have two further choices: to simply not do any formal modelling at all remaining in the world of natural language, or to ignore evidence of the phenomena and instead model their idea concerning the phenomena. In other words, to produce an abstract but strictly *analogical* model – a *way of thinking about the phenomena* expressed as a simulation. This latter kind of simulation does not directly relate to any data derived from observation but to an idea, which, in turn, relates to what is observed in a rich, informal manner. Of course there is nothing wrong with analogical thinking, it is a powerful source of ideas, but such a model is not amenable to scientific testing.

The introduction of accessible agent-based modelling opens up the world of social complexity to formal representation in a more natural and direct manner. Each entity in the target system can be represented by a separate entity (agent or object) in the model, each interaction between entities as a set of messages between the corresponding entities in the model. Each entity in the model can be different, with different behaviours and attributes. The behaviour of the modelled entities can be realised in terms of readily comprehensible rules rather than equations, rules that can be directly compared to accounts and evidence of the observed entities' behaviour. Thus the mapping between the target system and model is simpler and more obvious than when all the interactions and behaviour is "packaged up" into an analytic or statistical model. Formal modelling is freed from its analytical straight jacket, so that the most appropriate model can be formulated and explored. It is no longer necessary to distort a model with the introduction of overly strong assumptions simply in order to obtain analytic tractability. Also, agent-based modelling does not require high levels of mathematical skill and thus is more accessible to social scientists. The outcomes of such models can be displayed and animated in ways that make them more interpretable by experts and stakeholders (for good and ill).

It is interesting to speculate what Herbert Simon would have done if agent-based modelling was available to him. It is certainly the case that it brings together two of the research strands he played a large part in initiating: algorithmic models of aspects of cognition; and complex models that are able to take into account more of the available evidence. We must assume that he would have recognised and felt at home with such kinds of model. It is possible that he would not have narrowed his conception of substantive rationality to that of satisficing if he had other productive ways of formally representing the processes he observed in the way he observed them occurring.

It is certainly true that the battle he fought against "armchair theorising" (working from a neat set of assumptions that are independent of evidence) is still raging. Even in this volume you will find proponents (let's call them the *optimists*) that still hope that they can find some short-cut that will allow them to usefully capture social complexity within abstract and simple models (theory-like models), and those (the *pessimists*) that think our models will have to be complex, messy and specific (descriptive models) if they are going to usefully represent anything we observe in the social world. However, there is now the possibility of debate, since we can compare the results and success of the optimistic and pessimistic approaches and indeed they can learn from each other.

It seems that research into social complexity has reached a cusp, between the "revolutionary" and "normal" phases described by Kuhn (1962). A period of exploratory growth, opposed to previous orthodoxies, has occurred over the last 15–20 years, where it was sufficient to demonstrate a new kind of model, where opening up new avenues was more important than establishing or testing ideas about observed systems. Now attention is increasingly turning to the questions such as: how to productively and usefully simulate social complexity; how to do it with the greatest possible rigour; how to ensure the strongest possible relation to the evidence; how to compare different simulations; how to check them for unintentional errors; how to use simulation techniques in conjunction with others (analytic, narrative, statistical, discourse analysis, stakeholder engagement, data collection etc.). The field – if it is that – is maturing.

This handbook is intended to help in this process of maturation. It brings together summaries of the best thinking and practice in this area, from many of the top researchers. In this way it aims to help those entering into the field so that they do not have to reinvent the wheel. It will help those already in the field by providing accessible summaries of current thought. It aims to be a reference point for best current practice and a standard against which future methodological advances are judged.

## 1.3 The Structure of the Handbook

The material in this book is divided into four sections: *Introductory*, *Methodology*, *Mechanisms* and *Applications*. We have tried to ensure that each chapter within these sections covers a clearly delineated set of issues. To aid the reader each chapter starts with a very brief section called "Why read this chapter?" that sums up the reasons you would read it in a couple of sentences. This is followed by an abstract, which summarises the content of the chapter. Each chapter also ends with a section of "Further Reading" briefly describing things that a newcomer might read next if they are interested. This is separate from the list of references, which contains all the references mentioned in the chapter.

### *1.3.1 Introductory Section*

The introductory section includes three chapters: this chapter, a historical introduction (Chap. 2) that reviews the development of social simulation providing some context for the rest of the book, and an overview of the different kinds of simulation (Chap. 3).

### *1.3.2 Methodology Section*

The next section on methodology consists of nine chapters that aim to guide the reader through the process of simulating complex social phenomena. It starts with two approaches to designing and building simulation models: formal (Chap. 5) and informal (Chap. 4). The former being more appropriate where the goals and specification of the proposed simulation are known and fixed, the latter more appropriate in the case where possible models are being explored, in other words when the simulation model one wants cannot be specified in advance.

However carefully a modeller designs and constructs such simulations they are complex entities, which are difficult to understand completely. The next chapter (6) guides the reader through the ways in which a simulation model can be checked to ensure that it conforms to the programmer's intentions for it. All of the approaches described in these three chapters are aided by good, clear documentation. Chapter 7 describes a way of structuring and performing such documentation that helps to ensure that all necessary information is included without being an overly heavy burden.

Three chapters in this section are concerned with the results of simulations. Chapter 8 concentrates on the validation of simulation models: the many ways in which a model and the possible outputs from simulation runs can be related to data as a check that it is correct for its purpose. Chapter 9 explores ways of analysing and visualising simulation results, which is vital if the programmer or a wider audience is to understand what is happening within complex simulations. Chapter 12 looks at the wider question of the meaning and import of simulations, in other words the philosophy of social simulation including what sort of theorising they imply.

Two other chapters consider separate aspects, but ones that will grow in importance over time. Chapter 10 looks at participatory approaches to simulation, that is ways of involving stakeholders more directly in the model specification and/or development process. This is very different to an approach where the simulation model is built by expert researchers who judge success by the correspondence with data sets, and can almost become an intervention within a social process rather than a representation of it. Chapter 11 investigates how analytic approaches can be combined with simulation approaches, both using analytics to approximate and understand a simulation model as well as using simulation to test the assumptions within an analytic model.

### *1.3.3 Mechanisms Section*

The third section considers types of social mechanisms that have been used and explored within simulations. It does not attempt to cover all such approaches, but concentrates upon those with a richer history of use, where knowing about what has been done might be important and possibly useful.

Chapter 13 takes a critical look at mechanisms that may be associated with economics. Although this handbook is not about economic simulation[1] mechanisms from economics are often used within simulations with a broader intent. Unfortunately, this is often done without thinking so that, for example, an agent might be programmed using a version of economic rationality (i.e. considering options for actions and rating them as to their predicted utility) just because that is what the modellers know or assume. However, since economic phenomena are a subset of social phenomena this chapter does cover these.

Chapter 14 surveys a very different set of mechanisms, those of laws, conventions and norms. This is where behaviour is constrained from outside the individual in some way (although due to some decision to accept the constraint from the inside to differing degrees). Chapter 15 focuses on trust and reputation mechanisms; how people might come to judge that a particular person is someone they want to deal with.

Chapter 16 looks at a broad class of structures within simulations, those that represent physical space or distribution in some way. This is not a cognitive or social mechanism in the same sense of the other chapters in this section, but has implications for the kinds of interactions that can occur, and indeed facilitates some kinds of interaction due to partial isolation of local groups.

The last two chapters in this section examine ways in which groups and individuals might adapt. Learning and evolution are concepts that are not cleanly separable; evolution is a kind of learning by the collection of entities that are evolving and has

[1] There is an extensive handbook on this (Tesfatsion and Judd 2006).

been used to implement learning within an individual (e.g. regarding the set of competing strategies an individual has) as well as within a society. However, Chap. 17 investigates these concepts primarily from the point of view of algorithms for an individual to learn, while Chap. 18 looks at approaches that explicitly take a population and apply some selective pressures upon it, along with adding some sources of variation.

### *1.3.4 Applications Section*

The last section looks at eight areas where the techniques that have been described are being applied. We chose areas where there has been some history of application and hence some experience. Areas of application that are only just emerging are not covered here.

Chapter 19 reviews applications to ecological management. This is one of the oldest and most productive areas where simulation approaches have been applied. Since it is inevitable that the interaction of society and the environment is complex, analytic approaches are usually too simplistic and approaches that are better suited are needed.

Chapter 20 discusses how simulation approaches have begun to inform the design of organisations or its processes. Chapter 21 explores how a simulation-based understanding of ICT systems can enable new kinds of distributed systems to be designed and managed, while Chap. 22 looks at how simulation can help us understand animal interaction. Chapter 23 describes agent-based simulations as a useful tool to come to a complex understanding of how markets actually work (in contrast to their economic idealisations). Chapter 24 considers systems where people and/or goods are being moved within space or networks including logistics and supply chains.

The last two chapters look at understanding human societies. Chapter 25 focuses on a descriptive modelling approach to structures of power and authority, with particular reference to Afghanistan, whereas Chap. 26 reviews the different ways in which simulations have been used to understand human societies, briefly describing examples of each.

## References

Edmonds B (2010) Bootstrapping knowledge about social phenomena using simulation models. J Artif Soc Soc Simulat 13(1). http://jasss.soc.surrey.ac.uk/13/1/8.html

Granovetter M (1985) Economic action and social structure: the problem of embeddedness. Am J Sociol 91(3):481–510

Kuhn TS (1962) The structure of scientific revolutions. University of Chicago Press, Chicago

Newell A, Simon HA (1972) Human problem solving. Prentice-Hall, Englewood Cliffs

Sargent TJ (1993) Bounded rationality in macroeconomics: the Arne Ryde memorial lectures. Clarendon, Oxford

Sent E-M (1997) Sargent versus Simon: bounded rationality unbound. Camb J Econ 21:323–338

Simon HA (1947) Administrative behavior: a study of decision-making processes in administrative organizations. The Free Press, New York

Simon HA (1956) Rational choice and the structure of the environment. Psychol Rev 63(2):129–138

Simon HA (1976) Administrative behavior, 3rd edn. The Free Press, New York

Tesfatsion L, Judd KL (eds) (2006) Handbook of computational economics, vol 2: Agent-based computational economics, vol 13: Handbooks in economics. North Holland, Amsterdam

# Chapter 2
# Historical Introduction

Klaus G. Troitzsch

**Why Read This Chapter?** To understand the historical context of simulation in the social sciences and thus to better comprehend the developments and achievements of the field.

**Abstract** This chapter gives an overview of early attempts at modelling social processes in computer simulations. It discusses the early attempts, their successes and shortcomings and tries to identify some of them as forerunners of modern simulation approaches.

## 2.1 Overview

The chapter is organised as follows: the next section will discuss the early attempts at simulating social processes, mostly aiming at prediction and numerical simulation of mathematical models of social processes. Section 2.3 will then be devoted to the non-numerical and early agent-based approaches, while Sect. 2.4 will give a short conclusion, followed by some hints at further reading.

## 2.2 The First Two Decades

Simulation in the social sciences is nearly as old as computer simulation at large. This is partly due to the fact that some of the pioneers of computer science – such as John von Neumann, one of the founders of game theory – were at the same time pioneers in the formalisation of social science. In addition, Herbert A. Simon, one of the pioneers

K.G. Troitzsch (✉)
Institut für Wirtschafts- und Verwaltungsinformatik, Universität Koblenz-Landau, Koblenz, Germany
e-mail: kgt@uni-koblenz.de

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_2,

in formalising social science, was an early adopter of computer-assisted methods of building social theories. Thus the first two decades of computational social science saw mathematical models and their inelegant solutions, microsimulation and even the first agent-based models before the name of this approach was coined.

Among the first problems tackled with the help of computer simulation were predictions of the future of companies ("industrial dynamics", Forrester 1961), cities ("urban dynamics", Forrester 1969) and the world as a whole ("world dynamics", Forrester 1971) in the early 1960s and 1970s by Jay W. Forrester as well as predictions of the consequences of tax and transfer laws for both the individual household and the national economy in microanalytical simulation, an attempt that started as early as 1956 (Orcutt 1957). Other early attempts at the prediction of elections and referendum campaigns also became known in the 1960s, such as Abelson and Bernstein's simulation analysis of a fluoridation referendum campaign within the Simulmatics project directed by de Sola Pool (Abelson and Bernstein 1963). What all these early simulations have in common is that they were aimed at predicting social and economic processes in a quantitative manner, and that computer simulation was seen as a "substitute for mathematical derivations" (Coleman 1964, p. 528). Despite Simon and others having already taught computers to deal with non-numerical problems as early as 1955 (*"Logic Theorist, the first computer program that solved non-numerical problems by selective search"*, Simon 1996, pp. 189–190), 10 years later Coleman still believed that "the computer cannot solve problems in algebra; it can only carry out computations when actual numbers are fed into it (Coleman 1964, p. 529).

The remainder of this section will give a short overview of system dynamics and microanalytic simulation – simulation approaches that continue to be promoted by learned societies such as the System Dynamics Society and the International Microsimulation Association, each celebrating their 50th anniversary with international conferences held in Boston in July 2007 and Vienna in August 2007, respectively –, before going into the details of some other early models that remained more or less isolated and are now all but forgotten.

*System dynamics* was developed by Jay W. Forrester in the mid-1950s as a tool to describe systems which could be modelled with large sets of difference and differential equations containing functions whose mathematical treatment would have been difficult or impossible. The general idea behind system dynamics was, and is, that a system, without considering its components individually, could be described in terms of its aggregate variables and their changes over time. The best known examples of system dynamics models are Forrester's (1971) and Meadows et al. (1974) world models which were inspired by the Club of Rome and won public attention in the 1970s when they tried to forecast the world population, the natural resources, the industrial and agricultural capital and the pollution until the end of the twenty-first century by describing the annual change of these aggregate variables as functions of their current states and numerous parameters which had some empirical background.

*Microsimulation* was first described in papers by Orcutt (1957) who designed a simulation starting with a (sample of a) given population and simulating the individual fate of all the members of this population (sample) with the help of transition probabilities empirically estimated from official statistics. Transitions represent changes in the circumstances of an individual, e.g. switching to a different job, achieving a higher educational level, marriage, birth of a child or death. These models have mainly been used for predicting demographic changes and the effects of tax and transfer rules. Usually they do not take into account that the overall changes of the aggregated variables of the population (or the sample) may affect individual behaviour. Thus in the sense of Coleman (1990, p. 10) these models neglect the "downward causation" (i.e. the influence of the aggregate on the individual) and focus only on the "upward causation", namely the changes on the macro level, which are the result of the (stochastically simulated) behaviour of the individuals.

The fluoridation referendum campaign model already mentioned above was one of the first models that can be classified as an early predecessor of today's agent-based models. It consisted of a large number of representatives of people living in a community faced with the option of compulsory fluoridation of drinking water – an issue often discussed in the 1960s – which they would have to vote upon at the end of a longish campaign, in which the media and local politicians were publishing arguments in favour of or against this issue. In this model, 500 individuals are exposed to information spread by several communication channels or sources and additionally, they also exchange information among themselves. It depends on their simulated communication habits to which extent they actually receive this information and, moreover, to which extent this leads to changes in their attitudes towards the referendum issue. Abelson and Bernstein defined 51 rules of behaviour, 22 of which are concerned with the processing of information spread over the communication channels, whereas 27 rules are related to the information exchange among individuals. Another two rules determine the final voting behaviour at the end of the referendum campaign. The rules for processing the information from public channels and those for processing the information exchanged among individual citizens are quite similar; rule A3 and rule B2 read, for instance, "Receptivity to [source] *s* is an inverse function of the extremity of [individual] *i*'s attitude position."

While this early model did not endow the model individuals with an appropriate repertoire of behaviours, it nevertheless displays a relatively broad range of communication possibilities among the model individuals – something that was neither aimed at in the classical microanalytical simulation approach, nor in the cellular automata approach adopted in the early 1970s in Thomas Schelling's seminal paper on segregation. One of the shortcomings of Abelson and Bernstein's model in the eyes of its critics was the fact that it "has never been fully tested empirically" (Alker 1974, p. 146). They also contested the adequacy of its "static representations of citizen belief systems defined primarily in terms of assertions held, assertions acceptance predispositions, with associated, more general, conflict levels" (Alker 1974, p. 146). Moreover, the assertions were modelled numerically (not a problem with the proponents of a mathematical sociology who would even have used a large system of differential equations to model the citizens' attitude changes) where obviously real citizens' attitudes were never mapped on to the set of integer or real numbers. Nowak et al. (1990, p. 371) give further reasons for the fact that this approach was dropped for decades, "the ad hoc quality of many of the assumptions of the models, perhaps because of dissatisfaction with the plausibility of their outcomes despite their dependence on extensive parameter estimation, or perhaps because they were introduced at a time when computers were still cumbersome and slow and programming time-consuming and expensive."

Simulmatics suffered basically the same fate as Abelson and Bernstein's model: Simulmatics was set up "for the Democratic Party during the 1960 campaign. . . . The immediate goal of the project was to estimate rapidly, during the campaign, the probable impact upon the public, and upon small strategically important groups within the public, of different issues which might arise or which might be used by the candidates" (de Sola Pool and Abelson 1961, p. 167). The basic components of this simulation model were voter types, 480 of them, not individual voters, with their attitudes towards a total of 48 so-called "issue clusters", i.e. "political characteristics on which the voter type would have a distribution". Voter types were mainly defined by region, agglomeration structure, income, race, religion, gender and party affiliation. From different opinion polls and for different points of time these voter types were attributed four numbers per "issue cluster": the number of voters in this type and "the percentages pro, anti and undecided or confused on the issue" (168). For each voter type empirical findings about cross-pressure (e.g. anti-Catholic voters who had voted for the Democratic Party in the 1958 congressional elections and were likely to stay at home instead of voting for the Catholic candidate of the Democrats) were used during a simulation run to re-adjust the preferences of the voters, type by type. It is debatable whether this would classify as a simulation in current social simulation communities, but since this approach at least in some way resembles the classical static microsimulation, where researchers are interested in the immediate consequences of new tax or transfer laws with no immediate feedback, one could argue that Simulmatics was a simulation project – though with as little sophistication as static microsimulation.

Thus the first two decades of computer simulation in the social sciences were mainly characterised by two beliefs: that computer simulations were nothing but the numerical solution of more adequate mathematical models, and that they were most useful for predicting the outcome of social processes whose first few phases had already been observed. This was also the core of the discussion opened in 1968 by Hayward Alker who analysed, among others, the Abelson-Bernstein community referendum model and came to the conclusion that this "simulation cannot be 'solved': one must project what will be in the media, what elites will be doing, and know what publics already believe before even contingent predictions are made about community decisions. In that sense an open simulation is bad mathematics even if it is a good social system representation." (Alker 1974, p. 153)

## 2.3 Computer Simulation in Its Own Right

The Simulmatics Corporation mentioned in the previous subsection did not only work in the context of election campaigning, but later on also as a consulting agency in other political fields. Their Crisiscom model is another example of an early forerunner of current simulation models of negotiation and decision making processes. At the same time it is an early example of a simulation not aimed at prediction but at "our understanding of the process of deterrence by exploring how far the behaviour of political decision makers in crisis can be explained by psychological mechanisms." (de Sola Pool and Kessler 1965, p. 31) Crisiscom dealt with messages of the type "actor one is related to actor two", where the set of relations was restricted to just two relations: affect and salience. In some way, Crisiscom could also be used as part of a gaming simulation in which one or more of the actors were represented by human players, whereas the others were represented by the computer program – thus it can also be classified as a predecessor of participatory simulation (see Chap. 10 in this volume).

The 1970s and 1980s saw a number of new approaches to simulate abstract social processes, and most of them now were actual computer simulations, as – in terms of Thomas Ostrom – they used the "third symbol system" (Ostrom 1988, p. 384) directly by translating their ideas from the first symbol system, natural language, into higher level programming languages instead of using it as a machine to manipulate symbols of the second symbol system, mathematics. Although this was already true for Herbert Simon's Logic Theorist, the General Problem Solver and other early artificial intelligence programs, the direct use of the "third symbol system" in social science proper was not introduced before the first multilevel models and cellular automata that integrated at least primitive agents in the sense of software modules with some autonomy.

Cellular automata (Farmer et al. 1984; Ilachinski 2001) are a composition of finite automata which all follow the same rule, are ordered in a (mostly) two-dimensional grid and interact with (receive input from) their neighbours. The behaviour of the individual cells is usually quite simple: they only have a small number of states among which they switch according to relatively simple transition rules. Prime example is the famous game of life (Gardener 1970), where the cells are either alive or dead and change state according to two simple rules: (a) a cell stays alive if it has exactly two or three live neighbouring cells, otherwise it dies; (b) a dead cell bursts into life if there are exactly three live cells among its eight neighbours. The great variety of outcomes on the level of the cellular automaton as a whole enthused researchers in complexity science and laid the headstone for innumerable cellular automata in one or two dimensions.

One of the first applications of cellular automata to problems of social science is Thomas Schelling's (1971) segregation model, demo versions of which are nowadays part of any distribution of simulation tools used for programming cellular automata and agent-based models. This model shows impressively that segregation and the formation of ghettos is inevitable even if individuals tolerate a majority of neighbours different from themselves.

Another example is Bibb Latané's Dynamic Social Impact theory with the implementation of the SITSIM model (Nowak and Latané 1994). This model, similar to Schelling's, also ends up in clustering processes and in the emergence of local structures in an initially randomly distributed population, but unlike Schelling's segregation model (where agents move around the grid of a cellular automaton until they find themselves in an agreeable neighbourhood) the clustering in SITSIM comes from the fact that immobile agents adapt their attitudes to the attitudes they find in their neighbourhood according to the persuasive strength of their neighbours.

Other cellular automata models dealt with *n*-person cooperation games and integrated game theory into complex models of interaction between agents and their neighbourhoods. These models, too, usually end up in emergent local structures (Hegselmann 1996).

Another game-theory-related computer simulation, run by Axelrod (1984), showed the Tit-For-Tat strategy in the iterated prisoner's dilemma as superior to all other strategies represented in a computer tournament. The prisoner's dilemma had served game theorists, economists and social scientists as a prominent model of decision processes under restricted knowledge. The idea stems from the early 1950s, first written down by Albert Tucker, and is about "two men, charged with a joint violation of law, [who] are held separately by the police. Each is told that (1) if one confesses and the other does not, the former will be given a reward . . . and the latter will be fined . . . (2) if both confess, each will be fined . . . At the same time, each has good reason to believe that (3) if neither confesses, both will go clear." (Poundstone 1992, pp. 117–118) In the non-iterated version the rational solution is that both confess – but if they believe they can trust each other, they can both win, as both will go clear if neither confesses. Axelrod's question was under which conditions a prisoner in this dilemma would "cooperate" (with his accomplice, not with the police) and under which condition they would "defect" (i.e. confess, get a reward and let the accomplice alone in prison). Strategies in this tournament had to define which choice – cooperate or defect – each player would make, given the history of choices of both players, but not knowing the current decision of the partner. Then every strategy played the iterated game against every other strategy, with identical payoff matrices – and the Tit-For-Tat strategy proved to be superior to 13 other strategies proposed by economists, game theorists, sociologists, psychologists and mathematicians (and it was the strategy that had the shortest description in terms of lines-of-code). Although later on several characteristics of a number of the strategies proposed could be analysed mathematically, the tournament had at least the advantage of easy understandability of the outcomes – which, by the way, is another advantage of the "third symbol system" over the symbol system of mathematics.

Cellular automata later on became the environment of even more complex models of abstract social processes. They serve as a landscape where moving, autonomous, pro-active, goal-directed software agents harvest food and trade with each other. Sugarscape is such a landscape functioning as a laboratory for a "generative social science" (Epstein and Axtell 1996, p. 19) in which the researcher "grows" the emergent phenomena typical for real-world societies in a way that includes the explanation of these phenomena. In this artificial world, software agents find several types of food which they need for their metabolism, but in different proportions, which gives them an incentive to barter one kind of food, of which they have plenty, for another kind of food, which they urgently need.

**Table 2.1** Overview of important approaches to computational social science

| Approach | Used since | Characteristics |
| --- | --- | --- |
| System dynamics | Mid-1950s | Only one object (the system) with a large number of attributes |
| Microsimulation | Mid-1950s | A large number of objects representing individuals that do not interact, neither with each other nor with their aggregate, with a small number of attributes each, plus one aggregating object |
| Cellular automata | Mid-1960s | Large number of objects representing individuals that interact with their neighbours, with a very restricted behaviour rule, no aggregating object, thus emergent phenomena have to be visualised |
| Agent-based models | Early 1990s, with some forerunners in the 1960s | Any number of objects ("agents") representing individuals and other entities (groups, different kinds of individuals in different roles) that interact heavily with each other, with an increasingly rich repertoire of changeable behaviour rules (including the ability to learn from experience and/or others, to change their behavioural rules and to react differently to identical stimuli when the situation in which they are received is different) |

This kind of laboratory gives an insight under which conditions skewed wealth distributions might occur or be avoided; with some extensions (König et al. 2002) agents can even form teams led by agents who are responsible for spreading the information gained by their followers among their group.

## 2.4 Conclusion

This short guided tour through early simulation models tried to show the optimism of the early adopters of this method: "*If it is possible to reproduce, through computer simulation, much of the complexity of a whole society going through processes of change, and to do so rapidly, then the opportunities to put social science to work are vastly increased.*" (de Sola Pool and Abelson 1961, p. 183) 35 years later, Epstein and Axtell formulate nearly the same optimism when they list a number of problems that social sciences have to face – suppressing real-world agents' heterogeneity, neglecting non-equilibrium dynamics and being preoccupied with static equilibria – and claim that "the methodology developed [in Sugarscape] can help to overcome these problems" (Epstein and Axtell 1996, p. 2).

To complete this overview, Table 2.1 lists the approaches touched in this introductory chapter with their main features.

As one can easily see from this table, only the agent-based approach is able to "cover all the world" (Brassel et al. 1997), as only this approach can (a) include the features of all the other approaches, and (b) meet the needs of social science for models of individuals which are able to exchange symbolic messages that have to be interpreted by the recipients before they can take effect. When investigating large-scale social phenomena involving large numbers of individuals in more or less similar situations, then microsimulation, cellular automata, including sociophysics models (Chakrabarti et al. 2006; Ball 2005), or even system dynamics may provide a good (enough) approximation of what happens in human societies. But if we deal with small communities – including the local communities Abelson and Bernstein analysed –, then the process of persuasion, which needs at least one persuasive person and one or more persuadable persons, has to be taken into account, and this calls for agents of a richer structure than the early approaches could provide.

## Further Reading

Most of the literature suggested for further reading has already been mentioned. Epstein and Axtell's (1996) work on generating societies gives a broad overview of early applications of agent-based modelling. Epstein (2006) goes even further as he defines this approach as the oncoming paradigm in social science. For the state of the art of agent-based modelling in the social sciences at the onset of this approach, the proceedings of early workshops and conferences on computational social science are still worth reading (Gilbert and Doran 1994; Gilbert and Conte 1995; Conte et al. 1997; Troitzsch et al. 1996).

And many early papers on computational social science were recently republished (Gilbert 2010).

## References

Abelson RP, Bernstein A (1963) A computer simulation of community referendum controversies. Public Opin Q 27:93–122

Alker HR Jr (1974) Computer simulations: inelegant mathematics and worse social science. Int J Math Educ Sci Technol 5:139–155

Axelrod R (1984) The evolution of cooperation. Basic Books, New York

Ball P (2005) Critical mass – how one thing leads to another. Arrow Books, London

Brassel KH, Möhring M, Schumacher E, Troitzsch KG (1997) Agents cover all the world? In: Conte R, Hegselmann R, Terna P (eds) Simulating social phenomena (Lecture notes in economics and mathematical systems), vol 456. Springer, Berlin, pp 55–72

Chakrabarti BK, Chakraborti A, Chatterjee A (eds) (2006) Econophysics and sociophysics: trends and perspectives. Wiley, Weinheim

Coleman JS (1964) Introduction to Mathematical Sociology. The Free Press of Glencoe, New York

Coleman JS (1990) The foundations of social theory. Harvard University Press, Boston

Conte R, Hegselmann R, Terna P (eds) (1997) Simulating social phenomena (Lecture notes in economics and mathematical systems), vol 456. Springer, Berlin

de Sola Pool I, Abelson RP (1961) The simulmatics project. Public Opin Q 25:167–183 de Sola Pool I, Kessler A (1965) The Kaiser, the Czar, and the computer: information processing in a crisis. Am Behav Sci 8:32–38

Epstein JM (2006) Generative social science: studies in agent-based computational modeling. Princeton University Press, Princeton/Oxford

Epstein JM, Axtell R (1996) Growing artificial societies: social science from the bottom up. Brookings/MIT Press, Washington/Cambridge, MA

Farmer D, Toffoli T, Wolfram S (eds) (1984) Cellular automata: proceedings of an interdisciplinary workshop, Los Alamos, 7–11 Mar 1983. North-Holland, Amsterdam

Forrester JW (1961) Industrial dynamics. MIT/Wright Allen, Cambridge, MA

Forrester JW (1969) Urban dynamics. MIT/Wright Allen, Cambridge, MA

Forrester JW (1971) World dynamics. MIT/Wright Allen, Cambridge, MA

Gardener M (1970) The game of life. Sci Am 223(4):120–123

Gilbert N (ed) (2010) Computational Social Science Four-Volume Set. Sage, Los Angeles

Gilbert N, Conte R (eds) (1995) Artificial societies: the computer simulation of social life. UCL Press, London

Gilbert N, Doran JE (eds) (1994) Simulating societies: the computer simulation of social phenomena. UCL Press, London

Hegselmann R (1996) Cellular automata in the social sciences: perspectives, restrictions, and artefacts. In: Hegselmann R, Mueller U, Troitzsch KG (eds) Modelling and simulation in the social sciences from the philosophy of science point of view. Kluwer, Dordrecht, pp 209–234

Ilachinski A (2001) Cellular automata: a discrete universe. World Scientific, Singapore/River Edge/London/Hong Kong

König A, Möhring M, Troitzsch KG (2002) Agents, hierarchies and sustainability. In: Billari F, Prskawetz-Fürnkranz A (eds) Agent based computational demography. Physica, Berlin/Heidelberg, pp 197–210

Meadows DL et al (1974) Dynamics of growth in a finite world. Wright-Allen, Cambridge, MA

Nowak A, Latané B (1994) Simulating the emergence of social order from individual behaviour. In: Gilbert N, Doran JE (eds) Simulating societies: the computer simulation of social processes. University College of London Press, London, pp 63–84

Nowak A, Szamrej J, Latané B (1990) From private attitude to public opinion: a dynamic theory of social impact. Psychol Rev 97:362–376

Orcutt G (1957) A new type of socio-economic system. Rev Econ Stat 58:773–797

Ostrom TM (1988) Computer simulation: the third symbol system. J Exp Soc Psychol 24:381–392

Poundstone W (1992) Prisoner's dilemma: John von Neumann, game theory, and the puzzle of the bomb. Oxford University Press, Oxford

Schelling TC (1971) Dynamic models of segregation. J Math Sociol 1:143–186

Simon HA (1996) Models of my life. MIT Press, Cambridge, MA

Troitzsch KG, Mueller U, Gilbert N, Doran JE (eds) (1996) Social science microsimulation. Springer, Berlin

# Chapter 3
# Types of Simulation

**Paul Davidsson and Harko Verhagen**

**Why Read This Chapter?** To understand the different ways that computer simulation can differ in terms of (a) purpose, (b) targets for simulation, (c) what is represented, and (d) its implementation; and subsequently, to be more aware of the choices to be made when simulating social complexity.

**Abstract** This chapter describes the main purposes of computer simulation and gives an overview of the main issues that should be regarded when developing computer simulations. While there are two basic ways of representing a system in a simulation model – the equation-based or macroscopic approach and the individual-based or microscopic approach – this chapter (as the rest of the handbook) focuses on the latter. It discusses the various options a modeller faces when choosing how to represent individuals, their interactions and their environment in a simulation model.

## 3.1 Introduction

Simulation concerns the imitation of some aspects of the reality (past, present, or future) for some purpose. We should contrast computer simulation to *physical simulation* in which physical objects are substituted for the real thing. These physical objects are often chosen because they are smaller or cheaper than the actual object or system. When (some of) the objects in a physical simulation are humans, we may refer to this as *human simulation*. However, the focus of this book is on computer simulation, and in particular, computer simulation of social

P. Davidsson (✉)
Department of Computer Science, Malmö University, Malmö, Sweden
e-mail: paul.davidsson@mah.se

H. Verhagen
Stockholm University, Stockholm, Sweden
e-mail: verhagen@dsv.su.se

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_3,

complexity, which concerns the imitation of the behaviour of one or more groups of social entities and their interaction.

Computer simulation, as any other computer program, can be seen as a tool, which could be used professionally, or used in the user's spare time, e.g., when playing computer games. It is possible to distinguish between different types of professional users, e.g. *scientists* who use simulation in the research process to gain new knowledge, *policy makers* who use it for making strategic decisions, *managers (of a system)* who use it to make operational decisions, and *engineers* who use it when developing systems. We can also differentiate two user situations, namely the user as *participant* in the simulation and the user as *observer* of the simulation. Computer games and training settings are examples of the former, where the user is immerged in the simulation. In the case of using simulation as a tool for, say, scientific research or decision support, the user is an outside observer of the simulation. (In other words, we may characterize this difference as that between interactive simulations and batch simulations.)

The main task of computer simulation is the creation and execution of a formal model of the behaviour and interaction (of the entities) of the system being simulated. In scientific research, computer simulation is a research methodology that can be contrasted to empirically driven research.[1] As such, simulation belongs to the same family of research as analytical models. One way of formally modelling a system is to use a mathematical model, and then attempt to find analytical solutions enabling the prediction of the system's behaviour from a set of parameters and initial conditions. Computer simulation, on the other hand, is often used when simple closed form analytic solutions are not possible. Although there are many different types of computer simulation, they typically attempt to generate a sample of representative scenarios for a model in which a complete enumeration of all possible states would be prohibitive or impossible.

It is possible to make a general distinction between two ways of modelling the system to be simulated. One is to use mathematical models and is referred to as *equation-based* (or system dynamics or macro-level) simulation. In such models the set of individuals (the population of the system) is viewed as a structure that can be characterized by a number of variables. In the other way of modelling, which is referred to as *individual-based* (or agent-based or micro-level) simulation, the specific behaviours of specific individuals are explicitly modelled. In contrast to equation-based simulation, the structure is viewed as emergent from the interactions between the individuals and thus exploring the standpoint that complex effects need not have complex causes. We will here, as well as in the remainder of this book, focus on individual-based simulation.

In this chapter we will describe the main purposes of computer simulation and also give an overview of the main issues that should be regarded when developing computer simulations.

[1] This distinction is of course not set in stone. For an example of an evidence-driven approach to computer simulation see Chap. 25 in this volume (Geller and Moss 2013).

## 3.2 Purposes of Simulation

We can identify a number of distinct purposes of simulation. In general terms, simulation is almost always used for analyzing (some aspects of) a system, typically by predicting future states. More specifically, we may say that in the case when the user is *observing* the simulation, the purpose is often one of the following:

– *Management of a system*, where simulation of (parts of) this system is used to support operational decisions, i.e. which action to take, or strategic decisions, i.e. which policy to use. The chapters on application areas in this book provide some examples of this purpose; e.g., Chap. 19 addresses environmental management (Le Page et al. 2013).
– *Design or engineering of a system*, where simulation is used as a tool to support design decisions when developing a system. Chapter 21 illustrates how simulation can help in the design of distributed computer systems (Hales 2013). In fact, many new technical systems are distributed and involve complex interaction between humans and machines, which make individual-based simulation a suitable approach. The idea is to model the behaviour of the human users which is useful in situations where it is too expensive, difficult, inconvenient, tiresome, or even impossible for real human users to test out a new technical system. An example of this is the simulation of "intelligent buildings" where software agents model the behaviour of the people in the building (Davidsson 2000).
– *Evaluation and verification*, where simulation is used to evaluate a particular theory, model, hypothesis, or system, or compare two or more of these. Moreover, simulation can be used to verify whether a theory, model, hypothesis, system, or software is correct. An example of this purpose is found in Chap. 20 of this book on the assessment of (changes in) organizational design (Dignum 2013). More generally, in the context of social theory building, simulations can be seen as an experimental method or as theories in themselves (Sawyer 2003). In the former case, simulations are run e.g. to test the predictions of theories, whereas in the latter case the simulations themselves are formal models of theories. Formalizing the ambiguous, natural language-based theories of the social sciences helps to find inconsistencies and other problems, and thus contributes to theory building.
– *Understanding*, where simulation is used to gain deeper knowledge of a certain domain. In such explorative studies, there is no specific theory, model, etc. to be verified, but we want to study different phenomena (which may however lead to theory refinement). Chapter 22 in this volume provides a number of examples how simulation has helped in understanding animal social behaviour (Hemelrijk 2012).

The focus of this book is on the user as an observer, the role of the user as participant is just touched upon in Chap. 10 on participatory approaches (Barreteau et al. 2013). However, to give a more complete picture, we have identified the following purposes in the case when the user is participating in the simulation:

– *Education*, where simulation is used to explain or illustrate a phenomenon and deepen the user's theoretical knowledge. An example of this is the recently developed SimPort,[2] a multiplayer serious game where the players have to construct a port area in the vicinity of Rotterdam. One aim of this simulation-based tool is to give its users better insight into any unforeseen, undesirable and unintentional effects of one or more development strategies and design variations in the medium term (10–30 years) as a result of exogenous uncertainties (economic, market, technological) and due to strategic behaviour of the parties involved. Another example of individual-based simulation for educational purpose is the PSI agent (Künzel and Hämmer 2006) that supports acquiring theoretical insights in the realm of psychological theory. It enables students to explore psychological processes without ethical problems.
– *Training*, where simulation is used to improve a person's practical skills in a certain domain. The main advantage of using simulation for training purposes is to be part of a real-world-like situation without real-world consequences. An early work in this area was a tool to help train police officers to manage large public gatherings, such as crowds and protest marches (Williams 1993). Another example of agent-based simulation for training purposes is Steve, an agent integrated with voice synthesis software and virtual reality software providing a very realistic training environment. For instance, it has been applied to maintenance tasks in nuclear power plants (Méndez et al. 2003).
– *Entertainment*, where simulation is used just to please the user. There are a large number of popular simulation games available. These belong to genres like *construction and management simulations*, where players experience managing a government, a sports team, a business, or a city, *life simulations*, where players manage a life-form or ecosystem, such as the well-known "Sims" and its sequels, *vehicle simulations*, where players experience driving a vehicle, such as an airplane or a racing car, and of course different types of *war games*.

## 3.3 Types of Systems Simulated

It is possible to categorize the systems being simulated:

1. *Human-centered systems*, such as:

   – *Human societies*, consisting of a set of persons with individual goals. That is, the goal of different individuals may be conflicting. In Chap. 26 of this book, more information on the simulation of human societies is given (Edmonds et al. 2013).

[2] http://www.simport.eu/

  - *Organizations*, which we here define as structures of persons related to each other in order to purposefully accomplishing work or some other kind of activity. That is, the persons of an organization share some of their goals. Further details on the modelling and simulation of organizations are provided in Chap. 20 (Dignum 2013).
  - *Economic systems*, which are organized structures in which actors (individuals, groups, or enterprises) are trading goods or services on a market. Chapter 23 (Rouchier 2013) takes a closer look at markets.

2. *Natural systems*, such as:

  - *Animal societies*, which consist of a number of interacting animals, such as an ant colony or a colony of birds. Chapter 22 (Hemelrijk 2013) is devoted to simulation of animal societies.
  - *Ecological systems*, in which animals and/or plants are living and evolving in a relationship to each other and in dependence of the environment (even if humans also are part of the ecological system, they are often not part of these simulation models). In Chap. 19 (Le Page et al. 2013) more details on the simulation of ecological systems are discussed.

3. *Socio-technical systems*, which are hybrid systems consisting of both living entities (in most cases humans) and technical artefacts interacting with each other. Examples of this type of system are transportation and traffic systems concerning the movement of people, or goods in a transportation infrastructure such as a road network. Chapter 24 (Ramstedt et al. 2013) provides a review of simulation studies in these areas.
4. *Artificial societies*, which consist of a set of software and/or hardware entities, i.e. computer programs and/or robots, with individual goals. One type of artificial societies, namely distributed computer systems, is treated in Chap. 21 (Hales 2013).

In addition, there are systems that are interesting to simulate using a micro-level approach, but that we do not regard as social systems and are therefore not treated in this book. One class of such systems are *physiological systems*, which consist of functional organs integrated and co-operating in a living organism, e.g. subsystems of the human body. *Physical systems*, which are collections of passive entities following only physical laws, constitute another type of non-social systems.

## 3.4 Modelling

Let us now focus on how to model the system to be simulated. This depends on the type of system and the purpose of the simulation study. An individual- or agent-based model of a system consists of a set of entities and an environment in which the entities are situated. The entities are either *individuals* (agents) that have some decision-making capabilities, or *objects* (resources) that have no agency and are purely physical. There are a number of characteristics that can be used to differentiate between different types of models. We will first look at how individuals are being modelled, then on the interaction between the individuals, and finally how the environment is being modelled.

### *3.4.1 Individuals*

A model of an individual can range from being very simple, such a one binary variable (e.g. alive or dead) that is changed using only a single rule, to being very complex. The complexity of the model for a given simulation should be determined by the complexity of the individuals being simulated. Note, however, that very complex collective behaviour could be achieved from very simple individual models, if the number is sufficiently large.

We can distinguish between modelling the *state* of an individual and the *behaviour* of the individual, i.e. the decisions and actions it takes. The state of an individual, in turn, can be divided into the *physical* and the *mental* state. The description of the physical state may include the position of the individual, and features such as age, sex, and health status. The physical state is typically modelled as a feature vector, i.e. a list of attribute/value pairs. However, this is not always the case as in some domain the physical state of individual is not modelled at all. An example is the PSI agent mentioned earlier that was used to give students theoretical insights in the area of psychological theory.

Whereas the physical state is often simple to model, representing the mental state is typically much more complex, especially if the individuals modelled are human beings. A common approach is to model the beliefs, desires, and intentions of the individual, for instance by using the BDI model (Bratman 1987; Georgeff et al. 1998). Such a model may include the social state of the individual, i.e. which norms it adheres to, which coalitions it belongs to, etc. Although the BDI model is not based on any experimental evidence of human cognition it has proven to be quite useful in many applications. There has also been some work on incorporating emotions in models of the mental state of individuals (cf. Bazzan and Bordini 2001) as well as obligations, like the BOID model (Broersen et al. 2001), which extends the BDI with obligations.

Modelling the behaviours (and decisions) of the individuals can be done in a variety of ways, from simple probabilities to sophisticated reasoning and planning mechanisms. As an example of the former we should mention *dynamic microsimulation* (Gilbert and Troitzsch 2005), which was one of the first ways of performing individual-based simulation and is still frequently used. The purpose is to simulate the effect the passing of time has on individuals. Data (feature vectors) from a random sample from the population is used to initially characterize the simulated individuals. A set of *transition probabilities* is then used to describe how these features will change over a time period, e.g. there is a probability that an employed person becomes unemployed during a year. The transition probabilities are applied to the population for each individual in turn, and then repeatedly re-applied for a number of simulated time periods. In traditional micro-simulation, the behaviour of each individual is regarded as a "black box". The behaviour is modelled in terms of probabilities and no attempt is made to justify these in terms of individual preferences, decisions, plans, etc. Thus, better results may be gained if also the cognitive processes of the individuals were simulated.

Opening the black box of individual decision-making can be done in several ways. A basic and common approach is to use decision rules, for instance, in the form of a set of situation-action rules: If an individual and/or the environment is in state X then the individual will perform action Y. By combining decision rules and the BDI model quite sophisticated behaviour can be modelled. Other models of individual cognition used in agent-based social simulation include the use of Soar, a computer implementation of Allen Newell's unified theory of cognition (Newell 1994), which was used in Steve (discussed above). Another unified theory of individual cognition, for which a computer implementation exists, is ACT-R (Anderson et al. 2004), which is realized as a production system. A less general example is the Consumat model (Janssen and Jager 1999), a meta-model combining several psychological theories on decision making in a consumer situation. In addition, non-symbolic approaches such as neural networks have been used to model the agents' decision making (Massaguer et al. 2006).

As we have seen, the behaviour of individuals could be either *deterministic* or *stochastic*. Also, the *basis* for the behaviour of the individuals may vary. We can identify the following categories:

- *The state of the individual itself:* In most social simulation models the physical and/or mental state of an individual plays an important role in determining its behaviour.
- *The state of the environment:* The state of the environment surrounding the individual often influences the behaviour of an individual. Thus, an individual may act differently in different contexts although its physical and mental state is the same.
- *The state of other individuals:* One popular type of simulation model, where the behaviour of individuals is (solely) based on the state of other individuals, is those using *cellular automata* (Schiff 2008). Such a simulation model consists of a grid of cells representing individuals, each in one of a finite number of states. Time is discrete and the state of a cell at time $t$ is a function of the states of a finite number of cells (called its neighbourhood) at time $t - 1$. These neighbours are a fixed selection of cells relative to the specified cell. Every cell has the same rule for updating, based on the values in its neighbourhood. Each time the rules are applied to the whole grid a new generation is created. In this case, information about the state of other individuals can be seen as gained through observations. Another possibility to gain this information is through communication, and in this case the individuals do not have to be limited to the neighbours.

– *Social states (norms* etc.*) as viewed by the agent:* For simulation of social behaviour the agents need to be equipped with mechanisms for reasoning at the social level (unless the social level is regarded as emergent from individual behaviour and decision making). Several models have been based on theories from economy, social psychology, sociology, etc. Guye-Vuillème (2004) provides an example of this with his agent-based model for simulating human interaction in a virtual reality environment. The model is based on sociological concepts such as roles, values, and norms and motivational theories from social psychology to simulate persons with social identities and relationships.

In most simulation studies, the behaviour of the individuals is *static* in the sense that decision rules or reasoning mechanisms do not change during the simulation. However, human beings and most animals do have an ability to adapt and learn. To model *dynamic* behaviour of individuals through learning/adaptation can be done in many ways. For instance, both ACT-R and Soar have learning built in. Other types of learning include the internal modelling of individuals (or the environment) where the models are updated more or less continuously.

Finally, there are some more general aspects to consider when modelling individuals. One such aspect is whether all agents share the same behaviour or whether they behave differently, in other words, representation of behaviour is either *individual* or *uniform*. Another general aspect is the number of individuals modelled, i.e. the *size* of the model, which may vary from a few individuals to billions of individuals. Moreover, the population of individuals could be either *static* or *dynamic*. In dynamic populations, changes in the population are modelled, typically births and deaths.

### 3.4.2 Interaction Between Individuals

In dynamic micro-simulation simulated individuals are considered in isolation without regard to their interaction with others. However, in many situations the interaction between individuals is crucial for the behaviour at system level. In such cases better results will be achieved if the interaction between individuals was included in the model. Two important aspects of interaction are (a) who is interacting with whom, i.e. the *interaction topology*, and (b) the *form* of this interaction.

A basic form of interaction is *physical interaction* or interaction based on spatial proximity. As we have seen, this is used in simulations based on cellular automata, e.g. in the well-known Game of Life (Gardner 1970). The state of an individual is determined by how many of its neighbours are alive. Inspired by this work researchers developed more refined models, often modelling the social behaviour of groups of animals or artificial creatures. One example is the Boid model by Reynolds (1987), which simulates coordinated animal motion such as bird flocks and fish schools in order to study emergent phenomena. In these examples, the interaction topology is limited to the individuals immediately surrounding an individual. In other cases, as we will see below, the interaction topology is defined more generally in terms of a *(social) network*. Such a network can be either *static*, i.e. the topology does not change during a simulation, or *dynamic*. In these networks interaction is typically *language-based*. An example is the work by Verhagen (2001), where agents that are part of a group use direct communication between the group members to form shared group preferences regarding the decisions they make. Communication is steered by the structure of the social network regardless of the physical location of the agents within the simulated world. For a more detailed discussion of the different options to model interaction topologies see Chap. 16 in this volume (Amblard and Quattrochiocchi 2013).

### *3.4.3 The Environment*

The state of the environment is usually represented by a set of (global) parameters, e.g. temperature. In addition, there are a number of important aspects of the environment model, such as:

– *Spatial explicitness:* In some models, there is actually no notion of physical space at all. An example of a scenario where location is of less importance are "innovation networks" (Gilbert et al. 2001). Individual agents are high-tech firms that each have a knowledge base used to develop artefacts to launch on a simulated market. The firms are able to improve their products through research or by exchanging knowledge with other firms. However, in many scenarios location is very important, thus each individual (and sometimes objects) is assigned a specific location at each time step of the simulation. In this case, the individuals may be either static (the entity does not change location during the simulation) or mobile. The location could either be specified as an *absolute position* in the environment, or in terms of *relative positions* between entities. In some areas the simulation software is integrated with a Geographical Information System (GIS) in order to achieve closer match to reality (cf. Schüle et al. 2004).
– *Time:* There are in principle two ways to address time, and one is to ignore it. In static simulation time is not explicitly modelled; there is only a "before" and an "after" state. However, most simulations are dynamic, where time is modelled as a sequence of time steps. Typically, each individual may change state between each time step.
– *Exogenous events:* This is the case when the state of the environment, e.g. the temperature, changes without any influence/action from the individuals. Exogenous events, if they are modelled, may also change the state of entities, e.g. decay of resources, or cause new entities to appear. This is a way to make the environment stochastic rather than deterministic.

### 3.4.4 Factors to Consider When Choosing a Model

In contrast to some of the more traditional approaches, such as system dynamics, individual-based modelling does not yet have any standard procedures that can support the model development. (Although some attempts in this direction have been made, e.g. by Grimm et al. (2006) in the area of ecological systems.) In addition, it is often the case that the only formal description of the model is the actual program code. However, it may be useful to use the Unified Modelling Language (UML) to specify the model.

Some of the modelling decisions are determined by the features of the system to be simulated, in particular those regarding the interaction model and the environment model. The hardest design decision is often how the mental state and the behaviour of individuals should be modelled, in particular when representing human beings. For simpler animals or machines, a feature vector combined with a set of transitions rules is often sufficient. Depending on the phenomena being studied, this may also be adequate when modelling human beings. Gilbert (2006) provides some guidelines whether a more sophisticated cognitive model is necessary or not. He states that the most common reason for ignoring other levels is that the properties of these other levels can be assumed constant, and exemplifies this by studies of markets in equilibrium where the preferences of individual actors are assumed to remain constant. (Note, however, that this may not always be true). Another reason for ignoring other levels, according to Gilbert, is when there are many alternative processes at the lower level, which could give rise to the same phenomenon at the macro level. He illustrates this with the famous study by Schelling (1971) regarding residential segregation. Although Schelling used a very crude model of the mental state and behaviour of the individuals, i.e. ignoring the underlying motivations for household migration, the simulation results were valid (as the underlying motivations were not relevant for the purpose of Schelling's study).

On the other hand, there are many situations where a more sophisticated cognitive model is useful, in particular when the mental state or behaviour of the individual constraints or in other ways influences the behaviour at the system level. However, as Gilbert concludes, the current research is not sufficiently mature in order to give advice on which cognitive model to use (BDI, Soar, ACT-R, or other). Rather, he suggests that more pragmatic considerations should guide the selection.

The model of the environment is mostly dictated by the system to be simulated, with the modeller having to decide on the granularity of the values the environmental attributes can take. The interaction model is often chosen based on the theory or practical situation that lies at the heart of the simulation, but sometimes the limitations of the formal framework used restrict the possibilities. Here the modeller also has to decide upon the granularity of attribute values.

## 3.5 Implementation

We will now discuss some issues regarding the implementation (programming and running) of a simulator.

A simulator can be *time-driven*, where the simulated time is advanced in constant time steps, or *event-driven*, where the time is advanced based on the next event. In an event-driven simulation, a simulation engine drives the simulation by continuously taking the first event out of a time-ordered event list, and then simulating the effects on the system state caused by this event. Since time segments where no event takes place are not regarded, event-driven simulation is often more efficient than time-driven simulation. On the other hand, since time is incremented at a constant pace during a simulation in time-driven mode, this is typically a better option if the simulation involves user participation.

There are a number of *platforms* or toolkits for agent-based simulation available, such as Swarm, NetLogo and RePast (see Railsback et al. (2006) for a critical review of these and some other platforms). These are freely available, simplify the programming and can be of great help, in particular for modellers that are not skilled programmers. However, they all impose some limitations on what can be modelled, which may or may not be crucial for the application at hand. An approach without such limitation is of course to program the simulator from scratch using ordinary programming languages like Java or C, which is more difficult and time consuming. In some cases, e.g. if you want to distribute the simulation on a number of computers, it may be appropriate to use an agent platform, such as JADE. In this case, the individuals may be implemented as actual software agents. In particular, when the number of individuals simulated is large and/or the models of individuals are complex, it may be too time consuming to run the simulation on a single computer. Instead, one may distribute the computational load on several computers in order to get reasonable running times. It should be mentioned that there are some efforts on making agent-based simulation platforms run on large-scale computer networks such as Grids, see e.g. the work by Chen et al. (2008).

It is worth noting that the resulting software is an approximation of a simulation model, which in turn is an approximation of the actual system. Thus, there are several steps of verification and validation that need to be addressed in the development of a simulation model, as discussed in Chap. 8 (David 2013).

## 3.6 Conclusion

As we have seen, there are many different types of individual-based social simulation. In the table below, we provide a summary.

| Focus | Aspect | Options |
|---|---|---|
| Usage | Users | Scientists |
| | | Policy makers |
| | | Managers |
| | | Non-professionals |
| | Purposes | Management of a system |
| | | Design or engineering of a system |
| | | Evaluation and verification |
| | | Understanding |
| | | Education |
| | | Training |
| | | Entertainment |
| System simulated | Human-centered systems | Human societies |
| | | Organizations |
| | | Economic systems |
| | Natural systems | Animal societies |
| | | Ecological systems |
| | Socio-technical systems | |
| | Artificial systems | |
| Individual model | Individual physical state | Feature vector |
| | Individual mental state | Feature vector |
| | | BDI |
| | Individual behaviour | Transition probabilities |
| | | Decision rules |
| | | Cognitive model (Soar, ACT-R, etc.) |
| | Basis of behaviour | Own state |
| | | State of the environment |
| | | State of other individuals |
| | | Social states |
| | Uniformity | Uniform/non-uniform |
| | Population | Static/dynamic |
| Interaction model | Form of interaction | No interaction |
| | | Physical |
| | | Language-based |
| | Interaction topology | Static/dynamic |
| | | Neighbourhood/network |
| Environment model | Spatial explicitness | None |
| | | Relative positions |
| | | Absolute positions |
| | Time | Static/dynamic |
| | Exogenous events | Yes/no |
| Implementation | Simulation engine | Time-driven/event-driven |
| | Programming | MABS platform (NetLogo, Repast, etc.) |
| | | MAS platform (JADE, etc.) |
| | | From scratch (C, Java, etc.) |
| | Distributedness | Single computer/distributed |

## Further Reading

Gilbert and Troitzsch (2005) also have sections that describe the different kinds of simulation available. Railsback and Grimm (2011) present a complementary analysis, coming from ecological modelling. The introductory chapters in (Gilbert and Doran 1994) and (Conte and Gilbert 1995) map out many of the key issues and aspects in which social simulation has developed.

## References

Amblard F, Quattrociocchi W (2013) Social networks and spatial distribution. Chapter 16 in this volume

Anderson JR et al (2004) An integrated theory of the mind. Psychol Rev 111(4):1036–1060

Barreteau O et al (2013) Participatory approaches. Chapter 10 in this volume

Bazzan ALC, Bordini RH (2001) A framework for the simulation of agents with emotions: report on experiments with the iterated prisoners dilemma. In fifth international conference on autonomous agents, Montreal. ACM Press, New York, pp 292–299

Bratman ME (1987) Intentions, plans, and practical reason. Harvard University Press, Cambridge, MA

Broersen J, Dastani M, Huang Z, Hulstijn J, Van der Torre L (2001) The BOID architecture: conflicts between beliefs, obligations, intentions and desires. In fifth international conference on autonomous agents, Montreal. ACM Press, New York, pp 9–16

Chen D et al (2008) Large-scale agent-based simulation on the grid. Future Gen Comput Syst 24 (7):658–671

Conte R, Gilbert N (eds) (1995) Artificial societies: the computer simulation of social life. UCL Press, London

David N (2013) Validating simulations. Chapter 8 in this volume

Davidsson P (2000) Multi agent based simulation: beyond social simulation. In: Moss S, Davidsson P (eds) Multi agent based simulation (Lecture notes in computer science, 1979). Springer, Berlin, pp 98–107

Dignum V (2013) Organisational design. Chapter 20 in this volume

Edmonds B, Lucas P, Rouchier J, Taylor R (2013) Human societies: understanding observed social phenomena. Chapter 26 in this volume

Gardner M (1970) Mathematical games: the fantastic combinations of John Conway's new solitaire game "Life". Sci Am 223(4):120–124

Geller A, Moss S (2013) Modeling power and authority: an emergentist view from Afghanistan. Chapter 25 in this volume

Georgeff M, Pell B, Pollack M, Tambe M, Wooldridge M (1998) The belief-desire-intention model of agency. In: Muller J, Singh M, Rao A (eds) Intelligent agents V (Lecture notes in artificial intelligence, 1555). Springer, Berlin, pp 1–10

Gilbert N (2006) When does social simulation need cognitive models? In: Sun R (ed) Cognition and multi-agent interaction: from cognitive modelling to social simulation. Cambridge University Press, Cambridge, pp 428–432

Gilbert N, Doran J (eds) (1994) Simulating societies. UCL Press, London

Gilbert N, Troitzsch KG (2005) Simulation for the social scientist, 2nd edn. Open University Press/McGraw Hill Education, Maidenhead

Gilbert N, Pyka A, Ahrweiler P (2001) Innovation networks: a simulation approach. J Artif Soc Soc Simulat 4(3). http://jasss.soc.surrey.ac.uk/4/3/8.html

Grimm V et al (2006) A standard protocol for describing individual-based and agent-based models. Ecol Model 198:115–126

Guye-Vuillème A (2004) Simulation of nonverbal social interaction and small groups dynamics in virtual environments (Ph.D. thesis). Ècole Polytechnique Fédérale de Lausanne, No 2933

Hales D (2013) Distributed computer systems. Chapter 21 in this volume

Hemelrijk C (2013) Animal social behaviour. Chapter 22 in this volume

Janssen MA, Jager W (1999) An integrated approach to simulating behavioural processes: a case study of the lock-in of consumption patterns. J Artif Soc Soc Simulat 2(2). http://jasss.soc.surrey.ac.uk/2/2/2.html

Künzel J, Hämmer V (2006) Simulation in university education: the artificial agent PSI as a teaching tool. Simulation 82(11):761–768

Le Page C et al (2013) Environmental management. Chapter 19 in this volume

Massaguer D, Balasubramanian V, Mehrotra S, Venkatasubramanian N (2006) Multi-agent simulation of disaster response. In: Jennings NR, Tambe M, Ishida T, Ramchurn SD (eds) First international workshop on agent technology for disaster management. Hakodate, 8 May 2006, pp 124–130. http://users.ecs.soton.ac.uk/sdr/atdm/ws34atdm.pdf

Méndez G, Rickel J, de Antonio A (2003) Steve meets jack: the integration of an intelligent tutor and a virtual environment with planning capabilities. In: intelligent virtual agents (Lecture notes on artificial intelligence, 2792). Springer, Berlin, pp 325–332

Newell A (1994) Unified theories of cognition. Harvard University Press, Cambridge, MA

Railsback SF, Grimm V (2011) Agent-based and individual-based modeling: a practical introduction. Princeton University Press, Princeton

Railsback SF, Lytinen SL, Jackson SK (2006) Agent-based simulation platforms: review and development recommendations. Simulation 82(9):609–623

Ramstedt L, Törnquist Krasemann J, Davidsson P (2013) Movement of people and goods. Chapter 24 in this volume

Reynolds CW (1987) Flocks, herds, and schools: a distributed behavioural model. Comput Graph 21(4):25–34

Rouchier J (2013) Markets. Chapter 23 in this volume

Sawyer RK (2003) Artificial societies: multi-agent systems and the micro–macro link in sociological theory. Sociol Methods Res 31(3):325–363

Schelling TC (1971) Dynamic models of segregation. J Math Sociol 1:143–186

Schiff JL (2008) Cellular automata: a discrete view of the world. Wiley, Oxford

Schüle M, Herrler R, Klügl F (2004) Coupling GIS and multi-agent simulation: towards infrastructure for realistic simulation. In: Lindemann G, Denzinger J, Timm IJ, Unland R (eds) Multiagent system technologies, second German conference, MATES 2004 (LNCS, 3187). Springer, Berlin, pp 228–242

Verhagen H (2001) Simulation of the learning of norms. Soc Sci Comput Rev 19(3):296–306

Williams R (1993) An agent based simulation environment for public order management training. In: Western simulation multiconference, object-oriented simulation conference, Hyatt Regency, La Jolla, 17–20 Jan 1993, pp 151–156

# Part II
# Methodology

# Chapter 4
# Informal Approaches to Developing Simulation Models

**Emma Norling, Bruce Edmonds, and Ruth Meyer**

**Why Read This Chapter?** To get to know some of the issues, techniques and tools involved in building simulation models in the manner that probably most people in the field do this. That is, not using the "proper" computer science techniques of specification and design, but rather using a combination of exploration, checking and consolidation.

**Abstract** This chapter describes the approach probably taken by most people in the social sciences when developing simulation models. Instead of following a formal approach of specification, design and implementation, what often seems to happen in practice is that modellers start off in a phase of exploratory modelling, where they don't have a precise conception of the model they want but a series of ideas and/or evidence they want to capture. They then may develop the model in different directions, backtracking and changing their ideas as they go. This phase continues until they think they may have a model or results that are worth telling others about. This then is (or at least should be) followed by a consolidation phase where the model is more rigorously tested and checked so that reliable and clear results can be reported. In a sense what happens in this later phase is that the model is made so that it is *as if* a more formal and planned approach had been taken.

There is a danger of this approach: that the modeller will be tempted by apparently significant results to rush to publication before sufficient consolidation has occurred. There may be times when the exploratory phase may result in useful and influential personal knowledge but such knowledge is not reliable enough to be up to the more exacting standards expected of publicly presented results. Thus it is only with careful consolidation of models that this informal approach to building simulations should be undertaken.

---

E. Norling (✉) • B. Edmonds (✉) • R. Meyer
Manchester Metropolitan University, Centre for Policy Modelling, Business School, Oxford Road, Manchester M15 6BH, United Kingdom
e-mail: norling@acm.org; bruce@edmonds.name; r.meyer@mmu.ac.uk

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_4,

## 4.1 Introduction: Exploration and Consolidation Modelling Phases

Formal approaches to the development of computer programs have emerged through the collective experience of computer scientists (and other programmers) over the past half-century. The experience has shown that complex computer programs are very difficult to understand: once past a certain point, unless they are very careful, programmers lose control over the programs they build. Beyond a certain stage of development, although we may understand each part, each micro-step, completely we can lose our understanding of the program as a whole: the effects of the interactions between the parts of a program are unpredictable; they are emergent. Thus computer science puts a big emphasis on techniques that aim to ensure that the program does what it is intended to do as far as possible. However, even with the most careful methodology it is recognised that a large chunk of time will have to be spent debugging the program – we all *know* that a program cannot be relied on until it has been tested and fixed repeatedly.

However, it is fair to say that most computational modellers do not follow such procedures and methodologies all the time (although since people don't readily admit to how messy their implementation process actually is, we cannot know this, just as one does not know how messy people's homes are when there are no visitors). There are many reasons for this. Obviously those who are not computer scientists may simply not know these techniques. Then there are a large number of modellers who know of these techniques to some degree but judge that they are not necessary or not worth the effort. Such a judgement may or may not be correct. Certainly it is the case that people have a tendency to underestimate the complexity of programming and so think they can get away with not bothering with a more careful specification and analysis stage. There may also be times when there are good reasons not to follow such techniques.

A specification and design approach is simply not possible if you don't have a very clear idea of your goal. When modelling some target phenomena, one simply does not know beforehand which parts of the system will turn out to be important to the outcomes or even possible to computationally model. One of the big benefits of modelling phenomena computationally is that one learns a lot about what is crucial and possible *in the process of building a simulation model*. This is very unlike the case where one has a functional goal or specification for a program that can be analysed into sub-goals and processes etc. In (social) simulation, the degree to which formal approaches are useful depends somewhat on the goal of modelling. If the goal is very specific, for example understanding the effect of the recovery rate on the change in the number of infections in an epidemic, and the basic model structure is known then what is left is largely an *engineering* challenge. However, if the goal is general understanding of a particular process then there is no possible way of *systematically* determining what the model should be. Here the modelling is essentially a *creative* process, and the development of the model proceeds in parallel with the development of the understanding of the process; the model is itself a theory under development.

Thus what often seems to happen in practice is that modellers start off in a phase of exploratory modelling, where they don't have a precise conception of the model they want but a series of ideas and/or evidence they want to capture. They then may develop the model in different directions, backtracking and changing their ideas as they go. This phase continues until they think they may have a model or results that are worth telling others about. This then is (or at least should be) followed by a consolidation phase where the model is more rigorously tested and checked so that reliable and clear results can be reported. In a sense what happens in this later phase is that the model is made so that it is *as if* a more formal and planned approach had been taken.

There is nothing wrong with using an exploratory approach to model development. Unfortunately, it is common to see models and results that are publicly presented without a significant consolidation phase being undertaken. It is very understandable why a researcher might want to skip the consolidation phase: they may have discovered a result or effect that they find exciting and not wish to go through the relatively mundane process of checking their model and results. They may *feel* that they have discovered something that is of more general importance – however this *personal* knowledge, that may well inform their understanding, is not yet of a standard that makes it worthwhile for their peers to spend time understanding, *until* it has been more rigorously checked.

One of the problems with the activity of modelling is that it *does* influence how the modeller thinks. Paradoxically, this can also be one of the advantages of this approach. After developing and playing with a model over a period of time it is common to "see" the world (or at least the phenomena of study) in terms of the constructs and processes of that model. This is a strong version of Kuhn's "theoretical spectacles" (Kuhn 1969). Thus it is common for modellers to be convinced that they have found a *real* effect or principle during the exploration of a model, despite not having subjected their own model and conception to sufficient checking and testing – what can be called *modelling spectacles*. Building a model in a computer is almost always in parallel with the development of one's ideas about the subject being modelled. This is why it is almost inevitable that we think about the subject in terms of our models – this is at once a model's huge advantage but also disadvantage. As long as one is willing to be aware of the modelling spectacles and be critical of them, or try many different sets of modelling spectacles, the disadvantage can be minimised.

Quite apart from anything, presenting papers with no substantial consolidation is unwise. Such papers are usually painfully obvious when presented at workshops and easily criticised by referees and other researchers if submitted to a journal. It is socially acceptable that a workshop paper will not have as much consolidation as might be required of a journal article, since the criticism and evaluation of ideas and models at a workshop is part of its purpose, but presenting a model with an inadequate level of consolidation just wastes the other participants' time.

What steps then should modellers who follow such an informal approach take to ensure that their model *is* sufficiently consolidated to present to a wider audience? The first step is to ensure that they have a clear purpose for their model, as described below. Secondly, the modeller must be careful to identify the assumptions that are made during the construction of the model. Thirdly, the modeller must maintain control of the model while exploring different possibilities. And fourthly – and this is perhaps the

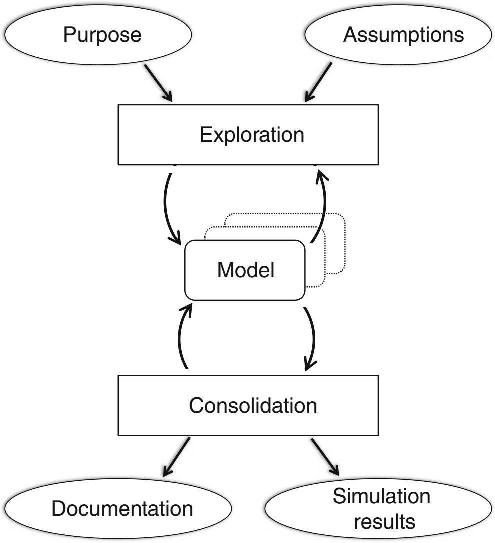



**Fig. 4.1** The exploration and consolidation approach to model development

most difficult – the modeller must maintain an understanding of the model. The following sections of this chapter discuss these points in more detail. Then there is the all-important consolidation phase (which may proceed in parallel with the former steps, rather than strictly sequentially), during which the modeller formalises the model in order to ensure that the results are sound and meaningful. Figure 4.1 illustrates this approach to model building.

## 4.2 Knowing the Purpose of the Model

There are many possible purposes for constructing a model. Although some models might be adapted for different purposes without too much difficulty, at any one time a model will benefit from having a clear purpose. One of the most common criticisms of modelling papers (after a lack of significant consolidation) is that the author has made a model but is unclear as to its purpose. This can be for several reasons, such as:

- The author may have simply modelled without thinking about why. (For example having vague ideas about a phenomenon, the modeller decides to construct a model without thinking about the questions one might want to answer about that phenomenon.)
- The model might have been developed for one purpose but is being presented as if it had another purpose.
- The model may not achieve any particular purpose and so the author might be forced into claiming a number of different purposes to justify the model.

The purpose of a model will affect how it is judged and hence *should* influence how it is developed and checked.

The classic reason for modelling is to predict some unknown aspect of observed phenomena – usually a future aspect. If you can make a model that does this for unknown data (data not known to the modellers before they published the model) then there can be no argument that such a model is (potentially) useful. Due to the fact that predictive success is a very strong test of a model *for which the purpose is prediction*, this frees one from an *obligation* as to the content or structure of the model.[1] In particular the assumptions in the model can be quite severe – the model can be extremely abstract as long as it actually does predict.

However, predictive power will not *always* be a measure of a model's success. There are many other purposes for modelling other than prediction. Epstein (2008) lists 16 other purposes for building a model, e.g. explanation, training of practitioners or education of the general public, and it is important to note that the measure of success will vary depending on the purpose.

With an explanatory model, if one has demonstrated that a certain set of assumptions can result in a set of outcomes (for example by exhibiting an acceptable fit to some outcome data) this shows that the modelled process is a *possible* explanation for those outcomes. Thus the model generates an explanation, but only in terms of the assumptions in the setup of the simulation. If these assumptions are severe ones, i.e. the model is very far from the target phenomena, the explanation it suggests in terms of the modelled process will not correspond to a real explanation in terms of observed processes. The chosen assumptions in an explanatory model are crucial to its purpose in contrast to the case of a predictive model – this is an example of how the purpose of a model might greatly influence its construction.

It does sometimes occur that a model made for one purpose can be adapted for another, but the results are often not of the best quality and it almost always takes more effort than one expects. In particular, using someone else's model is usually not very easy, especially if you are not able to ask the original programmer questions about it and/or the code is not very well documented.

## 4.3 Modelling Assumptions

Whilst the available evidence will directly inform some parts of a model design, other parts will not be so informed. In fact it is common that a large part of a simulation model is not supported by evidence. The second source for design decisions is the conceptions of the modeller, which may have come from ideas or structures that are available in the literature. However, this is *still* not sufficient to get a model working. In order to get a simulation model to run and produce results it will be necessary to add in all sorts of other details: these might include adding in a random process to "stand in"

[1] Of course a successfully predictive model raises the further question of *why* it is successful, which may motivate the development of further *explanatory* models, since a complete scientific understanding requires both prediction and explanation, but not necessarily from the same models (Cartwright 1983).

for an unknown decision or environmental factor, or even be a straight "kludge" because you don't know how else to program something. Even when evidence supports a part of the design there will necessarily be some interpretation of this evidence. Thus any model is dependent upon a whole raft of assumptions of different kinds.

If a simulation depends on many assumptions that are not relatable to the object or process it models, it is unlikely to be useful. However, just because a model has *some* assumptions in it this does not mean it should be disregarded. *Any* modelling is necessarily a simplification of reality, done within some context or other. Hence there will be the assumption that the details left out are not crucial to the aspect of the results deemed important, as well as those assumptions that are inherent in the specification of the context. This is true for any kind of modelling, not just social simulation. It is not sufficient to complain that a model has assumptions or does simplify, since modelling without this is impossible; one has to argue or show *why* the assumptions included will distort the results. (Equally, the author of a model should be able to justify the assumptions that have been made.)

What one can do is to try to make the assumptions as transparent, as clear and as explicit as possible. Thus future researchers will be better able to judge what the model depends upon and adapt the model if any of the assumptions turns out to be considered bad. The most obvious technique is to try to document and display the assumptions. This not only helps to defend the model against criticism but also helps one to think more clearly about the model.

Particularly in the early stages of constructing a model, it is common to make a number of "assumptions" about various processes that are involved. In a sense these are not strictly assumptions – they are just necessary simplifications made in order to get *something* running – but nevertheless are included here. The model builder might for example include a random term to substitute for an unknown process, or a particular value might be chosen for a constant without knowing if it is a suitable value. The modeller must carefully document such decisions *and be prepared to revisit them and adjust them as necessary*.

The next type of assumption to consider is that which is "forced" by the constraints of the programming system. This might be the simplification of a process due to computational power limitations, restrictions forced upon the modeller due to the data structures and/or algorithms available, or the desire to reuse another (sub-)model. Again, such decisions must be documented. While the modeller may feel that these decisions have been forced, their documentation can serve two purposes. Firstly, other modellers may have insights into the same programming system that will allow them to suggest alternate approaches. Secondly, modellers who wish to replicate the model using an alternate system may be able to better demonstrate the impact of these assumptions.

The third type of assumption to consider is the choice of relevant objects and processes. As mentioned previously, any modelling exercise is necessarily an abstraction, and one must leave out much of the detail of the real world. Of course it is impractical to document *every* detail that has been omitted, but the modeller should consider carefully which objects and processes may be relevant to the model, and document those that have been included and those that have been omitted. This documentation will then prove invaluable in the consolidation phase (see Sect. 4.6), when the modeller should explicitly test these assumptions.

The most difficult type of assumption to track and document is that which derives from the modeller's own personal biases. For example, the modeller may have an innate "understanding" of some social process that is used in the model without question. The modeller may also have been trained within a particular school that embraces a traditional set of assumptions. Such traditional assumptions may be so deeply ingrained that they are treated as fact rather than assumption, making them difficult to identify from within.

This final class of assumption may be difficult for the modeller to identify and document, but all others should be carefully documented. The documentation can then be used in the exploration and consolidation phases (see below), when the modeller checks these assumptions as much as possible, refining the model as necessary. The assumptions should also be clearly stated and justified when reporting the model and results.

## 4.4 Maintaining Control of the Model While Exploring

The second biggest problem in following the exploration and consolidation approach to model building (after that of giving in to the temptation to promote your results without consolidation) is that one loses control of the model while exploring, resulting in a tangle of bugs. Exploration is an essential step, testing the impact of the assumptions that have been made, but if not carefully managed can result in code that achieves nothing at all. Bugs can creep in, to an extent that fixing one merely reveals another, or the model can become so brittle that any further modifications are impossible to integrate, or the model becomes so flaky that it breaks in a totally unexpected manner. Although interactions between processes might be interesting and the point of exploration, too much unknown interaction can just make the model unusable. Thus it is important to keep as many aspects as possible under control as you explore, so you are left with something that you *can* consolidate!

The main technique for maintaining control of a model is doing *some* planning ahead and consolidation *as* you explore. This is a very natural way for a modeller to work – mixing stages of exploration and consolidation as they feel necessary and as matches their ambitions for the model. Each programmer will have a different balance between these styles of work. Some will consolidate immediately after each bit of development or exploration; some will do a lot of exploration, pushing the model to its limits and then reconstruct a large part of the model in a careful and planned way. Some will completely separate the two stages, doing some exploration, then completely rebuild their ideas in a formal planned way but now having a better idea (if they are correct) of: what they are aiming to achieve, what needs to go into the model (*and what not*), what is happening in the model as it runs, and which results they need to collect from it.

There is no absolute rule for how careful and planned one should be in developing a model, but roughly the more complex and ambitious it is the more careful one should be.

Whilst a “quick and dirty” implementation may be sufficient for a simple model, for others it is unlikely to get the desired results: it is too easy to lose understanding and control of the interactions between the various parts, and also the model loses the flexibility to be adapted as needed later on. At the other end of the spectrum, one can spend ages planning and checking a model, building in the maximum flexibility and modularity, only to find that the model does not give any useful results. This might be a valuable experience for the programmer, but does not produce interesting knowledge about the target phenomenon. This is the fundamental reason why exploration is so important: because one does not know which model to build before it is tried. This is particularly so for models that have emergent effects (like most of the ones discussed in this volume), and also for those where there is no benchmark (either formal or observed) against which to check them.

One important thing about the activity of modelling is that one has to be willing to throw a lot of model versions away. Exploratory modelling is an inherently difficult activity; most of the models built will either be the wrong structure or just not helpful with regard to the phenomena we seek to understand. Further, the modelling is constrained in many ways: in the time available for developing and checking them, in the amount of computational resources they require, in the evidence available to validate the model, in the necessary compromises that must be made when making a model, and in the need to understand (at least to some extent) the models we make. Thus the mark of a good modeller is that he or she throws away a *lot* of models and only publishes the results of a small fraction of those he or she builds. There is a temptation to laziness, to trying to ‘fix’ a model that is basically not right and thus save a lot of time, but in reality this often only wastes time. This relates to the *modelling spectacles* mentioned above: one becomes wedded to the structure one is developing and it takes a mental effort to start afresh. However, if it is to be effective, a corollary of an exploratory approach is being highly selective about what one accepts – junking a lot of models is an inevitable consequence of this.

Whatever balance you choose between exploration and consolidation, it is probably useful to always pause before implementing any key algorithm or structure in your model, thinking a little ahead to what might be the best way. This is an ingrained habit for experienced programmers but may take more effort for the beginner. The beginner may not *know* of different ways of approaching a particular bit of programming and so may need to do some research. This is why developing some knowledge of common algorithms and data structures is a good idea. There is a growing body of work on documenting programming ‘patterns’ – which seek to describe programming solutions at a slightly general level – which can be helpful, although none of these pattern catalogues have yet been written specifically with models of social complexity in mind (but see Grimm et al. 2005 for examples from ecology). Increasingly too researchers within this field are making their code, or at least descriptions of the algorithms used, available to wider audiences.

There are dangers of using someone else’s code or algorithm though. There is the danger of assuming that one understands an algorithm, relying on someone else’s description of it.[2] It is almost inconceivable that there will not be some unforeseen results of applying even a well-known algorithm in some contexts. When it comes to reusing code, the risk is even higher. Just as there are assumptions and simplifications in one's own code, so there will be in the code of others, and it is important to understand their implications. Parameters may need adjustment, or algorithms tweaking, in order for the code to work in a different context. Thus one needs to thoroughly understand at the very least the interface to the code, and perhaps also its details. In some cases the cost of doing this may well outweigh the benefits of reuse.

It is important to note that even though the approach presented here deviates from *formal* approaches to software development, this does not mean one should ignore the standard 'good practices' of computer programming. Indeed, due to the complexity of even the simplest models in this field, it is advisable to do *some* planning and design before coding. In particular, the following principles should be applied:

- Conceptualisation: any model will benefit greatly from developing a clear understanding of the model structure and processes before starting to program. This is often called a *conceptual model* and usually involves some diagramming technique. While computer scientists will tend to use UML for this purpose, any graphical notation that you are familiar with will do to sketch the main entities and their relationships on paper, such as mind maps or flow diagrams. Apart from helping a modeller to better understand what the model is about this will form a valuable part of the overall model documentation. See (Alam et al. 2010; appendix) for an example of using UML class and activity diagrams.
- Modularity: it is not always possible to cleanly separate different functions or parts of a model but where it is possible it is hugely advantageous to separate these into different modules, classes or functions. In this way the interactions with the other parts of your model are limited to what is necessary. It makes it much easier to test the module in isolation, facilitates diagnostics, and can make the code much simpler and easier to read.
- Clear structures/analogies: it is very difficult to understand what code does, and to keep in mind all the relevant details. A clear idea or analogy for each part of the simulation can help you keep track of the details as well as being a guide to programming decisions. Such analogies may already be suggested by the conceptions that the programmer (or others) have of the phenomena under study, but it is equally important not to *assume* that these are always right, even if this was your intention in programming the model.
- Clear benchmarks: if there is a set of reference data, evidence, theory or other model to which the simulation is supposed to adhere this can help in the development of a model, by allowing one to know when the programming has gone astray or is not working as intended. The clearest benchmark is a set of observed social phenomena, since each set of observations provides a new set of data for benchmarking.

[2] Of course this danger is also there for one's *own* programming: it is more likely, *but far from certain*, that you understand some code you have implemented or played with.

Similarly if a part of the model is supposed to extend another model then restricting the new model should produce the same outcomes as the original.[3]

- Self-documentation: if one is continuously programming a simulation that is not very complex then one *might* be able to recall what each chunk of code does. However when developing this type of simulation it is common to spend large chunks of time focusing on one area of a model before returning to another. After such a lapse, one will not necessarily remember the details of the revisited code, but making the code clear and self-documenting will facilitate it. This sort of documentation does not necessarily have to be full documentation, but could include: using sensible long variable and module names, adding small comments for particularly tricky parts of the code, keeping each module, class, function or method fairly simple with an obvious purpose, and having some system for structuring your code.
- Building in error checking: errors are inevitable in computer code. Even the best programmer can inadvertently introduce errors to his or her code. Some of these will be obvious but some might be subtle; difficult to isolate and time-consuming to eliminate. Detecting such errors as early as possible is thus very helpful and can save a lot of time. Including safeguards within your code that automatically detect as many of these errors as possible might seem an unnecessary overhead, but in the long run can be a huge benefit. Thus you might add extra code to check that all objects that *should* exist at a certain time do in fact exist, or that a message from one object to another is not empty, or that a variable that should only take values within a certain range *does* stay within this range. This is especially important in an exploratory development, where one might develop a section of code for a particular purpose, which then comes to be used for another purpose. In other words the computational context of a method or module has altered.

There are also many techniques that computer scientists may exhort you to use that are not necessarily useful, that may be more applicable to the development of software with more clearly defined goals. Thus do evaluate any such suggested techniques critically and with a large dose of common sense.

## 4.5 Understanding the Model

Understanding a model is so intertwined with controlling a model that it is difficult to cleanly separate the two. You cannot really control a complex model if you do not at least partially understand it. Conversely, you cannot deeply understand a model until you have experimented with it, which necessitates being able to control it to a considerable extent. However, since modelling complex social phenomena requires

[3] What the "same outcomes" here means depends on how close one can expect the restricted new model to adhere to the original, for example it might be the same but with different pseudo-random number generators.

(at least usually and probably always) complex models, *complete* understanding and/ or control is often unrealistic. Nevertheless, understanding your model as much as is practical is key to useful modelling. This is particularly true for exploratory modelling because it is the feedback between trying model variations and building an understanding of what these variations entail that makes this approach useful.

Understanding one's model is a struggle. The temptation is to try shallow approaches by only doing some quick graphs of a few global measures of output, hoping that this is sufficient to give a good picture of what is happening in a complex social simulation. Although information about the relationship of the set-up of a simulation and its global outcomes can be useful, this falls short of a full scientific understanding, which must explain *how* these are connected. If you have an idea of what the important features of your simulation model are, you might be able to design a measure that might be suitable for illustrating the nature of the processes in your model. However, a single number is a very thin indication of what is happening – this is OK if you *know* the measure is a good reflection of what is crucial in the simulation, but can tend to obscure the complexity if you are trying to *understand* what is happening.

To gain a deeper understanding, one *has* to look at the details of the interactions between the parts of the simulation as well as the broader picture. There are two main ways of doing this: case studies using detailed traces/records and complex visualisations.

A case study involves choosing a particular aspect of the simulation, say a particular individual, object or interaction; then following and understanding it, step-by-step, using a detailed trace of all the relevant events. Many programming environments provide tracing tools as an inbuilt feature, but not all social simulation toolkits have such a feature. In this latter case, the modeller needs to embed the tracing into the model, with statements that will log the relevant data to a file for later analysis. This "zooming in" into the detail is often very helpful in developing a good understanding of what is happening and is well worth while, *even* if you don't think you have any bugs in your code. However, in practice many people seek to avoid this mundane and slightly time-consuming task.

The second way to gain an understanding is to program a good dynamic visualisation of what is happening in the model. What exactly is "good" in this context depends heavily on the nature of the model: it should provide a meaningful view of the key aspects of the model as the simulation progresses. Many social simulation toolkits provide a range of visualisation tools to assist this programming, but the key is identifying the relevant processes and choosing the appropriate visualisation for them – a task that is not amenable to generic approaches. Thus you could have a 2D network display where each node is an individual, where the size, shape, colour, and direction of each node all indicate different aspects of its state, with connections drawn between nodes to indicate interactions, and so on. A good visualisation can take a while to design and program but it can crucially enhance the understanding of your simulation and in most cases is usable even when you change the simulation set-up. The chapter by Evans et al. (2013) in this volume discusses a range of visualisation techniques aimed at aiding the understanding of a simulation model.

## 4.6 The Consolidation Phase

The consolidation phase should occur after one has got a clear idea about what simulation one wants to run, a good idea of what one wants to show with it, and a hypothesis about what is happening. It is in this stage that one stops exploring and puts the model design and results on a more reliable footing. It is likely that even if one has followed a careful and formal approach to model building some consolidation will still be needed, but it is particularly crucial if one has developed the simulation model using an informal, exploratory approach. The consolidation phase includes processes of: *simplification*, *checking*, *output collection* and *documentation*. Although the consolidation phase has been isolated here, it is not unusual to include some of these processes in earlier stages of development, intermingling exploration and consolidation. In such circumstances, it is essential that a final consolidation pass is undertaken, to ensure that the model is truly robust.

*Simplification* is where one decides which features/aspects of the model you need for the particular paper/demonstration you have in mind. In the most basic case, this may just be a decision as to which features to ignore and keep fixed as the other features are varied. However this is not very helpful to others because (a) it makes the code and simulation results harder to understand (the essence of the demonstration is cluttered with excess detail) and (b) it means your model is more vulnerable to being shown to be brittle (there may be a hidden reliance on some of the settings for the key results). A better approach is to actually remove the features that have been explored but turned out to be unimportant, so that only what is important and necessary is left. This not only results in a simpler model for presentation, but is also a stronger test of whether or not the removed features were irrelevant.

The *checking* stage is where one ensures that the code does in fact correspond to the original intention when programming it and that it contains no hidden bug or artefact. This involves checking that the model produces “reasonable” outputs for both “standard” inputs and “extreme” inputs (and of course identifying what “standard” and “extreme” inputs and “reasonable” outputs are). Commonly this involves a series of parameter sweeps, stepping the value of each parameter in turn to cover as wide a combination as possible (limited usually by resources). When possible, the outputs of these sweeps should be compared against a standard, whether that is real world data on the target phenomenon, or data from a comparable (well-validated) model.

The *output collection* stage is where data from the various runs is collected and summarised in such a way that (a) the desired results are highlighted and (b) sufficient “raw” data is still available to understand how these results have been achieved. It would be impractical to record the details of every variable for every run of the simulation, but presenting results in summary form alone may hide essential details. At the very least, it is essential to record the initial parameter settings (including random seeds, if random numbers are used) so that the summary results may be regenerated. It may also be informative to record at least a small number of detailed traces that are illustrative of the simulation process (once one has determined which parameter configurations produce “interesting” results).

*Documentation* is the last stage to be mentioned here, but is something that should develop throughout the exploration and consolidation of a model. Firstly, as mentioned above, the code should be reasonably self-documenting (through sensible naming and clear formatting) to facilitate the modeller's own understanding. Secondly, the consolidated model should be more formally documented. This should include any assumptions (with brief justifications), descriptions of the main data structures and algorithms, and if third-party algorithms or code have been used, a note to their source. This may seem like unnecessary effort, particularly if the modeller has no intention of publicly releasing the code, but if questions arise some months or years down the track, such documentation can be invaluable, even for the original author's understanding.

Finally, the modeller must present the model and its results to a wider audience. This is essential to the process of producing a model, since one can only have some confidence that it has been implemented correctly when it has been replicated, examined and/or compared to other simulations by the community of modellers. The distribution of the model should include a description of the model *with sufficient detail* that a reader could re-implement it if desired. It should present the typical dynamics of the system, with example output and summaries of detailed output. The relevant parameters should be highlighted, contrasting those deemed essential to the results with those with little or no impact. The benchmark measurements should be summarised and presented. To maximise a simulation's use in the community the simulation should be appropriately licensed to allow others to analyse, replicate and experiment with it (Polhill and Edmonds 2007).

## 4.7 Tools to Aid Model Development

As indicated previously, there is now a variety of systems for aiding the development of complex simulations. These range from programming-language-based tracing and debugging tools through to frameworks designed explicitly for social simulation, which include libraries of widely used patterns. Learning to use a particular system or framework is a substantial investment and because of this, most people do not swap from system to system readily once they have mastered one (even when an alternate system may provide a far more elegant solution to a problem). Ideally, a modeller would evaluate a range of systems when embarking on a new project, and decide upon the most appropriate one for that project. In practice, most modellers simply continue to use the same system as they have used on previous projects, without considering alternatives. There is no simple answer as to which system is "best". The available options are constantly changing as new systems are developed and old ones stop being supported. The type of modelling problem will influence the decision. And indeed it is partly a personal decision, depending on the modeller's own personal style and preferences. However given that such an investment is involved in learning a new system, it is a good idea to make this investment in one that will have large payoffs; that will be useful for developing a wide range of models.

Systems for developing and testing simulations range from the very specific to those that claim to be fairly generally applicable. At the specific end there are simulators that are designed with a restricted target in mind – such as a grid-based simulation of land use change (e.g. FEARLUS,[4] Polhill et al. 2001, or SLUDGE, Parker and Meretzky 2004) – where most of the structures, algorithms and outputs are already built in. The user has some latitude to adapt the simulation for their own modelling ends, but the ease with which one can make small changes and quickly get some results may be at the cost of being stuck with inbuilt modelling assumptions, which may not be appropriate for the task at hand. The specificity of the model means that it is not easy to adapt the system beyond a certain point; it is not a *universal* system, capable in principle, of being adapted to any modelling goal. Thus such a specific modelling framework allows ease of use at the cost of a lack of flexibility.

At the other end of the spectrum are systems that aim to be general systems to support simulation work; that can, at least in principle, allow you to build any simulation that can be conceived. Such systems will usually be close to a computer programming language, and usually include a host of libraries and facilities for the modeller to use. The difficulty with this type of system is that it can take considerable effort to learn to use it. The range of features, tools and libraries that they provide take time to learn and understand, as does learning the best ways to combine these features. Furthermore, even if a system in principle makes it possible to implement a modelling goal, different systems have different strengths and weaknesses, making any particular system better for some types of models, and less good for others. Thus modellers will sometimes "fight the system," implementing workarounds so that their model *can* be implemented within the system in which they have invested so much time, when in fact the model could more efficiently be implemented in an alternative system.

Between these two extremes lie a host of intermediate systems. Because they are often open source, and indeed more specific modelling frameworks are commonly built *within* one of these generic systems, it is usually possible (given enough time and skill) to 'dig down' to the underlying system and change most aspects of these systems. However the fundamental trade-offs remain – the more of a simulation that is 'given,' the more difficult it will be to adapt and the more likely it is that assumptions that are not fully understood will affect results.

Thus it is impossible to simply dictate which the best system is to use for developing simulation models of social complexity; indeed there is no single system that is best under all circumstances. However the sorts of questions one should consider are clearer. They include:

- *Clear structure*: Is the way the system is structured clear and consistent? Are there clear analogies that help 'navigate' your way through the various choices you need to make? Is it clear how its structures can be combined to achieve more complex goals?
- *Documentation*: Is there a good description of the system? Is there a tutorial to lead you through learning its features? Are there good reference documents

[4] http://www.macaulay.ac.uk/fearlus/.

where you can look up individual features? Are there lots of well-documented examples you can learn from?

- *Adaptability*: Can the system be adapted to your needs without undue difficulty? Is the way it is structured helpful to what you want to do? Are the structures easily adaptable once implemented in your model? Does the system facilitate the modularisation of your model so that you can change one aspect without having to change it all?
- *Speed*: How long does it take to run a model? Speed of execution is particularly important when a variety of scenarios or parameters need to be explored, or when several runs are necessary per parameter configuration due to random processes in the model.
- *User community*: Do many people in your field use the system? Are there active mailing lists or discussion boards where you can ask for help? If you publish a model in that system is it likely that it will be accessible to others?
- *Debugging facilities*: Does the system provide inbuilt facilities for debugging and tracing your simulation? If not, are there perhaps generic tools that could be used for the purpose? Or would you have to debug/trace your model by manually inserting statements into your code?
- *Visualisation facilities*: Does the system provide tools and libraries to visualise and organise your results? Are there dynamic visualisation tools (allowing one to view the dynamics of the system as it evolves)? How quickly can you develop a module to visualise the key outputs of a simulation?
- *Batch processing facilities*: Is there a means of running the model a number of times, collecting and perhaps collating the results? Is it possible to automatically explore a range of parameters whilst doing this?
- *Data collection facilities*: Are the results collected and stored systematically so that previous runs can easily be retrieved? Is it possible to store them in formats suitable for input into other packages (for example for statistical analysis, or network analysis)?
- *Portability*: Is the system restricted to a particular platform or does it require special software to run? Even if all your development will be done on one particular machine, in the interests of reusability it is desirable to use a system that will run on multiple platforms, and that is not dependent on specialised commercial software.
- *Programming paradigm*: Different programming paradigms are more appropriate to different types of modelling problems. If for example you think of things in terms of "if-then" statements, a rule-based system might be the most appropriate for your modelling. If instead you visualise things as series of (perhaps branching) steps, a procedural one might be more appropriate. In practice most systems these days are not *purely* one paradigm or another, but they still have leanings one way or another, and this will influence the way you think about your modelling.
- *Timing*: How will time be handled in the simulation? Will it be continuous or stepped, or perhaps event-driven? Will all agents act "at once" (in practice, unless each agent is run on a separate processor they will be executed in some sense sequentially, even if conceptually within the model they are concurrent),

or do they strictly take turns? Will it be necessary to run the simulation in real time, or (many times) faster than real time?

Once one has considered these questions, and decided on the answers for the particular model in mind, the list of potential systems will be considerably shortened, and one should then be able to make an informed choice over the available options. The temptation, particularly when one is beginning to write models, is to go for the option that will produce the quickest results, but it is important to remember that sometimes a small initial investment can yield long-term benefits.

## 4.8 Conclusion

It is easy to try and rationalise bad practice. Thus it is tempting to try and prove that some of the more formal techniques of computer science are not applicable to building social simulations just because one cannot be bothered to learn and master them. It *is* true however that not all the techniques suggested by computer scientists are useful in an exploratory context, where one does not know in advance precisely what one wants a simulation to do. In these circumstances one has to take a looser and less reliable approach, and follow it with consolidation once one has a more precise idea of what one wants of the simulation. The basic technique is to mix bits of a more careful approach in with the experimentation in order to keep sufficient control. This has to be weighed against the time that this may take given one does not know which final direction the simulation will take. There is a danger of this approach: that the modeller will be tempted by apparently significant results to rush to publication before sufficient consolidation has occurred. There may be times when the exploratory phase may result in useful and influential *personal* knowledge but such knowledge is not reliable enough to be up to the more exacting standards expected of *publicly* presented results. Thus it is only with careful consolidation of models that this informal approach to building simulations should be undertaken.

## Further Reading

Outside the social sciences, simulation has been an established methodology for decades. Thus there is a host of literature about model building in general. The biggest simulation conference, the annual "Winter Simulation Conference", always includes introductory tutorials, some of which may be of interest to social scientists. Good examples are (Law 2008) and (Shannon 1998).

For a comprehensive review of the currently existing general agent-based simulation toolkits see (Nikolai and Madey 2009); other reviews focus on a smaller selection of toolkits (e.g. Railsback et al. 2006; Tobias and Hofmann 2004; Gilbert and Bankes 2002).

The chapters on checking your simulation model (Galán et al. 2013), documenting your model (Grimm et al. 2013) and model validation (David 2013) in this volume should be of particular interest for anyone intending to follow the exploration and consolidation approach to model development. However, if you would rather attempt a more formal approach to building an agent-based simulation model, the subsequent chapter (Jonker and Treur 2013) discusses one such approach in detail. You could also consult textbooks on methodologies for the design of multi-agent systems, such as (Luck et al. 2004), (Bergenti et al. 2004) or (Henderson-Sellers and Giorgini 2005). After all, any agent-based simulation model can be seen as a special version of a multi-agent system.

## References

Alam SJ, Geller A, Meyer R, Werth B (2010) Modelling contextualized reasoning in complex societies with "endorsements". J Artif Soc Soc Simul 13(4). http://jasss.soc.surrey.ac.uk/13/4/6.html

Bergenti F, Gleizes M-P, Zambonelli F (eds) (2004) Methodologies and software engineering for agent systems: the agent-oriented software engineering handbook. Kluwer, Boston

Cartwright N (1983) How the laws of physics lie. Clarendon, Oxford

David N (2013) Validating simulations. Chapter 8 in this volume

Epstein J (2008) Why model? J Artif Soc Soc Simul 11(4). http://jasss.soc.surrey.ac.uk/11/4/12.html

Evans A, Heppenstall A, Birkin M (2013) Understanding simulation results. Chapter 9 in this volume

Galán J et al (2013) Detecting and avoiding errors and artefacts. Chapter 6 in this volume

Gilbert N, Bankes S (2002) Platforms and methods for agent-based modelling. Proc Natl Acad Sci USA 99(3):7197–7198

Grimm V et al (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310:987–991

Grimm V, Polhill G, Touza J (2013) Documenting simulations using ODD. Chapter 7 in this volume

Henderson-Sellers B, Giorgini P (eds) (2005) Agent-oriented methodologies. Idea Group, Hershey

Jonker C, Treur J (2013) A formal approach to building compositional agent-based simulations. Chapter 5 in this volume

Kuhn T (1969) The structure of scientific revolutions. University of Chicago Press, Chicago

Law AM (2008) How to build valid and credible simulation models. In: Mason SJ et al. (eds) Proceedings of the 2008 winter simulation conference, Miami, FL. http://www.informs-sim.org/wsc08papers/007.pdf

Luck M, Ashri R, d'Inverno M (2004) Agent-based software development. Artech House, London

Nikolai C, Madey G (2009) Tools of the trade: a survey of various agent based modelling platforms. J Artif Soc Soc Simul 12(2). http://jasss.soc.surrey.ac.uk/12/2/2.html

Parker DC, Meretsky V (2004) Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. Agric Ecosyst Environ 101(2–3):233–250

Polhill JG, Edmonds B (2007) Open access for social simulation. J Artif Soc Soc Simul 10(3). http://jasss.soc.surrey.ac.uk/10/3/10.html

Polhill G, Gotts N, Law ANR (2001) Imitative versus non-imitative strategies in a land-use simulation. Cybern Syst 32(1–2):285–307

Railsback SF, Lytinen SL, Jackson SK (2006) Agent-based simulation platforms: review and development recommendations. Simulation 82:609–623

Shannon RE (1998) Introduction to the art and science of simulation. In: Medeiros DJ, Watson EF, Carson JS, Manivannan MS (eds) Proceedings of the 1998 winter simulation conference, Washington, D.C. http://www.informs-sim.org/wsc98papers/001.PDF

Tobias R, Hofmann C (2004) Evaluation of free Java-libraries for social-scientific agent based simulation. J Artif Soc Soc Simul 7(1). http://jasss.soc.surrey.ac.uk/7/1/6.html

# Chapter 5
# A Formal Approach to Building Compositional Agent-Based Simulations

**Catholijn M. Jonker and Jan Treur**

**Why Read This Chapter?** To be introduced to a more formal "computer-science" style of simulation design, especially suited to simulations of multi-level systems (e.g. firms, departments, and people).

**Abstract** This chapter is an introduction to a more formal approach to designing agent-based simulations of organisations (in the widest sense). The basic method is the iterative refinement of structure, process and knowledge, decomposing each abstraction into near-decomposable components that can be (for the most part) then considered separately. Within this over all framework there are two complementary approaches: designing the organisation first, and designing the individual agents first.

## 5.1 Introduction

This chapter outlines a more formal approach to designing an agent system, in this case an agent-based simulation. Its approach comes from computer science, and shows how one can develop a design for a simulation model in a staged and cautious manner. This is particularly appropriate when the simulation is complex, requiring the interaction of complex and intelligent entities, and the intended design of the model is essentially known or accessible. Thus it contrasts with and complements the previous chapter describing more informal and exploratory approaches to developing simulations (Norling et al. 2013).

C.M. Jonker (✉)
Man-Machine Interaction, Delft University of Technology, Mekelweg 4, Delft 2628 CD, The Netherlands
e-mail: C.M.Jonker@tudelft.nl

J. Treur
Department of Artificial Intelligence, Vrije Universiteit Amsterdam, De Boelelaan 1081a, Amsterdam 1081 HV, The Netherlands
e-mail: treur@few.vu.nl

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_5,

This chapter draws on approaches which were designed for engineering agent-based software systems (e.g. air traffic control) and applies them to the engineering of agent-based simulations since a simulation model *is* an example of a complex piece of software. Such a cautious approach will result in the development of the model taking more time and effort but also should result in a simulation model that is easier to maintain and adapt and create fewer bugs, or simulations artefacts (as described in Chap. 6, Galán et al. 2013).

The chapter is structured into three main sections, covering: compositional design, organisations and agents. The section on compositional design is the most general, and "works" for both the design of organisations and agents. However, there is no need to reinvent the wheel – the sections on designing organisations and agents raise more specific issues and aspects that experience has shown to be helpful when designing simulations of these entities. Although they do not recapitulate all in the section on compositional design they are essentially examples of that general approach, though they each only concentrate on part of the overall process. The differences are, roughly, that the organisational view starts with the organisation, then the roles within that organisation, working "downwards". It highlights the constraints from above that organisations place on their members, dealing with the nature of the agents after. The agent viewpoint starts with the agents and their properties first and then moves on to how they interact (including maybe as an organisation). Social phenomena often *do* involve both the "downward" actions of constraint and "immergence" as well as the upwards actions of emergence and collective outcomes. Thus both organisational and agent views are often necessary to be explicitly considered in simulations.

## 5.2 Principles of Compositional Design of Multi-agent Systems

Although any principled design method can be used for the development of simulations, we are concentrating in this chapter on those social systems which are compositional in nature, since compositional actors are common in the social world and this case also covers the principles in the simpler, non-compositional, case. The principles described are quite general in nature, but will be illustrated with respect to a particular compositional multi-agent development method, the *Design and Specification of Interacting Reasoning Components (DESIRE)* (Brazier et al. 1998).

The approach described here considers the design of autonomous interactive agents and explicitly models both the *intra-agent functionality* and the *inter-agent functionality*. Intra-agent functionality concerns the expertise required to perform the tasks for which an agent is responsible in terms of the knowledge, and reasoning and acting capabilities. The inter-agent functionality concerns the expertise required to perform and guide co-ordination, co-operation and other forms of social interaction in terms of knowledge, reasoning and acting capabilities.

In the approach here described, both the individual agents and the overall system are considered as compositional structures – hence all functionality is designed in terms of interacting, compositionally structured components. Complex distributed processes are the result of tasks performed by agents in interaction with their environment.

The process starts at the most general, aggregate and abstract level, works in ever increasing detail down through the specification of individual agents and finally the code that determines their behaviour. At each stage there are a number of decisions to be made, which correspond to specifying the features and properties that are described below. The decisions at the higher levels help frame those made for the next level, etc. until everything has been described. This should leave a documented "trace" of the decisions that are made during the simulation design, that will help with the description of the simulation and the explicit logging of assumptions made during the design process.

Once this has been done and a roughly working simulation obtained, the verification and validation procedure happens in the other direction, starting at testing and validating the smallest processes and components and building upwards to higher groups and units until the simulation as a whole has been checked.

### *5.2.1 The Design Process*

The design of a multi-agent simulation is an iterative process, which aims at the identification of the parties involved (i.e., human agents, system agents, external worlds), and the processes, in addition to the types of knowledge needed. Initially broad conceptual descriptions of specific processes and knowledge are attained. Further explication of these conceptual design descriptions results in more detailed design descriptions, most often in parallel with the development of the conceptual design. During the design of these models, prototype implementations or parts or sections of the overall model may be used to analyse or verify the resulting behaviour. On the basis of examination of these partial prototypes, new designs and prototypes are generated and examined, and so on and so forth. This *evolutionary development* of systems is characteristic to the whole approach. Thus, as with any idealised design methodology, in practice there is a lot of iterating back and forth between stages and levels until a satisfactory design is obtained. The concepts presented in this chapter are to help focus what might otherwise be an unstructured process.

We distinguish the following kinds of descriptions within the development process:

- Problem description – Sect. 5.2.3
- Conceptual design – Sect. 5.2.4
- Detailed design – Sect. 5.2.4
- Design rationale – Sect. 5.2.5
- Operational design

The *problem description* is a description of the target system to be simulated and the main issues and questions to be investigated, this usually includes the *requirements* imposed on the design – what the simulation must do to be useful in this regard. Starting from the problem description, the *design rationale* specifies the choices made during each of the levels of the design process, the reasons for those choices, and the assumptions behind those choices that will impinge on its use. In other words, the *design rationale* is the strategy for "solving" the *problem description* using the design along with the reasons and justification for that strategy.

The actual design process roughly proceeds from conceptual and abstract, down to the more concrete until one has almost written the simulation code itself. The *conceptual design* includes conceptual models (the main design ideas and structures) for each entity in the model: the organisation, its roles and groups, the individual agents, the external world, the interaction between agents, and the interaction between agents and the external world. In a sense a conceptual design could apply as much to a human who will have to fulfil a role as a programmed agent, it does not concern itself with exactly *how* these aspects are to be achieved, but more about *how* it relates to other key structures. The *detailed design* of a system, based on the conceptual design, specifies all aspects of a system's knowledge and behaviour. It describes how the agent's processes will achieve its role. This can be thought of as the step where one is thinking about how a computational agent might achieve what is described in the conceptual design, but it is probably independent of which computer language, or system the agent is destined to be implemented in. A detailed design is an adequate basis for the *operational design*, which deals with some of the nitty-gritty of a specific implementation in a particular system. It stops short of actual programming code, but would be enough for a programmer to implement the system. This final stage will not be discussed in this chapter since it will be different for each programming language or system that is used to implement the final simulation.

The sequence tends to progress roughly from the conceptual towards the concrete, however this is only a general rule; there is no immutable sequence of design: depending on the specific situation, different types of knowledge are available at different points during system design which means that stages need to be iterated or even that, at times, lower level necessities might drive the higher levels of design.

### 5.2.2 Compositionality of Processes and Knowledge

Compositionality is a general principle that refers to the use of components to structure a design. This process of composition can be extended downwards as far as it is useful, for example, components can themselves be compositional structures in which a number of other, more specific components are grouped. During the design all the components at the different levels are identified. Processes at each of these levels (except the lowest level) are modelled as (process) *components* composed of entities at the level one lower to the process. Clearly this approach depends upon the possibility of decomposing what is being modelled into separate entities that are somewhat independent, so that the interaction of these components can be considered in turn. In other words it is necessary that what is being modelled is a near-decomposable system (Simon 1962). Although this is not always the case, there are many social actors that, prima facie, *are* themselves composed of other actors, for example political parties or firms. Thus this is often a reasonable approach to take in the field of social simulation. Even when it is not clear that what is being modelled does divide so neatly, then this method provides a systematic approach to attempting to identify and analyse those parts that are amenable to design so that when they are all put together the desired behaviour will *emerge* from their interaction during the simulation. Such emergence is an indication of non-decomposability and can never be guaranteed – one has to find out by running the simulation model, i.e. performing simulation experiments. If the desired behaviour does not emerge then one has to try and work out the possible reasons for the failure and go back to earlier stages of the design and rethink the approach.

In this compositional approach, each process within a multi-agent system may be viewed as the result of interaction between more specific processes. A complete multi-agent system may, for example, be seen to be one single component responsible for the performance of the overall process. Within this one single component a number of agent components within a common internal environment may be distinguished, each responsible for a more specific process. Each agent component may, in turn, have a number of internal components responsible for more specific parts of this process. These components may themselves be composed, again entailing interaction between other more specific processes.

The knowledge and information that is stored, produced, and communicated is as important as the processes. The set of all terms, labels, types, structures etc. that is used to encode and process the knowledge needed in a specific domain or for the purposes of a particular agent may also be seen as a component, a *knowledge structure*. This knowledge structure can be composed of a number of more specific knowledge structures which, in turn, may again be composed of other even more specific knowledge structures.

*Compositionality of processes* and *compositionality of knowledge* are two independent dimensions of design. Thus a set of processes might be summarised as a single process when appropriate or, in the other direction, broken down into a system of sub-processes. The knowledge structures at one level might be adequate for the purposes of the compositional processes or, maybe, a finer system of description might be needed to be utilised by a finer grain of process representation. For example, some simulations might represent the spread of beliefs through a group as a simple contagion process, with simple entities representing the beliefs and who has them. A finer grained model might include more of a cognitive representation of the knowledge structures involved and how others are persuaded to accept these as the result of a dialogue process.

Compositionality is a means to achieve *information and process hiding* within a model: by defining processes and knowledge at different levels of abstraction, unnecessary detail can be hidden at those stages, allowing the broader considerations to be considered separately from the component details. Clearly in the realm of social simulation being able to satisfactorily express social processes without always going down to the details is necessary if the resulting model is to be feasible – clearly we cannot simulate social actors, going all the way down to the atoms they are made of. Compositionality also makes it possible to *integrate* different types of components in one agent, providing the structure and means by which they work together.

### *5.2.3 Problem Description*

There are many ways to write a *problem description*. Techniques vary in their applicability, depending on, for example, the situation, the task, or the type of knowledge on which the system developer wishes to focus. Therefore, no particular method will be described here. However, whichever way the problem description is developed it is crucial to capture the key requirements to be imposed on the system – that is, what one *wants to gain* from building and using the simulation. These requirements are part of the initial problem definition, but may also evolve during the development of a system. Different simulations of the *same phenomena* might well be appropriate because each might have different requirements. For example, a simulation model to predict where queues will form on a certain stretch of motorway will probably be different from one to predict whether different proportions of lorries might affect the throughput of traffic, even if both simulations are of traffic on the same stretch of road at the same times.

### *5.2.4 Conceptual and Detailed Design*

A conceptual and detailed design consists of specifications of the following three types:

- Process composition;
- Knowledge composition;
- The relation between process composition and knowledge composition.

These are discussed in turn in more detail below.

#### 5.2.4.1 Process Composition

Process composition identifies the relevant processes at different levels of (process) abstraction, and describes how a process can be defined in terms of lower level processes. Depending on the context in which a system is to be designed two different approaches can be taken: a *task perspective*, or a *multi-agent perspective*.

The task perspective refers to the approach, in which the processes needed to perform an overall task are distinguished first, which are then *delegated* to appropriate agents and the external world. In this approach the agents and the external world are designed later. The *multi-agent perspective* refers to the approach in which agents and an external world are distinguished first and afterwards the processes within each agent and within the external world.

### Identification of Processes at Different Levels of Abstraction

Processes can be described at different levels of abstraction; for example, the processes for the multi-agent system as a whole, processes within individual agents and the external world, processes within task-related components of individual agents. Thus in a traffic simulation system processes might include the introduction and removal of vehicles, the collection of statistics and the visualisations of the system state; individual agents representing vehicles might have processes for monitoring their speed and for deciding when to change lane; within these agents might be a reactive component that monitors and adjusts the speed reacting when other traffic gets too close, a learning component that remembers which lanes were faster in the past, and a reasoning component that decides when to change lanes.

### Relevant Aspects of a Process

The processes identified are modelled as *components*. For each process the *types of information* used by it for input and resulting as output are identified and modelled as *input and output interfaces* of the component (an interface is a protocol for the information and maybe some specification of a process to deal with it, translating or storing it). So in a traffic simulation the process in an agent may need the distance and the relative speed of any object in its path to be passed to it as an input and the reactions (accelerating, braking) may be passed as its outputs. Clearly in a simple simulation these interfaces will be trivial, but in more complex simulations or where a degree of modularity is required some effort in designing these interfaces might well pay off.

### Modelling Process Abstraction Levels

Each level of process is either an *abstraction* of lower levels of component and/or a specialisation of the levels above. These layers of abstraction only go down so far since processes are either *composed* of other components or they may be *primitive*. Primitive components may be either reasoning components (for example based on a knowledge base), or, alternatively, components capable of performing tasks such as calculation, information retrieval, optimisation, etc.

The identification of processes at different levels of abstraction results in the specification of components that can be used as building blocks, and which components are sub-components of which other component. The distinction of different levels of process abstraction results in hiding detail from the processes at the higher levels. Thus a process to decide whether to change lane in the traffic example might be composed of a process to access the memory of how fast each lane was in the past, an estimate of the average speed of the current lane, and how fast the traffic ahead is moving.

#### Composition

The way in which processes at one level of abstraction in a system are composed of processes at the adjacent lower abstraction level in the same system is called *composition*. This composition of processes is described not only by the component/sub-component relations, but in addition by the (possibilities for) *information exchange* between processes (the *static* aspects), and *task control knowledge* used to control processes and information exchange (the *dynamic* of the composition).

#### Information Exchange

A specification of information exchange defines which types of information can be transferred between components and the ways by which this can be achieved, called *information links*. Within each of the components *private* information links are defined to transfer information from one component to another. In addition, *mediating* links are defined to transfer information from the input interfaces of encompassing components to the input interfaces of the internal components, and to transfer information from the output interfaces of the internal components to the output interface of the encompassing components. That is the mediating links are those which pass information "up" and "down" the structure of the agent: to their components or up to the entity that they are a component of. Thus in the traffic example there might well be mediating information links from each vehicle up to the simulation to pass information about its current speed and position.

### 5.2.4.2 Knowledge Composition

Knowledge composition identifies the knowledge structures at different levels of abstraction, and describes how a knowledge structure can be defined in terms of lower level knowledge structures. The levels of knowledge abstraction may correspond to the levels of process abstraction, but this is not necessarily the case.

Identification of Knowledge Structures at Different Abstraction Levels

The two main structures used as building blocks to model knowledge are: *information types* and *knowledge bases*. These knowledge structures can be identified and described at different levels of abstraction. At the higher levels the details can be hidden. The resulting levels of knowledge abstraction can be distinguished for both information types and knowledge bases.

*Information Types*

An information type defines the sorts of terms that will be used describe objects or other terms, their kinds, and the relations or functions that can be defined on these objects.[1] Information types can be specified in graphical form, or in formal textual form. Thus the *speed* of objects is a type of knowledge in the traffic example, relatable to other speeds in terms of *relative speed*.

*Knowledge Bases*

Knowledge bases are structured collections of information held by agents. The specification of the knowledge bases use the information types just described. To specify a knowledge base one needs to say which information types are used in as well as the relationships between the concepts specified in the information types. Thus in a (somewhat complex) driver memory there might be three kinds of information: days of the week, times of the day, lane label and categories of speed. Each lane might relate to a set of past occasions composed of day of the week, time of day and speed category.

Composition of Knowledge Structures

Information types can be composed of more specific information types, following the principle of compositionality discussed above. Similarly, knowledge bases can be composed of more specific knowledge bases. Thus in the example of memory about past lane speeds the sets of past occasions might be a list of limited size ordered first by level of annoyance and secondly by recency.

#### 5.2.4.3 Relation Between Process Composition and Knowledge Composition

Each process in a process composition uses knowledge structures. These will be involved in the building and maintenance of the knowledge structures at their level, but could involve knowledge structures from higher or, occasionally, lower levels.

[1] Such sets of agreed terms are often called an "ontology" in computer science.

So the processes that comprise the cognitive processes of a traffic agent might well be involved in maintaining the memory of past lane speeds but might also relate to the positional clues that are associated with the highest levels of the simulation.

### *5.2.5 Design Rationale*

The *design rationale* is important because it makes explicit the reasoning "glue" that underpins some of the other parts. Essentially it answers the question "*given the modelling target and goals why have the design decisions been made?*" Thus it describes the relevant properties of the design in relation to the requirements identified in the problem description. It also documents the verification of the design – that is how one will check that the implemented system does *in fact* meet its specification, including the assumptions under which the desired properties will hold. All the important design decisions are made explicit, together with some of the alternative choices that could have been made, and the arguments in favour of and against the different options. At the operational level the design rationale includes decisions based on operational considerations, such as the choice to implement an agent's cognitive process in a particular way in order to make the simulation run at a reasonable speed.

### *5.2.6 Multi-agent Systems in the Simulation of Social Phenomena*

The method described above deals with the design process in terms of components and the interactions between those components. In this light, multi-agent systems are not considered specifically. However, in the context of simulating social phenomena, it comes out naturally that in many instances the appropriate "components" are the correlates of observed social actors. In other words it is almost always overwhelmingly sensible to model the target system as a multi-agent system, where agents in the model are representations of the actors (people, firms, etc.) that are known to exist. In a sense, in the social sphere almost everything is an agent, or conversely, agents are nothing special. It is simply that a component that is naturally thought of as having elements of cognition (learning, reasoning etc.) is an agent and will be endowed, as part of the simulation process, with many of the attributes that agents are expected to have (and are discussed later in this chapter). Representations of humans in a simulation will not include all aspects of cognition but, dependent on the modelling goals, might well be much simpler. On the other hand some non-human components, such as a firm, might be represented as an agent, being able to learn, react, and reason in an agent-like way.

Simulations are growing in complexity, not in the least because agents are asked to fulfil different roles over time, and to change their behaviour according to both their own internal learning mechanisms and changing role descriptions. Within the described approach it has become good practice to first design the organisation, and then the agents and their interaction in such a way that the agents realize the organisation. The next section explicitly considers organisations. The chapter on "Assessing Organisational Design" (Dignum 2013) follows the application of the ideas that are described here.

## 5.3 Organisations

The organisational approach to simulation design takes the observed and inferred organisational structures as the starting point and considers individual action and agency at a later stage. This is particularly suitable for situations that *seem* to be structured in this way, that is to say the roles and the requirements significantly characterise and constrain individual action. Clearly in many observed cases there is a complex mix of organisational constraint and emergence from individual action so the decision to adopt a primarily organisational approach is a pragmatic one. In many cases a mixture of organisation-based and agent-based approaches will be necessary.

Societies are characterised by complex dynamics involving interaction between many actors and groups of actors. If such complex dynamics take place in an completely unstructured, incoherent manner, then the actors involved will probably not be able to predict much, and not able to use and exploit any knowledge that they have in a useful way. However in many social situations this is not the case, social phenomena are full of structure, and even in initially unstructured situations social actors will often quickly develop norms, rules, habits etc. – effectively creating structure. Some sociologists (e.g. Luhman) have suggested that the *purpose* of human institutional structure is to manage the complexity, in other words to simplify social action and make planning possible. Organisational structure provides co-ordination of the processes in such a manner that the agents involved can function in a more adequate manner. The dynamics in many organisational structures are much more dependable and understood than in apparently entirely unstructured situations.

One key assumption of the organisational approach to simulation design is that the organisational structure itself is relatively stable, i.e., the structure may change, but the frequency and scale of change are assumed low compared to the more standard dynamics through the structure. Within the field of Organisation Theory such organisational structures regulating societal dynamics are studied (see e.g. Kreitner et al. 2001; Mintzberg 1979). In summary, organisational *structure* is used to help specify the *dynamics* (or organisational *behaviour*) of a desired type. A crucial issue for further analysis is how exactly structure is able to *affect* dynamics.

A number of organisation modelling approaches have been developed to simulate and analyse dynamics within organisations in society (e.g. Ferber and Gutknecht 1998; Hannoun et al. 1998, 2000; Hübner et al. 2002a b; Lomi and Larsen 2001; Moss et al. 1998; Prietula et al. 1997). Some of these approaches explicitly focus on modelling organisational structure, abstracting from the detailed dynamics. Other approaches put less emphasis on organisational structure but focus on the dynamics in the sense of implementing and experimenting with simulation models. Often these simulation models are based on some implementation environment and not specified in an implementation-independent manner using a formally defined conceptual language. However, there are some exceptions to this where the specification approach is supported by an implementation framework.[2] The Agent/Group/Role (AGR) approach (previously called Aalaadin) introduced in (Ferber and Gutknecht 1998) is a good example of the organisational approach. It focusses on organisational structure, abstracting from the details of the dynamics. It helps define a formal relation between the dynamic properties and the organisational structure (Ferber et al. 1999, 2000). The relevance for this chapter is that it shows how dynamics of the organisational structure itself can be modelled: agents can dynamically create, join, or quit groups. This is particularly relevant for simulating situations where the organisational structure is somewhat fluid.

In this section the "dynamics specification" approach exemplified by AGR is presented.[3] The organisational structure is discussed and its parts defined. Then the dynamics of the organisation is discussed in terms of dynamic properties that can be associated to each element of the organisational structure. These dynamic properties can help the simulation and analysis of empirical or simulated traces. The various compositional levels within an organisation are related to the organisational dynamics via a series of relationships. Finally, as a prerequisite to realising an organisation the requirements of the agents are specified from their roles within the organisation model.

### *5.3.1 Specification of Organisation Structure*

In this approach, an organisation is viewed as a framework for activity and interaction through the definition of groups, roles and their relationships. By avoiding an agent-oriented viewpoint, an organisation is regarded as a structural relationship between agents. In this way the organisation is described solely on the basis of its structure, i.e. by the way groups and roles are arranged to form a whole, without being concerned with the way agents actually behave. That is the systems will be analysed from the outside, as a set of interaction modes. The specific architecture of the agents is purposely not addressed in the organisational model.

---

[2] The Strictly Declarative Modelling Language SDML (Moss et al. 1998) and the use of the agent-oriented modelling approach DESIRE in social simulation as presented in (Brazier et al. 2001) are two examples.

[3] For more information on the use of AGR, see (Jonker and Treur 2003).

The three primitive definitions are:

- The *agents*. The model places no constraints on the internal architecture of agents. An agent is only specified as an active communicating entity which plays roles within groups. This agent definition is intentionally general to allow agent designers to adopt the most accurate definition of agent-hood relative to their application. In other words, the specification of the agent is left as flexible as possible, given the organisational constraints upon its roles.
- *Groups* are sets of agents. Each agent is part of one or more groups. In its most basic form, the group is only a way to tag a set of agents. An agent can be a member of *several* groups at the same time. A major point of these groups is that they can freely overlap.
- A *role* is an abstract representation of an agent function, service or identification within a group. Each agent can have multiple roles and each role handled by an agent is local to a group. Roles could be assigned beliefs; that is, they could reason about whether they should have a particular belief given a certain role. These beliefs can be seen as an additional requirement on the agents playing that role.

Organisation structure is often show as a diagram (for example, as kind of labelled graph; see Fig. 5.3 in Sect. 5.3.5) consisting of roles, groups, and interactions, and of relationships between these elements.

Within AGR an organisation structure consists of a set of groups, the roles in each group and the agents fulfilling those roles. To complete the picture relationships between roles can be specified.

### 5.3.2 Organisation Structure

An AGR specification of an organisation structure is defined by the following: groups, roles, (intergroup) interactions, transfers (intra-group interactions), which roles are in which groups, the roles that are the source of interactions, the roles that are the destination of interactions, the roles that are the source of transfers, and the roles that are the destination of transfers. Transfers, under this scheme, are within a group as opposed to interactions which may be between groups. Thus it is necessary that the source and destination of all transfers belong to the same group. Although intergroup interactions are defined above as between two roles, this can easily be generalised to intergroup interactions involving more than two roles.

### 5.3.3 Dynamic Properties of an Organisation

After the foundation of an organisation structure has been defined, the foundations for specification of dynamic properties in an organisation are addressed. The aim is not only to cover simple types of dynamics, such as simple reactive behaviour, but also more complex dynamics, necessary for the simulation of realistic organisations. The challenge here is to incorporate somehow the organisational structure within the formal description of the organisation's internal dynamics. To this aim, the following approach is introduced:

For each element within the organisational structure characterise its dynamics by a specific set of dynamic properties.

This is based on the structural relations between elements in an organisational structure. Then:

Identify relationships between the sets of dynamic properties corresponding with these elements;

In general, the dynamics of an element within an organisation structure can be characterised by describing how the states of the elements change over time. For a role the 'state' needs to include descriptions of for both the input and the output of the role. Transfers and intergroup interactions are assumed to operate only on input and output states of roles. These roles do not have their own internal state, so no further state is needed to be described for such transfers and intergroup interactions.

An organisational structure defines relations between different elements in an organisation. The dynamics of these different elements are characterised by their dynamic properties. An organisational structure has the aim of keeping the overall dynamics of the organisation manageable. For this reason the structural relations between the different elements within the organisational structure have to somehow impose constraints on or dependencies between their dynamics. Logical relations defined between sets of dynamic properties allow the use of logical methods to analyse, verify and validate organisation dynamics in relation to organisation structure. Within AGR organisation models three aggregation levels are involved:

- The organisation as a whole (the highest aggregation level)
- The level of a group
- The level of a role within a group

A general pattern for the dynamics in the organisation as a whole in relation to the dynamics in groups is as follows:

*Dynamic properties for the groups* AND *dynamic properties for intergroup role interaction*
⇨ *Dynamic properties for the organisation*

Moreover, dynamic properties of groups can be related to dynamic properties of roles as follows:

*Dynamic properties for roles* AND *dynamic properties for transfer between roles*
⇨ *Dynamic properties for a group*

The idea is that these are properties dynamically relating a number of roles within one group.

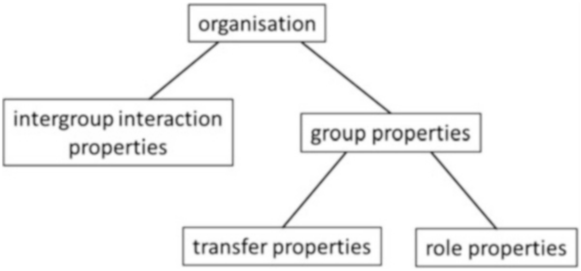



**Fig. 5.1** Overview of inter-level relations between dynamic properties within an AGR organisation model

**Table 5.1** Types of dynamic properties for an AGR organisation model

| Property type | Relating | |
|---|---|---|
| Role r | Role r input | Role r output |
| Transfer from r1 to r2 | Role r1 output | Role r2 input |
| Group G | Input or output of roles in G | |
| Intragroup interaction | Role r1 output | Role r2 output |
| Intergroup interaction | Role r1 input | Role r2 output |
| Organisation | Input or output of roles in O | |

An overview of the logical relationships between dynamic properties at different aggregation levels is depicted as an AND-tree in Fig. 5.1.[4]

To define states the notion of state property is useful. The notion of *trace* as a sequence of states over a time frame is used to formalise the dynamics. To formally specify dynamic properties that are essential in an organisation, an expressive language is needed. One can do this using a formal language,[5] however in this chapter this will not be used to retain its accessibility.

We distinguish five kinds of dynamic properties that might be described during a specification (or at least thought about). Not all of these are always necessary. A summary of them is displayed in Table 5.1.

#### 5.3.3.1 Role Dynamic Properties

The *role dynamic properties* relate input to output of that role. This is a subset of the dynamic properties of that role; it is a concern of that role only. For example, the gossip role behaviour: 'whenever somebody tells you something, you will tell it to everybody else' is expressed in terms of input of the role leading to output of the role in a reactive manner.

[4] For formalisation details of the logical relationships put forward above, see (Jonker and Treur 2003).

[5] E.g. the Temporal Trace Language (TTL), which defines the dynamics in terms of a "leads to" relation (Jonker et al. 2001). A specification of dynamic properties in *leads to* format has as advantages that it is executable and that it can often easily be depicted graphically.

#### 5.3.3.2 Transfer Dynamic Properties

Transfer properties relate output of the source roles to input of the destination roles. That is they represent the dynamic properties of transfers from one role to another.

Typically, these sets contain properties such as: information is indeed transferred from source to destination, transfer is brought about within *x* time, arrival comes later than departure, and information departs before other information also arrives before that other information.

#### 5.3.3.3 Group Dynamic Properties

Group dynamic properties relate input and/or output of roles within a group, it relates the roles within the group. An example of a group property is: "if the manager asks anyone within the group to provide the secretary with information, then the secretary will receive this information".

A special case of a group property is an *intragroup interaction* relating the outputs of two roles within a group. A typical (informal) example of such an intragroup interaction property is: "if the manager says 'good afternoon', then the secretary will reply with 'good afternoon' as well". Other examples may involve statistical information, such as "3 out of the 4 employees within the organisation never miss a committed deadline".

#### 5.3.3.4 Intergroup Interaction Dynamic Properties

Intergroup interaction properties relate the input of the source role in one group to the output of the destination role in another group. Note that intergroup interaction is specified by the interaction of roles within the group, and not the groups themselves. Sometimes there are specialist roles for such intergroup interaction. For example, a project leader is asked by one of the project team members (input of role 'project leader' within the project group) to put forward a proposal in the meeting of project leaders (output of role 'member' within the project leaders group).

#### 5.3.3.5 Organisation Dynamic Properties

Organisation dynamic properties relate to input and/or output of roles within the organisation. A typical (informal) example of such a property is: "if within the organisation, role A promises to deliver a product, then role B will deliver this product".

The different types of dynamic properties all relate to different combinations of input and output. Table 5.1 provides an overview of these combinations. Note that with respect to simulation, the above dynamics definition can contain elements that are redundant: a smaller subset of dynamical properties could form an executable specification of the dynamics of an AGR type organisation – *not all of the above are always needed*.

For example, an organisation could be simulated on the basis of the role dynamic properties, the transfer dynamic properties and the intergroup interactions. The group dynamic properties, including the intragroup role interaction properties, and the organisation properties should emerge in the execution. However specifying them in advance can be used to check what is expected and help verify the simulation.

In order to make an executable organisation model the dynamical properties need to be chosen from those properties that can be executed.

### *5.3.4 Organisation Realisation*

In this section criteria are discussed when allocation of a set of agents to roles is appropriate to realize the organisation dynamics, illustrated for the AGR approach. One of the advantages of an organisation model is that it abstracts from the specific agents fulfilling the roles. This means that all dynamic properties of the organisation remain the same, independent of the particular allocated agents. However, the behaviours of these agents have to fulfil the dynamic properties of the roles and their interactions that have been already specified. The organisation model can be (re)used for any allocation of agents to roles for which:

- For each role, the allocated agent's behaviour satisfies the dynamic role properties,
- For each intergroup role interaction, one agent is allocated to both roles and its behaviour satisfies the intergroup role interaction properties, and
- The communication between agents satisfies the respective transfer properties.

To satisfy the relationships specified above there needs to be a relevant overlap between the agent's ontologies and the role ontologies,[6] i.e. there must be some common referents so that the interactions of the agents are well defined with respect to their roles. Moreover, note that if one agent performs two roles in the group then dynamic properties of communication from itself to itself are required, i.e. that it will receive (at its input state) what it communicates (at its output state): 'it hears itself talking'. The logical relationships can be depicted as in the extension of Fig. 5.1 shown as Fig. 5.2.

Alternatively, if the roles in an intergroup interaction would not be fulfilled by one agent, but by several, this would create a mystery, since input to one agent creates output for another agent, even though the agents are not connected by any

[6] For a more detailed discussion on this issue, see (Sichman and Conte 1998).

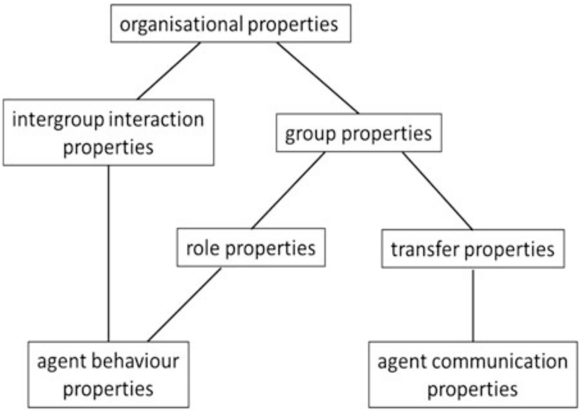



**Fig. 5.2** Inter-level relations between dynamic properties for a realised organisation model

transfer since the roles they fulfil are from separate groups. This would suggest that the organisation structure is not complete. The whole idea of specifying the organisational approach through roles is that all communication and interaction is somehow made explicit – in an AGR organisation model it is assumed that the roles in an intergroup interaction are fulfilled by one agent.

### 5.3.5 *Organisational Example*

Here the organisational approach to simulation specification is illustrated. This shows how an organisation that is the target of simulation can be analysed into groups, roles and processes. This analysis can then be the basis for the design of the simulation implementation and finally its code. As described this is essentially a top-down analytic approach (in contrast to the more bottom-up agent approach described at the end of this chapter).

In this example, a factory and some of its components are considered. This factory is organised at the highest level according to two divisions: the division that produces certain components (division A) and the division that assembles these components into products (division B). At the lower level, division A is organised in two departments: the work planning department for division A (dep. A1) and the component production department (dep. A2). Similarly, division B is organised in two department roles: one for assembly work planning (dep. B1) and one for product production (dep. B2). This example is illustrated in Fig. 5.3.

Here the two divisions are modelled as groups (depicted by the larger ovals), with the departments as their roles (depicted by smaller ovals within larger ones). A third group, the Connection Group C, models the communication between the two divisions. This group consists of the two roles ‘division A representative’ and ‘division B representative’. Intergroup role interactions (depicted by pairs of dashed lines) are modelled between the role ‘department A1’ in the division A

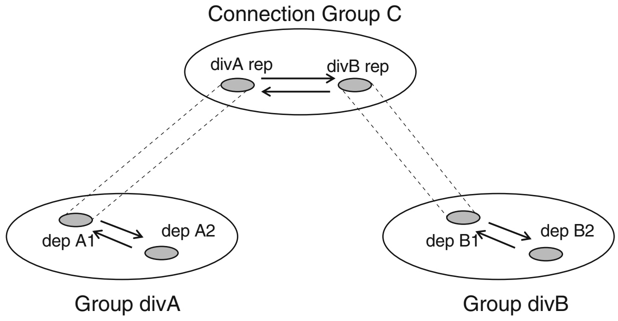



**Fig. 5.3** Organisational example. The *smaller ovals* indicate roles, the *bigger ovals* groups. Connections are indicated by the two types of *lines* (*dashed* indicates an intergroup interaction, *solid arrow* indicates a transfer). Membership of a role to a group is indicated by drawing the *smaller role oval* within the *bigger group oval*

group and the role 'division A representative' within the connection group, and between the role 'department B1' in the division B group and the role 'division B representative' within the connection group. Intragroup role transfers model communication between the two roles within each of the groups (depicted by the arrows).

Connections have destination roles (indicated by the arrow points) and source roles (where the arrow originates). Based on the semantic structures of many-sorted predicate logic a more precise formal definition is the following.

#### 5.3.5.1 Groups and Roles in Organisational Example

The example has the following groups, roles, and relationships between them:

- Groups = {divA, divB, C},
- Roles = {depA1, depA2, depB1, depB2, divArep, divBrep},
- Intergroup_interactions = {iAC, iCA, iBC, iCB}
- Transfers = {tA12, tA21, tB12, tB21},
- Some of the relationships are:

| Within divA | Organisation level |
|---|---|
| Role_in(depA1, divA) | Source_of_interaction(divA, iAC) |
| Role_in(depA2, divA) | Destination_of_interaction(C, iAC) |
| Source_of_transfer(depA1, tA12) | Source_of_interaction(C, iCA) |
| Destination_of_transfer(depA2, tA12) | Destination_of_interaction(divA, iCA) |
| Source_of_transfer(depA2, tA21) | |
| Destination_of_transfer(depA1, tA21) | |

#### 5.3.5.2 Dynamic Properties in Organisational Example

To get the idea, consider the special case of an intragroup role interaction from role r1 to role r2, characterised by dynamic properties that relate output of one role r1 to output of another role r2. Assuming that transfer from output of r1 to input of r2 is adequate and simply copies the information, this property mainly depends on the dynamics of the role r2. Therefore in this case the relationship has the form:

*Dynamic properties for role r2 AND*
*Dynamic properties for transfer from role r1 to role r2*
⇨ *Dynamic properties of intragroup interaction from r1 to r2*

Role Dynamic Properties

*DP(depA1) Progress Information Generates Planning in depA1*

If within division A department A1 receives progress information on component production,then an updated planning will be generated by department A1 taking this most recent information into account.

Group Dynamic Properties

*DP(A) A Progress Information Generation*

This property is the conjunction of the following two properties.

*DP1(A) Initial A Progress Information Generation*

Department A1 receives initial progress information on component production processes, involving already available components.

*DP2(A) Subsequent A Progress Information Generation*

Within the division A group, for any component production planning generated by department A1, incorporating a specific required set of components, progress information on the production of these components will be received by department A1.

Intergroup Interaction Dynamic Properties

*Intergroup Role Interaction between A and C: IrRI(A, C)*

For the connectivity between the groups A and C, the following intergroup role interaction properties are considered, one from A to C, and one from C to A.

*IrRI(depA1, divArep) Progress Information Provision A to B*

If within division A progress information on component production is received by department A1, then within the connection group this will be communicated by the division A representative to the division B representative.

*IrRI(divArep, depA1)* *B Progress Information Incorporation by A*

If within the connection group the division A representative receives information from the division B representative on which components are needed, then within division A a component production planning will be generated by department A1 taking these into account.

Organisational Dynamic Properties

*DP(F)* *Overall Progress Notification*

If a request for a product is made (by a client), then progress information will be provided (for the client).

Realisation

The following allocation of agents agentA1, agentA2, agentB1, agentB2 to roles is possible:

| | | |
|---|---|---|
| *agentA1 – depA1* | *agentB1 – depB1* | *agentA1 – divArep* |
| *agentA2 – depA2* | *agentB2 – depB2* | *agentB1 – divBrep* |

To realise the organisation model, for example agentA1 has to satisfy the following dynamic properties:

*DP(agentA1)*
*If* *agent A1 receives progress information on component production,*
*Then* *an updated planning will be generated by agent A1 taking this most recent information into account.*
*IrRI(agentA1)*
*If* *progress information on component production is received by agent A1,*
*Then* *this will be communicated by agent A1 to agent B1*
*If* *agent A1 receives information on which components are needed,*
*Then* *a component production planning will be generated by agent A1 taking these components into account*

#### 5.3.5.3 Conclusion of Organisational Example

One can see how the above analysis is getting us closer to the implementation of a simulation of this organisation. Given details of an organisation this could continue down the organisational structure. Clearly this kind of analysis is more appropriate when the structure of the organisation is known, and much less appropriate when the structure is only partially known or, indeed, emergent. However, even in those cases it could guide the documentation capturing how and which aspects of an organisation's official structure was translated into a simulation.

## 5.4 Organisation Design by Requirements Refinement

The previous sections address the question of how the structure and the behaviour of a given organisation can be modelled. This section takes the design perspective. This perspective does not assume a given organisation, but aims at creating a new organisation *in silica*. Whilst on the whole in simulation we are aiming to capture aspects of existing organisations one might want to design one in a number of circumstances.

For example, one may not know the internal structure of an organisation which is part of the system one is trying to model but only how it communicates or relates to the actors around it. In this case to get the simulation to run one would have to invent the organisation. Obviously the danger in this case is that the organisational structure that is chosen might subtly affect how it interacts with other agents and thus have an impact upon the simulation outcomes. However in some cases the internal workings of an organisation *are* effectively insulated from how it interacts with the outside world by regulation and self-interest – in these cases one might well have no choice but to invent its workings on the basis of its external constraints and common knowledge of how such things are arranged.

Another case is where one is not attempting to represent anything that is observed but rather exploring the space of possible organisations. For example one might wish to know which of several possible organisational structures might be best according to some specified criteria. Such "artificial societies" or "thought experiments" are reasonable common, however their relevance is questionable. If a small change in the environment or other aspect (e.g. reliability of internal communication) means that what was a good organisational structure now fails, then the results of such artificial studies are difficult to apply to other cases. In other words the general applicability of such studies is hard to establish. On the other hand if one has a good knowledge of the environment and characteristics where the results of the simulation experiments are to be applied and one does extensive 'what if' experiments testing the robustness of the designs to such small changes then this can be a helpful way forward.

Such an design process starts by specifying requirements for the overall organisation behaviour. The requirements express the dynamic properties that should 'emerge' if appropriate organisational building blocks, such as roles, groups, transfers, group interactions, role behaviours, and so on, are glued together in an appropriate manner in an organisation model. In addition, other requirements on behavioural and structural aspects of the organisation to be created may be imposed. Given these requirements on overall organisation behaviour (and, perhaps, some additional requirements), organisational structure and organisational behaviour are designed in such a manner that the requirements are fulfilled. The approach described in this section is the method of requirements refinement, which is illustrated for an example.

### 5.4.1 Designing by Requirements Refinement

In Sect. 5.3.3 a scheme for specifying the dynamic properties and relationships at different levels of aggregation was described; overall organisational behaviour can be related to dynamic group properties and group interaction properties via the following pattern:

*Dynamic properties for the groups AND dynamic properties for group interaction ⇨ Dynamic properties for the organisation*

This scheme is also useful for the design perspective. Consider a design problem for which the requirements of the overall behaviour are given in the form of dynamic properties. This scheme says that to fulfil these overall dynamic properties, dynamic properties of certain groups and group interactions together imply the organisation behaviour requirements. This process is called *requirements refinement* in that the requirements for the whole organisation are reduced to that of its constituent groups and the interactions between these groups. It thus provides a new, refined set of requirements in terms of the behaviour of groups and group interaction.

Clearly if one has a multi-level organisation with component sub-organisations as well as groups one has a choice as to how *best* to fill in the detail of one's design. One can decide to reduce it first to the behaviour of its constituent groups but it is also possible to first refine requirements for the behaviour of the organisation as a whole to the requirements on the behaviour of parts of the organisation, before further refinement is made to refinements for groups. In each case this is a pragmatic decision and will depend on the organisation being designed.

Subsequently, the required dynamic properties of groups can be refined to dynamic properties of certain roles and transfers, making use of:

*Dynamic properties for roles AND dynamic properties for transfer between roles ⇒ Dynamic properties for a group*

This provides requirements on role behaviour and transfer that together imply the requirements on the behaviour of the group. Again it is possible to first refine requirements on the behaviour of a group to requirements of the behaviour of parts of the group, before further refinement to role behaviour requirements is made, depending on what is best in each case.

An overview of the inter-level relationships between these dynamic properties at different aggregation levels is depicted in Fig. 5.1, repeated for your convenience

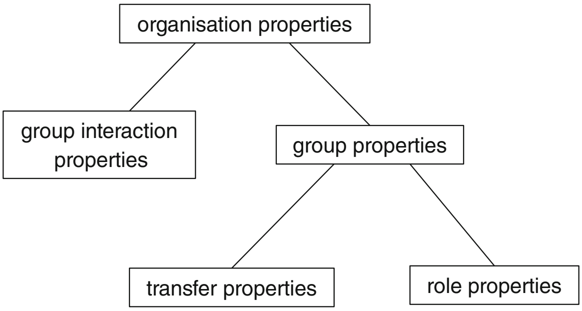



**Fig. 5.4** Overview of interlevel relationships between dynamic properties within an organisation model

here as Fig. 5.4. In summary, from the design perspective, a *top-down refinement approach* can be followed. That is, the requirements on overall organisational behaviour can be first refined to requirements on behaviour of groups and group interaction, and then the requirements on behaviour of groups can be refined to requirements on roles and transfers. Notice that as part of this refinement process the organisational structure (e.g., the groups and roles) is defined.

A design problem statement consists of:

- A set of requirements (in the form of dynamic properties) that the overall organisational behaviour has to fulfil
- A partial description of (prescribed) organisational structure that has to be incorporated
- A partial description of (prescribed) dynamic properties of parts of the organisation that have to be incorporated; e.g., for roles, for transfers, for groups, for group interactions.

A *solution specification* for a design problem is a specification of an organisation model (both structure and behaviour) that fulfils the imposed requirements on overall organisation behaviour, and includes the given (prescribed) descriptions of organisation structure and behaviour. Here 'fulfilling' the organisation behaviour requirements means that the dynamic properties for roles, transfers, and group interactions within the organisation model imply the behaviour requirements.

In specific circumstances, part of the organisational structure and/or behaviour may already be prescribed by requirements. For example, the organisational structure may already be prescribed; in such a case only the organisation dynamics is designed, for the given organisational structure. Other, more specific cases are, for example, *role behaviour design* and *interaction protocol design*.

#### 5.4.1.1 Role Behaviour Design

For role behaviour design the organisational structure and the transfers and interactions are completely prescribed. However, appropriate dynamic properties for the different roles have yet to be found, to satisfy the requirements for the

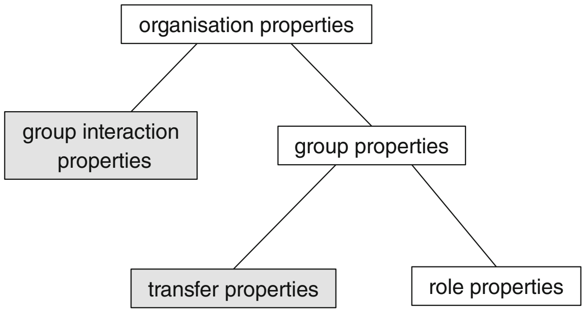



**Fig. 5.5** Role behaviour design

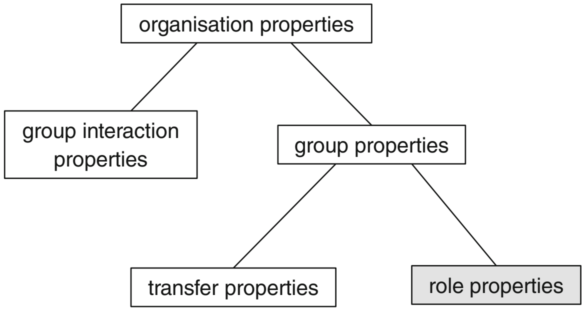



**Fig. 5.6** Organisation design

organisational behaviour that are imposed; see Fig. 5.5. Here (and in Fig. 5.6) the grey rectangles indicate what is already given as prescribed and the transparent rectangle what has to be designed.

#### 5.4.1.2 Interaction Protocol Design

For interaction protocol design the organisational structure and role dynamics are completely prescribed, but appropriate transfer and interaction dynamics have to be found to satisfy given requirements for the organisational behaviour that are imposed; see Fig. 5.6.

## 5.5 The Agent Approach

The agent approach is contrary to the organisational approach. It starts with the agents and its properties and attempts to work upwards towards the whole system. This is more useful in situations where the "top down" social constraints upon action are weak or non-existent, and it is the "upwards" emergence of outcome from the many micro-level interactions that is more important. Clearly, if one *was* in a situation where top down social and/or organisational constraints were severe then one would have no guarantee that working bottom-up in this manner one would be able to meet those constraints at the higher levels of the structure. It would be like trying to organise the production of a kind of car by the random meeting of people with particular parts and skills without any planning. However, especially in the *development of new social structure* such "bottom-up" processes can be crucial, so that the agent approach can be appropriate for investigating such issues. Often, for social phenomena some mix of both approaches is necessary, first a bit of one, then a bit of the other etc.

### 5.5.1 Some Agent Notions

The term agent has been used for a wide variety of applications, including: simple batch jobs, simple email filters, mobile applications, intelligent assistants, and large, open, complex, mission critical systems (such as systems for air traffic control).[7] Some of the key concepts concerning agents lack universally accepted definitions. In particular, there is only partial agreement on what an agent is. For example, simple batch jobs are termed agent because they can be scheduled in advance to perform tasks on a remote machine, mobile applications are termed agent because they can move themselves from computer to computer, and intelligent assistants are termed agents because they present themselves to human users as believable characters that manifest intentionality and other aspects of a mental state normally attributed only to humans. Besides this variety in different appearances of agents, the only precise description of the agents involved is their implementation code. As a result, existing agent architectures are only comparable in an informal manner – just because something is *called* an agent-architecture or an agent does not mean that it is suitable for simulating a human or social actor. Especially if the goal of the agent-based system is a complex simulation, a principled, design-oriented description of the organisation, and of the agents in it at a conceptual and logical level is of the essence, since the control, testing, verification and validation of such complex simulations is an issue. Spending time on a more formal and staged approach to simulation design can make life a lot easier later. Due to the organisational nature, and the complexity of intelligent agents and their interaction, a more formal compositional design method for agents is necessary.

As agents show a variety of appearances, perform a multitude of tasks, and their abilities vary significantly, attempts have been made to define what they have in common. The *weak notion of agent* is seen as a reference. The weak notion of agent is a notion that requires the behaviour of agents to exhibit at least the following four types of behaviour:

[7] Many of the notions discussed in this and the following section are adopted from (Wooldridge and Jennings 1995), (Nwana 1996), (Nwana and Ndumu 1998) and (Jennings and Wooldridge 1998).

- Autonomous behaviour
- Responsive behaviour (also called reactive behaviour)
- Pro-active behaviour
- Social behaviour

*Autonomy* relates to control: although an agent may interact with its environment, the processes performed by an agent are in full control of the agent itself. A*utonomous* behaviour is defined as:

> ... where the system is able to act without the direct intervention of humans (or other agents) and should have control over its own actions and internal state.

This means that an agent can only be requested to perform some action, and:

> The decision about whether to act upon the request lies with the recipient.

Examples of autonomous computer processes are: process control systems (e.g., thermostats, missile guiding systems, and nuclear reactor control systems), software daemons (e.g., one that monitors a user's incoming email and obtains their attention by displaying an icon when new, incoming email is detected), and operating systems.

Many processes that exhibit autonomous behaviour are called agents. However, if such agents do not exhibit flexible behaviour, they are not, in general, considered to be *intelligent* agents. An intelligent agent is a computer system that is capable of flexible autonomous actions in order to meet its design objectives – indeed Randall Beer (1990) defined intelligence as "the ability to display adaptive behaviour". Intelligence requires flexibility with respect to autonomous actions, meaning that intelligent agents also need to exhibit responsive, social, and pro-active behaviour.

An agent exhibits *responsive* (or *reactive*) behaviour if it reacts or responds to new information from its environment. Responsive behaviour is where:

> Agents perceive their environment (which may be the physical world, a user, a collection of agents, the Internet, etc.) and respond in a timely fashion to changes that occur in it.

A barometer is a simple example of a system that exhibits responsive behaviour: It continually receives new information about the current air pressure and responds to this new information by adjusting its dial.

*Pro-active* behaviour is where:

> Agents do not simply act in response to their environment, but are be able to exhibit opportunistic, goal-directed behaviour and take the initiative where appropriate.

Pro-active behaviour in some sense is the most difficult of the required types of behaviour for an agent defined according to the weak agent notion. For example, pro-active behaviour can occur simultaneously with responsive behaviour. It is possible to respond to incoming new information in an opportunistic manner according to some goals. Also initiatives can be taken in response to incoming new information from the environment, and thus this behaviour resembles responsive behaviour. However, it is also possible to behave pro-actively when no new information is received from the environment. This last behaviour can by no means be called responsive behaviour.

An agent exhibits *social* behaviour if it communicates and co-operates with other agents. Jennings and Wooldridge define social behaviour as when:

> Agents are able to interact, when they deem appropriate, with other artificial agents and humans in order to complete their own problem solving and to help others with their activities.

An example of an agent that exhibits social behaviour is a car: it communicates with its human user by way of its dials (outgoing communication dials: speed, amount of fuel, temperature) and its control mechanisms (incoming communication control mechanisms: pedals, the steering wheel, and the gears). It co-operates with its human user, e.g., by going in the direction indicated by the user, with the speed set by that user.

Agents can also be required to have additional characteristics. Here three of these characteristics are discussed: adaptivity, pro-creativity, and intentionality.

*Adaptivity* is a characteristic that is vital in some systems. An adaptive agent learns and improves with experience. This behaviour is vital in environments that change over time in ways that would make a non-adaptive agent obsolete or give it no chance of survival. This characteristic is modelled in simulations of societies of small agents, but also, for example, in adaptive user interface agents.

*Pro-creativity* is of similar importance to find agents that satisfy certain conditions. The chance of survival is often measured in terms of a fitness function. This characteristic is found in various simulations of societies of small agents (see the literature in the area of Artificial Life). A computer virus is an infamous form of a pro-creative agent.

An *intentional system* is defined by Dennett to be an entity

> ... whose behaviour can be predicted by the method of attributing beliefs, designs and rational acumen.

Mentalistic and intentional notions such as *beliefs*, *desires*, *intentions*, *commitments*, *goals*, *plans*, *preference*, *choice*, *awareness*, may be assigned to agents. The *stronger notion of agenthood,* in which agents are described in terms of this type of notions, provides additional metaphorical support for the design of agents.

### 5.5.2 Representative Agents

Of course a software agent need not have all the characteristics of something it *represents*. Thus, depending on the model purpose, it is quite possible to represent an intelligent actor by a relatively simply object, that might not even be meaningfully called an agent. For example, if simulating the movement of a crowd or traffic, it might not be necessary to include much, if any, of the features discussed above.

So sometimes something is called an agent because it *represents* an agent, but this usage conflates what is being modelled and the model, which may cause confusion. How and when it is necessary to include the various features that are known about the actor being modelled in the model is a crucial modelling question. However it is not a question to which there is a general answer since this goes to the heart of the difficulty of making sense of the complexity that we observe.

The power of the agent idea and approach to programming, clearly comes from the apparent efficacy of observed social actors that seem to be able to organise themselves in useful and adaptive ways, that they have the characteristics listed above. This differs qualitatively from more traditional computer science approaches to programming. It also comes from the power of the analogy with humans to guide the direction of programming – we have a deep mundane knowledge of how humans work (c.f. Dennett 1996) and this can help in design decisions, for example whether a certain feature or ability is necessary to model a particular social system. In this sense the idea of an agent can be thought of as a stance, in comparison to the "intentional" stance mentioned above – it may be useful to think of a computational object as an agent, having some of the sort of properties we know real human actors have. However there are clearly more and less useful applications of this stance: I may think of a light switch as an agent, but it has limited usefulness in terms of understanding or modelling it. In the contrary direction on can think of a fully autonomous and intelligent agent such as a human as a merely a physical particle for some circumstances, however this is open to question, depending upon the purpose and target of the exercise. Clearly, in many circumstances and for many purposes, treating complex social actors as if they were simple (for example acted upon in the manner of a simple linear influence plus some randomness) is insufficient since individual complexity can impinge upon the social complexity that results.

### *5.5.3 Agent Properties*

The notions of agency discussed above are highly abstract notions. In order to design agents, it is necessary to be familiar with a number of primitive agent concepts.[8] These primitive concepts serve as an ontology or vocabulary used to express analyses and designs of applications of agents and multi-agent systems. Two classes of primitive notions are distinguished: those used to describe the behaviour of agents in terms of their external (or public) states and interactions (Sect. 5.5.3.1), and those used to describe the behaviour of agents in terms of their internal (or private) states, and processes (Sect. 5.5.3.2). To illustrate these concepts, some example agents are discussed in Sect. 5.5.4.

[8] The material in this section is based on (Brazier et al. 2000).

#### 5.5.3.1 External Primitive Concepts

Two types of interaction of an agent with its environment are distinguished, depending on whether the interaction takes place with an agent or with something else (called an *external world*), for example a database, or the material world. For each of these two types of interaction specific terminology is used.

Interaction with the External World

Two primitive types of interaction with the external world are distinguished. The first type of interaction, *observation*, changes the information the agent has about the world, but does not change the world state itself, whereas the second type, *performing an action*, does change the world state, but does not change the information the agent has about the world. Combinations of these primitive types of interaction are possible; for example, performing an action, and observing its results.

Observation

In which ways is the agent capable of observing or sensing its environment? Two types of observation can be distinguished: the agent passively receives the results of observations without taking any initiative or control to observe (*passive observation*), or the agent actively initiates and controls which observations it wants to perform; this enables the agent to focus its observations and limit the amount of information acquired (*active observation*).

Execution of Actions in the External World

An agent may be capable of making changes to the state of its environment by initiating and executing specific types of actions.

Communication with Other Agents

Two directions of communication are distinguished, which can occur together: *outgoing communication* (is the agent capable of communicating to another agent; to which ones?), and *incoming communication* (is the agent capable of receiving communication from another agent; from which ones?).

#### 5.5.3.2 Internal Primitive Concepts

A description in terms of the external primitive concepts abstracts from what is inside the agent. In addition to descriptions of agents in terms of the external concepts, descriptions in terms of internal concepts are useful. The following internal primitive agent concepts are distinguished.

World and Agent Models

An agent may create and maintain information on (a model of) external *world* based on its observations of that world, on information about that world communicated by other agents, and its own knowledge about the world. The agent may also create and maintain information on (models of) *other agents* in its environment based on its observations of these agents as they behave in the external world, on information about these agents communicated by other agents, and knowledge about the world.

Self Model and History

Some agents create and maintain information on (a model of) their own characteristics, internal state, and behaviour. Or the agent creates and maintains a history of the world model, or agent models, or self model, or own and group processes.

Goals and Plans

To obtain pro-active, goal-directed behaviour, an agent represents, generates, and uses explicit goals and its own plans of action in its processing.

Group Concepts

Besides individual concepts, agents can use group concepts that allow it to co-operate with other agents. For example, joint goals: is the agent capable of formulating or accepting and using goals for a group of agents, i.e., goals that can only be achieved by working together? Or joint plans: is the agent capable of representing, generating, and using plans of action for joint goals, i.e., involving which actions are to be performed by which agents in order to achieve a certain joint goal? Also commitments to joint goals and plan, negotiation protocols and strategies can be useful group concepts for agents, depending on their role and function.

**Table 5.2** External primitive concepts for an elevator

| External primitive concepts | Elevator |
|---|---|
| Interaction with the world | |
| Passive observations | Presence of objects between doors (optically) |
| | Total weight |
| | Its position |
| Active observations | Presence of objects between the doors (mechanically) |
| Performing actions | Moving |
| | Opening and closing doors |
| Communication with other agents | |
| Incoming communication | From users in the elevator |
| | Where they want to go (pushing button in elevator) |
| | From users outside |
| | Where they want to be picked up (pushing button outside elevator) |
| Outgoing communication | To users in the elevator |
| | Where we are (display) |
| | There is overweight (beep) |
| | To users outside |
| | Where is the elevator (display) |
| | In which direction it moves (display) |

## 5.5.4 *Example of the Agent Approach: An Elevator*

Let us illustrate the agent concepts introduced above by an example: an elevator is analysed from the agent perspective using these basic concepts (Table 5.2). This might be an element in the simulation of how people move around a building. The advantage of using an elevator as an example is that it does interact with users as an agent, but it is well known and simple enough to make a clear example of specifying an agent using the agent-oriented approach.

### 5.5.4.1 External Primitive Concepts (Table 5.2)

Observation

Observations are performed continually. However, it receives passive observation results on the presence of objects between the doors (an optical sensor), the total weight of its contents, and its position in the building (at which floor). Besides it is able to perform active observation: the presence of objects between the doors (a mechanical sensor which is moved in the door opening just ahead of the doors themselves).

**Table 5.3** Internal primitive concepts for an elevator

| Internal primitive concepts | Elevator |
|---|---|
| World model | The current floor, max load, current load |
| Agent models | A user wants to be picked up from floor X |
| | A user wants to go to floor Y |
| Self model | When maintenance is next due |
| History | When maintenance was last performed |
| Goals | To go to floor X to pick up somebody |
| | To go to floor Y to deliver somebody |
| Plans | The order in which the required floors are visited |
| | Sometimes: the speed that is taken |
| Group concepts | |
| Joint goals | With other elevators to transport people and goods as efficiently as possible |
| Joint plans | Some elevators are capable of distributing the work |
| Commitments | The elevators then commit to their part of the work |
| Negotiation protocol | To reach a good distribution, they may have to negotiate |
| Negotiation strategies | To reach a good distribution, they may have to negotiate |

Performing Actions

Its actions are moving itself (and people) vertically from one position to another and opening and closing doors.

Incoming Communication

The elevator receives communication from users by buttons that have been pressed outside (to call the lift and indicate the direction they wish to go in) and inside the lifts (providing information about the floor to which they wish to be transported).

Outgoing Communication

The elevator communicates to a user by indicating which floor the lift is on (both inside and outside the lifts) and sounding beeps (information about overload) (Table 5.2).

#### 5.5.4.2 Internal Primitive Concepts (Table 5.3)

World and Agent Models

Elevators know where they are, to do this they keep track of which floor they are on based on their actions (going two floors up, going one floor down) they perform.

**Table 5.4** Types of behaviour for an elevator

| Types of behaviour | Elevator |
|---|---|
| Autonomy | Yes |
| Responsiveness | In reaction to user requests |
| | In immediate reaction to observed objects between the doors |
| Pro-activeness | Taking the initiative to go to a normally busy floor, if empty and not being called by a user |
| Social behaviour | Co-operation with users, and, sometimes, with other elevators |
| Own adaptation and learning | Often not possible |

Furthermore, the elevator knows if the weight in the lift is over its maximum limit. The agent information of the user goals (where they want to go) may be maintained as well.

### Self Model and History

The agent does not know what actions it previously performed to perform its current task. It might have an explicit representation of when it has last received maintenance.

### Goals and Plans

Modern elevators make use of the explicit goals (adopted from the goals communicated by the users). The goals are used to determine which actions to perform. They may even make plans for reaching these goals: determine the order of actions, for example when one of the users has the goal to be at a higher floor and another on a lower floor.

### Group Concepts

The elevator co-operates with its users. The elevator might also be designed to co-operate with other elevators so that they could strategically distribute themselves over the floors. The goals adopted from the goals communicated by the users are *joint goals* (joint with the users), and sometimes even joint with the other elevators. Modern elevators are capable of distributing the work load, and thus of making *joint plans*. To achieve the joint goals an elevator must *commit* to its part of the work as specified in the joint plans. To make a joint plan, the elevators might negotiate using a particular *strategy* as to which elevator goes where. Negotiation is only possible if a *negotiation protocol* is followed.

#### 5.5.4.3 Types of Behaviour (Table 5.4)

Autonomy

As soon as it is activated, no system or human is controlling its machinery, and (normally) it is not switched off and on by the user. The elevator has full control of its motor, doors, and lights.

Pro-activeness

The simplest elevators stay where they are (some take the initiative to close their doors) when no longer in use, but more intelligent elevators go to a strategic floor (e.g., the ground floor).

Reactiveness

The elevator reacts to the immediate stimuli of buttons pressed, therefore, it is shows reactive behaviour. People often have to wait for the elevator as the elevator picks up people on other floors, however, the elevator does not forget a signal and will, eventually, come to the requested floor.

Social Behaviour

The elevator co-operates with users and, sometimes, with other elevators.

Own Adaptation and Learning

Simple elevators are not capable of adjusting their own behaviour to new situations, nor are they capable of learning. However, it is possible to conceive of more intelligent elevators that can learn the rush hours for the different floors.

#### 5.5.4.4 Conclusion of Elevator Example

One can see that the above analysis has clarified what is needed to implement a model of this element within a simulation. The objects, properties and processes of the simulation code is an easy step from here. One also sees that some of the complexities of the agent have been considered before the implementation starts, in this way cleaner and more maintainable code might be produced, fewer mistakes made in the implementation and some of the complexities in terms of user-lift interaction considered and anticipated. Of course in the case of simulating a human agent there are likely to be many unknowns in terms of their goals, group concepts etc. – unless the simulators simply add in their informed guesses they will have a considerable job finding evidence to guide them in the answers to fill in within such an analysis. Thus this kind of analysis is not a total solution when simulating complex social actors whose attributes and internal states may be largely unknown.

## 5.6 Conclusion

More formal approaches to simulation design can help make complex implementations manageable and can probably save one time in the longer-run. It also makes the simulation easier to check, validate, re-implement and further develop. These approaches do this in a three principled ways. *Firstly*, by encouraging the more complete documentation of the intentions and decisions of a designer/implementer. One can see a lot of this chapter as a check-list of all the aspects one might think about and record. *Secondly*, it helps encourage doing this in an explicit and exact manner. We have not displayed some of the formal notation that can be used in this chapter, since we did not want to overwhelm the reader, however for those who wish to utilise this style of design to the fullest will naturally find them themselves adopting a variety of formal languages and diagrams in pursuit of precision. *Thirdly*, it suggests a method by which a complex specification can be iteratively refined from abstract and high-level entities towards a detailed implementation. In this way the design decisions do not all have to be made simultaneously but can be made in stages.

## Further Reading

For readers interested in software engineering approaches Bergenti et al. (2004) give a thorough introduction to and overview of current methodologies. Gilbert and Terno (2000) offer suggestions on techniques for building and structuring agent-based simulation models, particularly geared towards use in the social sciences.

In addition to methodologies, a lot of work has been done in the development of programming languages and platforms to support the implementation of multi-agent systems and models. Bordini et al. (2010) focus on a comprehensive presentation of MAS programming, including four approaches that are based on formal methods, whereas Railsback et al. (2006) provide a review of platforms for agent-based simulations.

## References

Beer RD (1990) Intelligence as adaptive behavior: an experiment in computational neuroethology. Academic, Boston

Bergenti F, Gleizes M-P, Zambonelli F (eds) (2004) Methodologies and software engineering for agent systems: the agent-oriented software engineering handbook. Kluwer, Boston

Bordini RH, Dastani M, Seghrouchni AEF (eds) (2010) Multi-agent programming: languages, platforms and applications. Springer, Berlin

Brazier FMT, Jonker CM, Treur J (1998) Principles of compositional multi-agent system development. In: Cuena J (ed) IT & KNOWS – information technology and knowledge systems, IFIP world computer congress 1998 (Schriftenreihe der Österreichischen Computer Gesellschaft, 122). OCG, Wien, pp 347–360

Brazier FMT, Jonker CM, Treur J (2000) Compositional design and reuse of a generic agent model. Appl Artif Intell J 14:491–538

Brazier FMT, van Eck PAT, Treur J (2001) Modelling a society of simple agents: from conceptual specification to experimentation. J Appl Intell 14:161–178

Daniel C. Dennett (1996), *The Intentional Stance* (6th printing), Cambridge, Massachusetts: The MIT Press, ISBN 0-262-54053-3 (First published 1987)

Dignum V (2013) Assessing organisational design. Chapter 20 in this volume

Ferber J, Gutknecht O (1998) A meta-model for the analysis and design of organisations in multi-agent systems. In: Demazeau Y (ed) Proceedings of the third international conference on multi-agent systems (ICMAS'98), Paris, 3–7 July 1998. IEEE Computer Society Press, pp 128–135

Ferber J, Gutknecht O (1999) Operational semantics of a role-based agent architecture. In: Jennings NR, Lespérance Y (eds) Intelligent agents VI, proceedings of the 6th international workshop on agent theories, architectures and languages, ATAL'99 (Lecture notes in Computer Science), vol 1757. Springer, Berlin, pp 205–217

Ferber J, Gutknecht O, Jonker CM, Müller JP, Treur J (2000) Organization models and behavioural requirements specification for multi-agent systems. In: Proceedings of the fourth international conference on MultiAgent systems (ICMAS), Boston, 10–12 July 2000. IEEE Computer Society, pp 387–388

Galán J et al (2013) Checking simulations: detecting and avoiding errors and artefacts. Chapter 6 in this volume

Gilbert N, Terno P (2000) How to build and use agent-based models in social science. Mind Soc 1(1):57–72

Hannoun M, Sichman JS, Boissier O, Sayettat C (1998) Dependence relations between roles in a multi-agent system: towards the detection of inconsistencies in organization. In: Sichman JS, Conte R, Gilbert N (eds) Multi-agent systems and agent-based simulation, proceedings of the first international workshop on multi-agent based simulation, MABS'98. (Lecture notes in artificial intelligence), vol 1534. Springer, Berlin, pp 169–182

Hannoun M, Boissier O, Sichman JS, Sayettat C (2000) MOISE: an organizational model for multi-agent systems. In: Monard MC, Sichman JS (eds) Advances in artificial intelligence, international joint conference 7th Ibero-American conference on AI, 15th. Brazilian symposium on AI (IBERAMIA-SBIA 2000), Atibaia, 19–22 Nov 2000. Proceedings (Lecture notes in computer science), vol 1952. Springer, Berlin, pp 156–165

Hübner JF, Sichman JS, Boissier O (2002a) A model for the structural, functional and deontic specification of organizations in multiagent systems. In: Bittencourt G, Ramalho GL (eds) Advances in artificial intelligence. Proceedings of 16th Brazilian symposium on artificial intelligence (SBIA'02), Porto de Galinhas/Recife, Brazil, 11–14 Nov 2002. (Lecture notes in computer science), vol 2507. Springer, Berlin, pp 439–448

Hübner JF, Sichman JS, Boissier O (2002b) MOISE+: towards a structural, functional and deontic model for MAS organizations. In: Castelfranchi C, Johnson WL (eds) Proceedings of the first international joint conference on autonomous agents and multi-agent systems, AAMAS 2002, Bologna, 15–19 July 2002. ACM Press, pp 501–502

Jennings NR, Wooldridge M (eds) (1998) Agent technology: foundations, applications, and markets. Springer, Berlin

Jonker CM, Treur J (2003) Relating structure and dynamics in organisation models. In: Sichman JS, Bousquet F, Davidsson P (eds) Multi-agent-based simulation II, third international workshop, MABS 2002, Bologna, July 2002, revised papers (Lecture notes in AI), vol 2581. Springer, Berlin, pp 50–69

Jonker CM, Treur J, Wijngaards WCA (2001) Temporal languages for simulation and analysis of the dynamics within an organisation. In: Dunin-Keplicz B, Nawarecki E (eds) From theory to practice in multi-agent systems, proceedings of the second international workshop of central and eastern Europe on multi-agent systems, CEEMAS'01 (Lecture notes in computer science), vol 2296. Springer, Berlin, pp 151–160

Kreitner R, Kunicki A (2001) Organisational behavior. McGraw-Hill, New York

Lomi A, Larsen ER (2001) Dynamics of organizations: computational modeling and organization theories. AAAI Press, Menlo Park

Mintzberg H (1979) The structuring of organisations. Prentice Hall, Englewood Cliffs

Moss S, Gaylard H, Wallis S, Edmonds B (1998) SDML: a multi-agent language for organizational modelling. Comp Math Org Theory 4(1):43–70

Norling E, Edmonds B, Meyer R (2013) Informal approaches to developing simulation models. Chapter 4 in this volume

Nwana HS (1996) Software agents: an overview. Knowl Eng Rev 11(3):205–244

Nwana HS, Ndumu DT (1998) A brief introduction to software agent technology. In: Jennings M, Wooldridge NR (eds) *Agent Technology: Foundations, Applications and Markets*. Springer Verlag, Berlin, pp 29–47

Prietula M, Gasser L, Carley K (1997) Simulating organizations. MIT Press, Cambridge, MA

Railsback SF, Lytinen SL, Jackson SK (2006) Agent-based simulation platforms: review and development recommendations. Simulation 82(9):609–623

Sichman JS, Conte R (1998) On personal and role mental attitudes: a preliminary dependence-based analysis. In: Oliveira F (ed) Advances in AI. Proceedings of the 14th Brazilian symposium on artificial intelligence, SBIA'98 (Lecture notes in artificial intelligence), vol 1515. Springer, Berlin, pp 1–10

Simon HA (1962) The architecture of complexity. Proc Am Philos Soc 106(6):467–482

Wooldridge M, Jennings NR (1995) Agent theories, architectures, and languages: a survey. In: Wooldridge M, Jennings NR (eds) Intelligent agents, proceedings of the first international workshop on agent theories, architectures and languages, ATAL'94 (Lecture notes in AI), vol 890. Springer, Berlin, pp 1–39

# Chapter 6
# Checking Simulations: Detecting and Avoiding Errors and Artefacts

**José M. Galán, Luis R. Izquierdo, Segismundo S. Izquierdo, José I. Santos, Ricardo del Olmo, and Adolfo López-Paredes**

**Why Read This Chapter?** Given the complex and exploratory nature of many agent-based models, checking that the model performs in the manner intended by its designers is a very challenging task. This chapter helps the reader to identify some of the possible types of error and artefact that may appear in the different stages of the modelling process. It will also suggest some activities that can be conducted to detect, and hence avoid, each type.

**Abstract** The aim of this chapter is to provide the reader with a set of concepts and a range of suggested activities that will enhance his or her ability to understand agent-based simulations. To do this in a structured way, we review the main concepts of the methodology (e.g. we provide precise definitions for the terms 'error' and 'artefact') and establish a general framework that summarises the process of designing, implementing, and using agent-based models. Within this framework we identify the various stages where different types of assumptions are usually made and, consequently, where different types of errors and artefacts may appear. We then propose several activities that can be conducted to detect each type of error and artefact.

J.M. Galán (✉) • L.R. Izquierdo • J.I. Santos • R. del Olmo
Departamento de Ingeniería Civil, Universidad de Burgos, Burgos E-09001, Spain
e-mail: jmgalan@ubu.es; luis@izquierdo.name; jisantos@ubu.es; rdelolmo@ubu.es

S.S. Izquierdo • A. López-Paredes
Departamento de Ingeniería Civil, Universidad de Burgos, Valladolid E-47011, Spain
e-mail: segis@eis.uva.es; adolfo@insisoc.org

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_6,

## 6.1 Introduction

Agent-based modelling is one of multiple techniques that can be used to conceptualise social systems. What distinguishes this methodology from others is the use of a more direct correspondence between the entities in the system to be modelled and the agents that represent such entities in the model (Edmonds 2001). This approach offers the potential to enhance the transparency, soundness, descriptive accuracy, and rigour of the modelling process, but it can also create difficulties: agent-based models are generally complex and mathematically intractable, so their exploration and analysis often require computer simulation.

The problem with computer simulations is that understanding them in reasonable detail is not as straightforward an exercise as one could think (this also applies to one's own simulations). A computer simulation can be seen as the process of applying a certain function to a set of inputs to obtain some results. This function is usually so complicated and cumbersome that the computer code itself is often not that far from being one of the best descriptions of the function that can be provided. Following this view, understanding a simulation would basically consist in identifying the parts of the mentioned function that are responsible for generating particular (sub)sets of results.

Thus, it becomes apparent that a prerequisite to understand a simulation is to make sure that there is no significant disparity between what we think the computer code is doing and what is actually doing. One could be tempted to think that, given that the code has been programmed by someone, surely there is always at least one person – the programmer – who knows precisely what the code does. Unfortunately, the truth tends to be quite different, as the leading figures in the field report:

> You should assume that, no matter how carefully you have designed and built your simulation, it will contain bugs (code that does something different to what you wanted and expected). (Gilbert 2007)
>
> An unreplicated simulation is an untrustworthy simulation – do not rely on their results, they are almost certainly wrong. ('Wrong' in the sense that, at least in some detail or other, the implementation differs from what was intended or assumed by the modeller). (Edmonds and Hales 2003)
>
> Achieving internal validity is harder than it might seem. The problem is knowing whether an unexpected result is a reflection of a mistake in the programming, or a surprising consequence of the model itself. [. . .] As is often the case, confirming that the model was correctly programmed was substantially more work than programming the model in the first place. (Axelrod 1997a)

In the particular context of *agent-based* simulation, the problem tends to be exacerbated. The complex and exploratory nature of most agent-based models implies that, before running a model, there is almost always some uncertainty about what the model will produce. Not knowing a priori what to expect makes it difficult to discern whether an unexpected outcome has been generated as a legitimate result of the assumptions embedded in the model or, on the contrary, it is due to an error or an artefact created in its design, its implementation, or in the running process (Axtell and Epstein 1994: 31; Gilbert and Terna 2000).

Moreover, the challenge of understanding a computer simulation does not end when one is confident that the code is free from errors; the complex issue of identifying what parts of the code are generating a particular set of outputs remains to be solved. Stated differently, this is the challenge of discovering what assumptions in the model are causing the results we consider significant. Thus, a substantial part of this non-trivial task consists in detecting and avoiding artefacts: significant phenomena caused by accessory assumptions in the model that are (mistakenly) deemed irrelevant. We explain this in detail in subsequent sections.

The aim of this chapter is to provide the reader with a set of concepts and a range of suggested activities that will enhance his ability to understand simulations. As mentioned before, simulation models can be seen as functions operating on their inputs to produce the outputs. These functions are created by putting together a range of different assumptions of very diverse nature. Some assumptions are made because they are considered to be an essential feature of the system to be modelled; others are included in a somewhat arbitrary fashion to achieve completeness – i.e. to make the computer model run –, and they may not have a clear referent in the target system. There are also assumptions – e.g. the selection of the compiler and the particular pseudo-random number generator to be employed – that are often made, consciously or not, without fully understanding in detail how they work, but *trusting* that they operate in the way we think they do. Finally, there may also be some assumptions in a computer model that not even its own developer is aware of, e.g. the use of floating-point arithmetic, rather than real arithmetic.

Thus, in broad terms, understanding simulations requires identifying what assumptions are being made, and assessing their impact on the results. To achieve this, we believe that it is useful to characterise the process by which assumptions accumulate to end up forming a complete model. We do this in a structured way by presenting a general framework that summarises the process of creating and using agent-based models through various stages; then, within this framework, we characterise the different types of assumptions that are made in each of the stages of the modelling process, and we identify the sort of errors and artefacts that may occur; we also propose activities that can be conducted to avoid each type of error or artefact.

The chapter is structured as follows: the following section is devoted to explaining what we understand by modelling, and to argue that computer simulation is a useful tool to explore formal models, rather than a distinctively new symbolic system or a uniquely different reasoning process, as it has been suggested in the literature. In Sect. 6.3 we explain what the essence of agent-based modelling is in our view, and we present the general framework that summarises the process of designing, implementing, and using agent-based models. In Sect. 6.4 we define the concepts of error and artefact, and we discuss their relevance for validation and verification. The framework presented in Sect. 6.3 is then used to identify the various stages of the modelling process where different types of assumptions are made and, consequently, where different types of errors and artefacts may appear. We then propose various activities aimed at avoiding the types of errors and artefacts previously described, and we conclude with a brief summary of the chapter.

## 6.2 Three Symbolic Systems Used to Model Social Processes

Modelling is the art of building models. In broad terms, a model can be defined as an abstraction of an observed system that enables us to establish some kind of inference process about how the system works, or about how certain aspects of the system operate.

Modelling is an activity inherent to every human being: people constantly develop mental models, more or less explicit, about various aspects of their daily life. Within science in particular, models are ubiquitous. Many models in the "hard" sciences are formulated using mathematics (e.g. differential equation models and statistical regressions), and they are therefore formal, but it is also perfectly feasible – and acceptable – to build non-formal models within academia; this is often the case in disciplines like history or sociology – consider e.g. a model written in natural language that tries to explain the expansion of the Spanish Empire in the sixteenth century, or the formation of urban "tribes" in large cities.

We value a model to the extent that it is useful – i.e. in our opinion, what makes a model good is its fitness for purpose. Thus, the assessment of any model can only be conducted relative to a predefined purpose. Having said that, there is a basic set of general features that are widely accepted to be desirable in any model, e.g. accuracy, precision, generality, and simplicity (see Fig. 6.1). Frequently some of these features are inversely related; in such cases the modeller is bound to compromise to find a suitable trade-off, considering the perceived relative importance of each of these desirable features for the purpose of the model (Edmonds 2005).

Some authors (Gilbert 1999; Holland and Miller 1991; Ostrom 1988) classify the range of available techniques for modelling phenomena in which the social dimension is influential according to three symbolic systems.

One possible way of representing and studying social phenomena is through verbal argumentation in natural language. This is the symbolic system traditionally used in historical analyses, which, after a process of abstraction and simplification, describe past events emphasising certain facts, processes, and relations at the expense of others. The main problem with this type of representation is its intrinsic lack of precision (due to the ambiguity of natural language) and the associated difficulty of uncovering the exact implications of the ideas put forward in this way. In particular, using this symbolic system it is often very difficult to determine the whole range of inferences that can be obtained from the assumptions embedded in the model in reasonable detail; therefore it is often impossible to assess its logical consistency, its scope, and its potential for generalisation in a formal way.

A second symbolic system that is sometimes used in the Social Sciences, particularly in Economics, is the set of formal languages (e.g. leading to models expressed as mathematical equations). The main advantage of this symbolic system derives from the possibility of using formal deductive reasoning to infer new facts from a set of clearly specified assumptions; formal deductive reasoning guarantees that the obtained inferences follow from the axioms with logical consistency. Formal languages also facilitate the process of assessing the generality of a

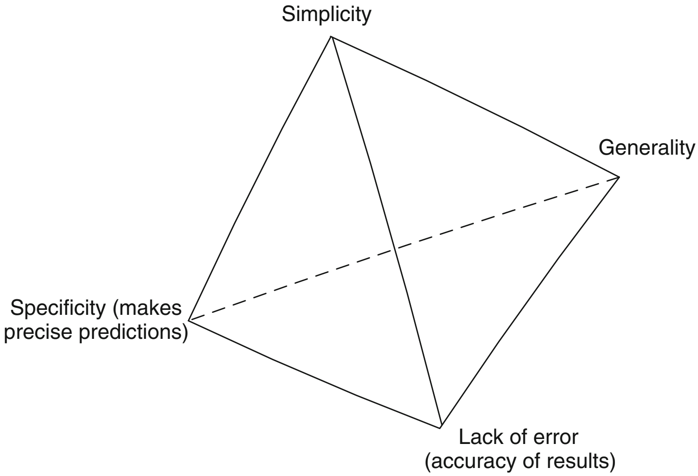



**Fig. 6.1** The trade-off between various desirable features depends on the specific case and model. There are not general rules that relate, not even in a qualitative fashion, all these features. The figure shows a particular example from Edmonds (2005) that represents the possible equilibrium relationships between some features in a particular model

model and its sensitivity to assumptions that are allowed to change within the boundaries of the model (i.e. parameter values and non-structural assumptions).

However, the process of reducing social reality to formal models is not exempt from disadvantages. Social systems can be tremendously complex, so if such systems are to be abstracted using a formal language (e.g. mathematical equations), we run the risk of losing too much in descriptiveness. To make things worse, in those cases where it appears possible to produce a satisfactory formal model of the social system under investigation, the resulting equations may be so complex that the formal model becomes mathematically intractable, thus failing to provide most of the benefits that motivated the process of formalisation in the first place. This is particularly relevant in the domain of the Social Sciences, where the systems under investigation often include non-linear relations (Axtell 2000). The usual approach then is to keep on adding simplifying hypotheses to the model – thus making it increasingly restrictive and unrealistic – until we obtain a tractable model that can be formally analysed with the available tools. We can find many examples of such assumptions in Economics: instrumental rationality, perfect information, representative agents, etc. Most often these concepts are not included because economists think that the real world works in this way, but to make the models tractable (see for instance Conlisk (1996), Axelrod (1997a), Hernández (2004), Moss (2001, 2002)). It seems that, in many cases, the use of formal symbolic systems tends to increase the danger of letting the pursuit for tractability be the driver of the modelling process.

But then, knowing that many of the hypotheses that researchers are obliged to assume may not hold in the real world, and could therefore lead to deceptive conclusions and theories, does this type of modelling representation preserve its advantages? Quoting G.F. Shove, it could be the case that sometimes "*it is better to be vaguely right than precisely wrong*".

The third symbolic system, computer modelling, opens up the possibility of building models that somewhat lie in between the descriptive richness of natural language and the analytical power of traditional formal approaches. This third type of representation is characterized by representing a model as a computer program (Gilbert and Troitzsch 1999). Using computer simulation we have the potential to build and study models that to some extent combine the intuitive appeal of verbal theories with the rigour of analytically tractable formal modelling.

In Axelrod's (1997a) opinion, computational simulation is the third way of doing science, which complements induction –the search for patterns in data– and deduction –the proof of theorems from a set of fixed axioms. In his opinion, simulation, like deduction, starts from an explicit set of hypotheses but, rather than generating theorems, it generates data that can be inductively analysed.

While the division of modelling techniques presented above seems to be reasonably well accepted in the social simulation community –and we certainly find it useful–, we do not fully endorse it. In our view, computer simulation does not constitute a distinctively new symbolic system or a uniquely different reasoning process by itself, but rather a (very useful) tool for exploring and analysing formal systems. We see computers as inference engines that are able to conduct algorithmic processes at a speed that the human brain cannot achieve. The inference derived from running a computer model is constructed by example and, in the general case, reads: *the results obtained from running the computer simulation follow (with logical consistency) from applying the algorithmic rules that define the model on the input parameters*[1] *used*.

In this way, simulations allow us to explore the properties of certain formal models that are intractable using traditional formal analyses (e.g. mathematical analyses), and they can also provide fundamentally new insights even when such analyses are possible. Like Gotts et al. (2003), we also believe that mathematical analysis and simulation studies should not be regarded as alternative and even opposed approaches to the formal study of social systems, but as complementary. They are both extremely useful tools to analyse formal models, and they are complementary in the sense that they can provide fundamentally different insights on one same model.

To summarise, a computer program is a formal model (which can therefore be expressed in mathematical language, e.g. as a set of stochastic or deterministic equations), and computer simulation is a tool that enables us to study it in ways that go beyond mathematical tractability. Thus, the final result is a potentially more realistic – and still formal – study of a social system.

[1] By *input parameters* in this statement we mean "everything that may affect the output of the model", e.g. the random seed, the pseudo-random number generator employed and, potentially, information about the microprocessor and operating system on which the simulation was run, if these could make a difference.

## 6.3 Agent Based Modelling

### *6.3.1 Concept*

As stated before, modelling is the process of building an abstraction of a system for a specific purpose (see Edmonds (2001) and Epstein (2008) for a list of potential applications). Thus, in essence, what distinguishes one modelling paradigm from another is precisely the way we construct that abstraction from the observed system.

In our view, agent-based modelling is a modelling paradigm with the defining characteristic that entities within the target system to be modelled – and the interactions between them – are explicitly and individually represented in the model (see Fig. 6.2). This is in contrast to other models where some entities are represented via average properties or via single representative agents. In many other models, entities are not represented at all, and it is only processes that are studied (e.g. a model of temperature variation as a function of pressure), and it is worth noting that such processes may well be already abstractions of the system.[2] The specific process of abstraction employed to build one particular model does not necessarily make it better or worse, only more or less useful for one purpose or another.

The specific way in which the process of abstraction is conducted in agent-based modelling is attractive for various reasons: it leads to (potentially) formal yet more natural and transparent descriptions of the target system, provides the possibility to model heterogeneity almost by definition, facilitates an explicit representation of the environment and the way other entities interact with it, allows for the study of the bidirectional relations between individuals and groups, and it can also capture emergent behaviour (see Epstein 1999; Axtell 2000; Bonabeau 2002). Unfortunately, as one would expect, all these benefits often come at a price: most of the models built in this way are mathematically intractable. A common approach to study the behaviour of mathematically intractable formal models is to use computer simulation. It is for this reason that we often find the terms "agent-based modelling" and "agent-based simulation" used as synonyms in the scientific literature (Hare and Deadman 2004).

Thus, to summarise our thoughts in the context of the classification of modelling approaches in the social sciences, we understand that the essence of agent-based modelling is the individual and explicit representation of the entities and their interactions in the model, whereas computer simulation is a useful tool for studying the implications of formal models. This tool happens to be particularly well suited to explore and analyse agent-based models for the reasons explained above. Running an agent-based model in a computer provides a formal proof that a particular micro-specification is *sufficient* to generate the global behaviour that is observed

[2] The reader can see an interesting comparative analysis between agent-based and equation-based modelling in (Parunak et al. 1998).

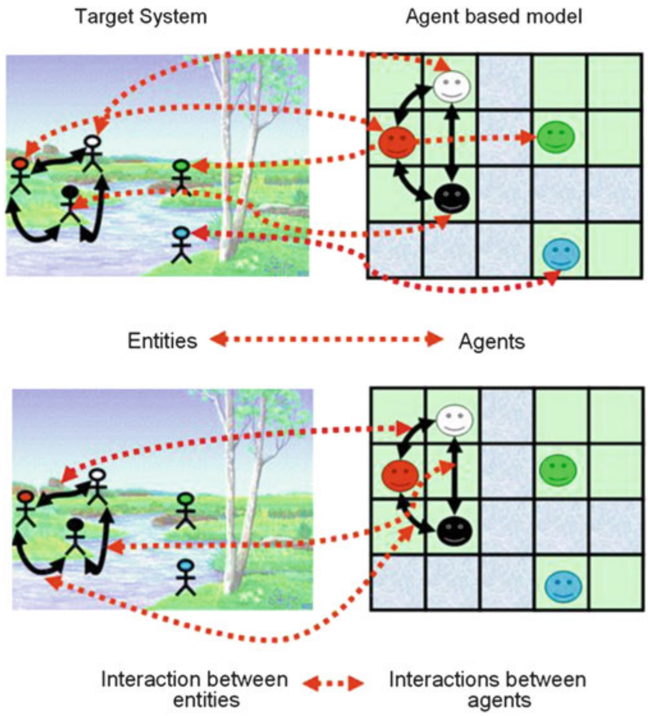



**Fig. 6.2** In agent based modelling the entities of the system are represented explicitly and individually in the model. The limits of the entities in the target system correspond to the limits of the agents in the model, and the interactions between entities correspond to the interactions of the agents in the model (Edmonds 2001)

during the simulation. If a model can be run in a computer, then it is in principle possible to express it in many different formalisms, e.g. as a set of mathematical equations. Such equations may be very complex, difficult to interpret and impossible to solve, thus making the whole exercise of changing formalism frequently pointless, but what we find indeed useful is *the thought* that such an exercise *could* be undertaken, i.e. an agent-based model that can be run in a computer is not that different from the typical mathematical model. As a matter of fact, it is not difficult to formally characterise most agent-based models in a general way (Leombruni and Richiardi 2005).

### 6.3.2 *Design, Implementation, and Use of an Agent-Based Model*

Drogoul et al. (2003) identify three different roles in the design, implementation, and use of a typical agent-based model: the *thematician* (domain expert), the *modeller*, and the *computer scientist*. It is not unusual in the field to observe that

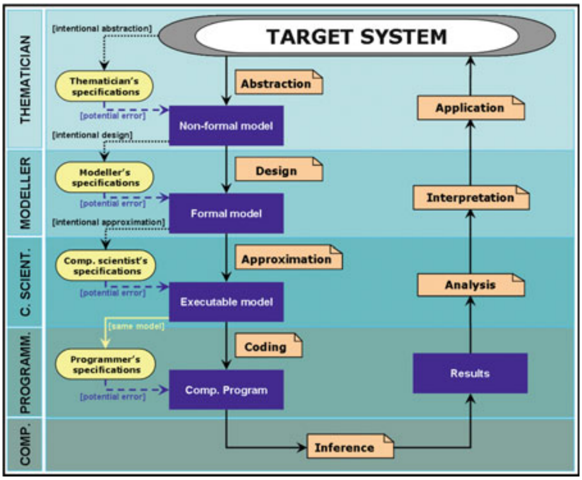



**Fig. 6.3** Different stages in the process of designing, implementing and using an agent-based model

one single person undertakes several or even all of these roles. We find that these three roles fit particularly well into the framework put forward by Edmonds (2001) to describe the process of modelling with an intermediate abstraction. Here we marry Drogoul et al.'s and Edmonds' views on modelling by dissecting one of Drogoul et al.'s roles and slightly expanding Edmonds' framework (Fig. 6.3). We then use our extended framework to identify the different types of assumptions that are made in each of the stages of the modelling process, the errors and artefacts that may occur in each of them, and the activities that can be conducted to avoid such errors and artefacts. We start by explaining the three different roles proposed by Drogoul et al. (2003).

The role of the *thematician* is undertaken by experts in the target domain. They are the ones that better understand the target system, and therefore the ones who carry out the abstraction process that is meant to produce the first conceptualisation of the target system. Their job involves defining the objectives and the purpose of the modelling exercise, identifying the critical components of the system and the linkages between them, and also describing the most prominent causal relations. The output of this first stage of the process is most often a non-formal model expressed in natural language, and it may also include simple conceptual diagrams, e.g. block diagrams. The non-formal model produced may describe the system using potentially ambiguous terms (such as e.g. learning or imitation, without fully specifying how these processes actually take place).

The next stage in the modelling process is carried out by the role of the *modeller*. The modeller's task is to transform the non-formal model that the *thematician* aims to explore into the (formal) requirement specifications that the *computer scientist* – the third role – needs to formulate the (formal) executable model. This job involves (at least) three major challenges. The first one consists in acting as a mediator between two domains that are very frequently fundamentally different (e.g. sociology and computer science). The second challenge derives from the fact that in most cases the *thematician*'s model is not fully specified, i.e. there are many formal models that would conform to it.[3] In other words, the formal model created by the *modeller* is most often just one of many possible particularisations of the *thematician*'s (more general) model. Lastly, the third challenge appears when the *thematician*'s model is not consistent, which may perfectly be the case since his model is often formulated using natural language. Discovering inconsistencies in natural language models is in general a non-trivial task. Several authors (e.g. Christley et al. (2004), Pignotti et al. (2005), and Polhill and Gotts (2006)) have identified ontologies to be particularly useful for this purpose, especially in the domain of agent-based social simulation. Polhill and Gotts (2006) write:

> An ontology is defined by Gruber (1993) as "a formal, explicit specification of a shared conceptualisation". Fensel (2001) elaborates: ontologies are formal in that they are machine readable; explicit in that all required concepts are described; shared in that they represent an agreement among some community that the definitions contained within the ontology match their own understanding; and conceptualisations in that an ontology is an abstraction of reality. (Polhill and Gotts 2006, p. 51)

Thus, the modeller has the difficult –potentially unfeasible– task of finding a set of (formal and consistent) requirement specifications[4] where each individual requirement specification of that set is a legitimate particular case of the *thematician*'s model, and the set as a whole is *representative* of the *thematician*'s specifications (i.e. the set is sufficient to fully characterise the *thematician*'s model to a satisfactory extent).

Drogoul et al.'s third role is the *computer scientist*. Here we distinguish between *computer scientist* and *programmer*. It is often the case that the modeller comes up with a formal model that cannot be implemented in a computer. This could be, for example, because the model uses certain concepts that cannot be operated by present-day computers (e.g. real numbers, as opposed to floating-point numbers), or because running the model would demand computational requirements that are not yet available (e.g. in terms of memory and processing capacity). The job of the *computer scientist* consists in finding a suitable (formal) approximation to the *modeller*'s formal model that can be executed in a computer (or in several computers) given the available technology. To achieve this, the *computer scientist* may have to approximate and simplify certain aspects of the *modeller*'s formal model, and it is his job to make sure that these simplifications are not affecting the results significantly. As an example, Cioffi-Revilla (2002) warns about the potentially significant effects of altering system size in agent-based simulations.

The Navier–Stokes equations of fluid dynamics are a paradigmatic case in point. They are a set of non-linear differential equations that describe the motion of a fluid. Although these equations are considered a very good (formal and fully specified)

[3] Note that the *thematician* faces a similar problem when building his non-formal model. There are potentially an infinite number of models for one single target system.

[4] Each individual member of this set can be understood as a different model or, alternatively, as a different parameterisation of one single –more general– model that would itself define the whole set.

model, their complexity is such that analytical closed-form solutions are available only for the simplest cases. For more complex situations, solutions of the Navier–Stokes equations must be estimated using approximations and numerical computation (Heywood et al. 1990; Salvi 2002). Deriving such approximations would be the task of the *computer scientist*'s role, as defined here.

One of the main motivations to distinguish between the *modeller*'s role and the *computer scientist*'s role is that, in the domain of agent-based social simulation, it is the description of the *modeller*'s formal model what is usually found in academic papers, even though the *computer scientist*'s model was used by the authors to produce the results in the paper. Most often the *modeller*'s model (i.e. the one described in the paper) simply *cannot* be run in a computer; it is the (potentially faulty) implementation of the *computer scientist*'s approximation to such a model what is really run by the computer. As an example, note that computer models described in scientific papers are most often expressed using equations in real arithmetic, whereas the models that actually run in computers almost invariably use floating-point arithmetic.

Finally, the role of the *programmer* is to implement the *computer scientist*'s executable model. In our framework, by definition of the role *computer scientist*, the model he produces must be executable and fully specified, i.e. it must include all the necessary information so given a certain input the model always produces the same output. Thus, the executable model will have to specify in its definition everything that could make a difference, e.g. the operating system and the specific pseudo-random number generator to be used. This is a subtle but important point, since it implies that the *programmer*'s job does not involve any process of abstraction or simplification; i.e. the executable model and the *programmer*'s specifications are by definition *the same* (see Fig. 6.3). (We consider two models to be *the same* if and only if they produce the same outputs when given the same inputs.) The *programmer*'s job consists "only" in writing the executable model in a programming language.[5] If the *programmer* does not make any mistakes, then the implemented model (e.g. the code) and the executable model will be the same.

Any mismatch between someone's specifications and the actual model he passes to the next stage is considered here an error (see Fig. 6.3). As an example, if the code implemented by the programmer is not the same model as his specifications, then there has been an implementation error. Similarly, if the *computer scientist*'s specifications are not complete (i.e. they do not define a unique model that produces a precise set of outputs for each given set of inputs) we say that he has made an error since the model he is producing is necessarily fully specified (by definition of the role). This opens up the question of how the executable model is defined: the executable model is *the same model* as the code if the *programmer* does not

[5] There are some interesting attempts with INGENIAS (Pavón and Gómez-Sanz 2003) to use modelling and visual languages as programming languages rather than merely as design languages (Sansores and Pavón 2005; Sansores et al. 2006. These efforts are aimed at automatically generating several implementations of one single executable model (in various different simulation platforms).

make any mistakes. So, to be clear, the distinction between the role of *computer scientist* and *programmer* is made here to distinguish (a) errors in the implementation of a fully specified model (which are made by the *programmer*) from (b) errors derived from an incomplete understanding of how a computer program works (which are made by the *computer scientist*). An example of the latter would be one where the *computer scientist*'s specifications stipulate the use of real arithmetic, but the executable model uses floating-point arithmetic.

It is worth noting that in an ideal world the specifications created by each role would be written down. Unfortunately the world is far from ideal, and it is often the case that the mentioned specifications stay in the realm of mental models, and never reach materialisation.

The reason for which the last two roles in the process are called 'the *computer scientist*' and the '*programmer*' is because, as mentioned before, most agent-based models are implemented as computer programs, and then explored through simulation (for tractability reasons). However, one could also think of e.g. a mathematician conducting these two roles, especially if the formal model provided by the *modeller* can be solved analytically. For the sake of clarity, and without great loss of generality, we assume here that the model is implemented as a computer program and its behaviour is explored through computer simulation.

Once the computer model is implemented, it is run, and the generated results are analysed. The analysis of the results of the computer model leads to conclusions on the behaviour of the *computer scientist*'s model and, to the extent that the *computer scientist*'s model is a valid approximation of the *modeller*'s formal model, these conclusions also apply to the *modeller*'s formal model. Again, to the extent that the formal model is a legitimate particularisation of the non-formal model created by the *thematician*, the conclusions obtained for the *modeller*'s formal model can be interpreted in the terms used by the non-formal model. Furthermore, if the *modeller*'s formal model is representative of the *thematician*'s model, then there is scope for making general statements on the behaviour of the *thematician*'s model. Finally, if the *thematician*'s model is satisfactorily capturing social reality, then the knowledge inferred in the whole process can be meaningfully applied to the target system.

In the following section we use our extended framework to identify the different errors and artefacts that may occur in each of the stages of the modelling process and the activities that can be conducted to avoid such errors and artefacts.

## 6.4 Errors and Artefacts

### *6.4.1 Definition of Error and Artefact and Their Relevance for Validation and Verification*

Since the meanings of the terms validation, verification, error, and artefact are not uncontested in the literature, we start by stating the meaning that *we* attribute to each of them. For us, validation is the process of assessing how useful a model is for a certain purpose. A model is valid to the extent that it provides a satisfactory range of accuracy consistent with the intended application of the model (Kleijnen 1995; Sargent 2003).[6] Thus, if the objective is to accurately represent social reality, then validation is about assessing how well the model is capturing the essence of its empirical referent. This could be measured in terms of goodness of fit to the characteristics of the model's referent Moss et al. (1997).

Verification – sometimes called "internal validation", e.g. by Taylor (1983), Drogoul et al. (2003), Sansores and Pavón (2005), or "internal validity", e.g. by Axelrod (1997a) – is the process of ensuring that the model performs in the manner intended by its designers and implementers (Moss et al. 1997). Let us say that a model is *correct* if and only if it would pass a verification exercise. Using our previous terminology, an expression of a model in a language is correct if and only if it is *the same* model as the developer's specifications. Thus, it could well be the case that a correct model is not valid (for a certain purpose). Conversely, it is also possible that a model that is not correct is actually valid for some purposes. Having said that, one would think that the chances of a model being valid are higher if it performs in the manner intended by its designer. To be sure, according to our definition of validation, what we want is a valid model, and we are interested in its correctness only to the extent that correctness contributes to make the model valid.

We also distinguish between errors and artefacts (Galán et al. 2009). *Errors* appear when a model does not comply with the requirement specifications self-imposed by its own developer. In simple words, an error is a mismatch between what the developer thinks the model is, and what it actually is. It is then clear that there is an error in the model if and only if the model is not correct. Thus, verification is the process of looking for errors. An example of an implementation error would be the situation where the *programmer* intends to loop through the whole list of agents in the program, but he mistakenly writes the code so it only runs through a subset of them. A less trivial example of an error would be the situation where it is believed that a program is running according to the rules of real arithmetic, while the program is actually using floating-point arithmetic (Izquierdo and Polhill 2006; Polhill and Izquierdo 2005; Polhill et al. 2005, 2006).

In contrast to errors, *artefacts* relate to situations where there is no mismatch between what the developer thinks a model is and what it actually is. Here the mismatch is between the set of assumptions in the model that the developer thinks are producing a certain phenomenon, and the assumptions that are the actual cause of such phenomenon. We explain this in detail. We distinguish between *core* and *accessory* assumptions in a model. *Core* assumptions are those whose presence is believed to be important for the purpose of the model. Ideally these would be the only assumptions present in the model. However, when producing a formal model it is often the case that the developer is bound to include some additional assumptions for the only purpose of making the model complete. We call these *accessory*

[6] See a complete epistemic review of the validation problem in Kleindorfer et al. (1998).

assumptions. Accessory assumptions are not considered a crucial part of the model; they are included to *make the model work*. We also distinguish between *significant* and *non-significant* assumptions. A *significant* assumption is an assumption that is the cause of some significant result obtained when running the model. Using this terminology, we define *artefacts* as significant phenomena caused by *accessory* assumptions in the model that are (mistakenly) deemed *non-significant*. In other words, an artefact appears when an accessory assumption that is considered non-significant by the developer is actually significant. An example of an artefact would be the situation where the topology of the grid in a model is accessory, it is believed that some significant result obtained when running the model is independent of the particular topology used (say, e.g. a grid of square cells), but it turns out that if an alternative topology is chosen (say, e.g. hexagonal cells) then the significant result is not observed.

The relation between artefacts and validation is not as straight-forward as that between errors and verification. For a start, artefacts are relevant for validation only to the extent that identifying and understanding causal links in the model's referent is part of the purpose of the modelling exercise. We assume that this is the case, as indeed it usually is in the field of agent-based social simulation. A clear example is the Schelling-Sakoda model of segregation, which was designed to investigate the causal link between individual preferences and global patterns of segregation (Sakoda 1971; Schelling 1971, 1978). The presence of artefacts in a model implies that the model is not representative of its referent, since one can change some accessory assumption (thus creating an alternative model which still includes all the core assumptions) and obtain significantly different results. When this occurs, we run the risk of interpreting the results obtained with the (non-representative) model beyond its scope (Edmonds and Hales 2005). Thus, to the extent that identifying causal links in the model's referent is part of the purpose of the modelling exercise, the presence of artefacts decreases the validity of the model. In any case, the presence of artefacts denotes a misunderstanding of what assumptions are generating what results.

### *6.4.2 Appearance of Errors and Artefacts*

The dynamics of agent-based models are generally sufficiently complex that model developers themselves do not understand in exhaustive detail how the obtained results have been produced. As a matter of fact, in most cases if the exact results and the processes that generated them were known and fully understood in advance, there would not be much point in running the model in the first place. Not knowing exactly what to expect makes it impossible to tell whether any unanticipated results derive exclusively from what the researcher believes are the core assumptions in the model, or whether they are due to errors or artefacts. The question is of crucial importance since, unfortunately, the truth is that there are many things that can go wrong in modelling.

Errors and artefacts may appear at various stages of the modelling process (Galán and Izquierdo 2005). In this section we use the extended framework explained in the previous section to identify the critical stages of the modelling process where errors and artefacts are most likely to occur.

According to our definition of artefact – i.e. significant phenomena caused by accessory assumptions that are not considered relevant –, artefacts *cannot* appear in the process of abstraction conducted by the *thematician*, since this stage consists precisely in distilling the *core* features of the target system. Thus, there should not be accessory assumptions in the *thematician*'s model. Nevertheless, there could still be issues with validation if, for instance, the *thematician*'s model is not capturing social reality to a satisfactory extent. Errors could appear in this stage because the *thematician*'s specifications are usually expressed in natural language, and rather than being written down, they are often transmitted orally to the modeller. Thus, an error (i.e. a mismatch between the *thematician*'s specifications and the non-formal model received by the *modeller*) could appear here if the *modeller* misunderstands some of the concepts put forward by the *thematician*.

The *modeller* is the role that may introduce the first artefacts in the modelling process. When formalising the *thematician*'s model, the *modeller* will often have to make a number of additional assumptions so the produced formal model is fully specified. By our definition of the two roles, these additional assumptions are not crucial features of the target system. If such accessory assumptions have a significant impact on the behaviour of the model and the *modeller* is not aware of it, then an artefact has been created. This would occur if, for instance, (a) the *thematician* did not specify any particular neighbourhood function, (b) different neighbourhood functions lead to different results, and (c) the *modeller* is using only one of them and believes that they all produce essentially the same results.

Errors could also appear at this stage, although it is not very likely. This is so because the specifications that the *modeller* produces must be formal, and they are therefore most often written down in a formal language. When this is the case, there is little room for misunderstanding between the *modeller* and the computer scientist, i.e. the *modeller*'s specifications and the formal model received by the *computer scientist* would be the same, and thus there would be no error at this stage.

The role of the *computer scientist* could introduce artefacts in the process. This would be the case if, for instance, his specifications require the use of a particular pseudo-random number generator, he believes that this choice will not have any influence in the results obtained, but it turns out that it does. Similar examples could involve the arbitrary selection of an operating system or a specific floating-point arithmetic that had a significant effect on the output of the model.

Errors can quite easily appear in between the role of the *computer scientist* and the role of the programmer. Note that in our framework any mismatch between the *computer scientist*'s specifications and the executable model received by the *programmer* is considered an error. In particular, if the *computer scientist's* specifications are not executable, then there is an error. This could be, for instance, because the *computer scientist's* specifications stipulate requirements that cannot be executed with present-day computers (e.g. real arithmetic), or because it does not specify all the necessary information to be run in a computer in an unequivocal way (e.g. it does not specify a particular pseudo-random number generator). The error then may affect the validity of the model significantly, or may not.

Note from the previous examples that if the *computer scientist* does not provide a fully executable set of requirement specifications, then he is introducing an error, since in that case the computer program (which is executable) would be necessarily different from his specifications. On the other hand, if he does provide an executable model but in doing so he makes an arbitrary accessory assumption that turns out to be significant, then he is introducing an artefact.

Finally, the *programmer* cannot introduce artefacts because his specifications are the same as the executable model by definition of the role (i.e. the *programmer* does not have to make any accessory assumptions). However, he may make mistakes when creating the computer program from the executable model.

### 6.4.3 Activities Aimed at Detecting Errors and Artefacts

In this section we identify various activities that the different roles defined in the previous sections can undertake to detect errors and artefacts. We consider the use of these techniques as a very recommendable and eventually easy to apply practice. In spite of this, we should warn that, very often, these activities may require a considerable human and computational effort.

#### 6.4.3.1 *Modeller*'s Activities

- Develop and analyse new formal models by implementing alternative accessory assumptions while keeping the core assumptions identified by the *thematician*. This exercise will help to detect artefacts. Only those conclusions which are not falsified by any of these models will be valid for the *thematician*'s model. As an example, see Galán and Izquierdo (2005), who studied different instantiations of one single conceptual model by implementing different evolutionary selection mechanisms. Takadama et al. (2003) conducted a very similar exercise implementing three different learning algorithms for their agents. In a collection of papers, Klemm et al. (2003a, b, c, 2005) investigate the impact of various accessory assumptions in Axelrod's model for the dissemination of culture (Axelrod 1997b). Another example of studying different formal models that address one single problem is provided by Kluver and Stoica (2003).
- Conduct a more exhaustive exploration of the parameter space within the boundaries of the *thematician*'s specifications. If we obtain essentially the same results using the wider parameter range, then we will have broadened the scope of the model, thus making it more representative of the *thematician*'s model. If, on the other hand, results change significantly, then we will have

identified artefacts. This type of exercise has been conducted by e.g. Castellano et al. (2000) and Galán and Izquierdo (2005).

- Create abstractions of the formal model which are mathematically tractable. An example of one possible abstraction would be to study the *expected* motion of a dynamic system (see the studies conducted by Galán and Izquierdo (2005), Edwards et al. (2003), and Castellano et al. (2000) for illustrations of mean-field approximations). Since these mathematical abstractions do not correspond in a one-to-one way with the specifications of the formal model, any results obtained with them will not be conclusive, but they may suggest parts of the model where there may be errors or artefacts.
- Apply the simulation model to relatively well understood and predictable situations to check that the obtained results are in agreement with the expected behaviour (Gilbert et al. 2000).

#### 6.4.3.2 *Computer Scientist*'s Activities

- Develop mathematically tractable models of certain aspects, or particular cases, of the *modeller*'s formal model. The analytical results derived with these models should match those obtained by simulation; a disparity would be an indication of the presence of errors.
- Develop new executable models from the *modeller*'s formal model using alternative modelling paradigms (e.g. procedural vs. declarative). This activity will help to identify artefacts. As an example, see Edmonds and Hales' (2003) reimplementation of Riolo et al. (2001) model of cooperation among agents using tags. Edmonds reimplemented the model using SDML (declarative), whereas Hales reprogrammed the model in Java (procedural).
- Rerun the same code in different computers, using different operating systems, with different pseudo-random number generators. These are most often accessory assumptions of the executable model that are considered non-significant, so any detected difference will be a sign of an artefact. If no significant differences are detected, then we can be confident that the code comprises all the assumptions that could significantly influence the results. This is a valuable finding that can be exploited by the *programmer* (see next activity). As an example, Polhill et al. (2005) explain that using different compilers can result in the application of different floating-point arithmetic systems to the simulation run.

#### 6.4.3.3 *Programmer*'s Activities

- Re-implement the code in different programming languages. Assuming that the code contains all the assumptions that can influence the results significantly, this activity is equivalent to creating alternative representations of the same executable model. Thus, it can help to detect errors in the implementation. There are several examples of this type of activity in the literature. Bigbee et al. (2007) reimplemented Sugarscape (Epstein and Axtell 1996) using MASON. Xu et al. (2003) implemented one single model in Swarm and Repast. The reimplementation exercise conducted by Edmonds and Hales (2003) applies here too.

- Analyse particular cases of the executable model that are mathematically tractable. Any disparity will be an indication of the presence of errors.
- Apply the simulation model to extreme cases that are perfectly understood (Gilbert et al. 2000). Examples of this type of activity would be to run simulations without agents or with very few agents, explore the behaviour of the model using extreme parameter values, or model very simple environments. This activity is common practice in the field.

## 6.5 Summary

The dynamics of agent-based models are usually so complex that their own developers do not *fully* understand how they are generated. This makes it difficult, if not impossible, to discern whether observed significant results are legitimate logical implications of the assumptions that the model developer is interested in or whether they are due to errors or artefacts in the design or implementation of the model.

Errors are mismatches between what the developer believes a model is and what the model actually is. Artefacts are significant phenomena caused by accessory assumptions in the model that are (mistakenly) considered non-significant. Errors and artefacts prevent developers from correctly understanding their simulations. Furthermore, both errors and artefacts can significantly decrease the validity of a model, so they are best avoided.

In this chapter we have outlined a general framework that summarises the process of designing, implementing, and using agent-based models. Using this framework we have identified the different type of errors and artefacts that may occur in each of the stages of the modelling process. Finally, we have proposed several activities that can be conducted to avoid each type of error or artefact. Some of these activities include repetition of experiments in different platforms, reimplementation of the code in different programming languages, reformulation of the conceptual model using different modelling paradigms, and mathematical analyses of simplified versions or particular cases of the model. Conducting these activities will surely increase our understanding of a particular simulation model.

**Acknowledgements** The authors have benefited from the financial support of the Spanish Ministry of Education and Science (projects DPI2004–06590, DPI2005–05676 and TIN2008–06464–C03–02), the Spanish Ministry for Science and Innovation (CSD2010–00034) within the framework of CONSOLIDER-INGENIO 2010 and of the JCyL (projects VA006B09, BU034A08 and GREX251–2009). We are also very grateful to Nick Gotts, Gary Polhill, Bruce Edmonds and Cesáreo Hernández for many discussions on the philosophy of modelling.

## Further Reading

Gilbert (2007) provides an excellent basic introduction to agent based modelling. Chapter 3 summarizes the different stages involved in an agent-based modelling project, including verification and validation. The paper entitled "Some myths and common errors in simulation experiments" (Schmeiser 2001) discusses briefly some of the most common errors found in simulation from a probabilistic and statistical perspective. The approach is not focused specifically on agent based modelling but on simulation in general. Yilmaz (2006) presents an analysis of the life cycle of a simulation study and proposes a process-centric perspective for the validation and verification of agent-based computational organization models. Finally, Chap. 8 in this volume (David 2013) discusses validation in detail.

## References


Axelrod RM (1997a) Advancing the art of simulation in the social sciences. In: Conte R, Hegselmann R, Terna P (eds) Simulating social phenomena, vol 456, Lecture notes in economics and mathematical systems. Springer, Berlin, pp 21–40

Axelrod RM (1997b) The dissemination of culture: a model with local convergence and global polarization. J Confl Resolut 41(2):203–226

Axtell RL (2000) Why agents? On the varied motivations for agent computing in the social sciences. In: Macal CM, Sallach D (eds) Proceedings of the workshop on agent simulation: applications, models, and tools. Argonne National Laboratory, Argonne, pp 3–24

Axtell RL, Epstein JM (1994) Agent based modeling: understanding our creations. The Bulletin of the Santa Fe Institute, Winter, pp 28–32

Bigbee T, Cioffi-Revilla C, Luke S (2007) Replication of sugarscape using MASON. In: Terano T, Kita H, Deguchi H, Kijima K (eds) Agent-based approaches in economic and social complex systems IV: post-proceedings of the AESCS international workshop 2005. Springer, Tokyo, pp 183–190

Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 99(2):7280–7287

Castellano C, Marsili M, Vespignani A (2000) Nonequilibrium phase transition in a model for social influence. Phys Rev Lett 85(16):3536–3539

Christley S, Xiang X, Madey G (2004) Ontology for agent-based modeling and simulation. In: Macal CM, Sallach D, North MJ (eds) Proceedings of the agent 2004 conference on social dynamics: interaction, reflexivity and emergence. Argonne National laboratory/The University of Chicago, Chicago. http://www.agent2005.anl.gov/Agent2004.pdf

Cioffi-Revilla C (2002) Invariance and universality in social agent-based simulations. Proc Natl Acad Sci U S A 99(3):7314–7316

Conlisk J (1996) Why bounded rationality? J Econ Lit 34(2):669–700

David N (2013) Validating simulations. Chapter 8 in this volume

Drogoul A, Vanbergue D, Meurisse T (2003) Multi-agent based simulation: where are the agents? In: Sichman JS, Bousquet F, Davidsson P (eds) Proceedings of MABS 2002 multi-agent-based simulation, vol 2581, Lecture notes in computer science. Springer, Bologna, pp 1–15

Edmonds B (2001) The use of models: making MABS actually work. In: Moss S, Davidsson P (eds) Multi-agent-based simulation, vol 1979, Lecture notes in artificial intelligence. Springer, Berlin, pp 15–32

Edmonds B (2005) Simulation and complexity: how they can relate. In: Feldmann V, Mühlfeld K (eds) Virtual worlds of precision: computer-based simulations in the sciences and social sciences. Lit-Verlag, Münster, pp 5–32
Edmonds B, Hales D (2003) Replication, replication and replication: some hard lessons from model alignment. J Artif Soc Soc Simulat 6(4). http://jasss.soc.surrey.ac.uk/6/4/11.html
Edmonds B, Hales D (2005) Computational simulation as theoretical experiment. J Math Sociol 29:1–24
Edwards M, Huet S, Goreaud F, Deffuant G (2003) Comparing an individual-based model of behaviour diffusion with its mean field aggregate approximation. J Artif Soc Soc Simulat 6(4). http://jasss.soc.surrey.ac.uk/6/4/9.html
Epstein JM (1999) Agent-based computational models and generative social science. Complexity 4(5):41–60
Epstein JM (2008) Why model? J Artif Soc Soc Simul 11(4). http://jasss.soc.surrey.ac.uk/11/4/12.html
Epstein JM, Axtell RL (1996) Growing artificial societies: social science from the bottom up. Brookings Institution Press/MIT Press, Cambridge, MA
Fensel D (2001) Ontologies: a silver bullet for knowledge management and electronic commerce. Springer, Berlin
Galán JM, Izquierdo LR (2005) Appearances can be deceiving: lessons learned re-implementing Axelrod's 'Evolutionary approach to norms'. J Artif Soc Soc Simulat 8(3). http://jasss.soc.surrey.ac.uk/8/3/2.html
Galán JM et al (2009) Errors and artefacts in agent-based modelling. J Artif Soc Soc Simulat 12(1). http://jasss.soc.surrey.ac.uk/12/1/1.html
Gilbert N (1999) Simulation: a new way of doing social science. Am Behav Sci 42(10):1485–1487
Gilbert N (2007) Agent-based models. Sage, London
Gilbert N, Terna P (2000) How to build and use agent-based models in social science. Mind Soc 1 (1):57–72
Gilbert N, Troitzsch KG (1999) Simulation for the social scientist. Open University Press, Buckingham
Gotts NM, Polhill JG, Adam WJ (2003) Simulation and analysis in agent-based modelling of land use change. In: Online proceedings of the first conference of the European Social Simulation Association, Groningen, 18–21 Sept 2003. http://www.uni-koblenz.de/~essa/ESSA2003/gotts_polhill_adam-rev.pdf
Gruber TR (1993) A translation approach to portable ontology specifications. Knowl Acquis 5 (2):199–220
Hare M, Deadman P (2004) Further towards a taxonomy of agent-based simulation models in environmental management. Math Comput Simulat 64(1):25–40
Hernández C (2004) Herbert A Simon, 1916–2001, y el Futuro de la Ciencia Económica. Revista Europea De Dirección y Economía De La Empresa 13(2):7–23
Heywood JG, Masuda K, Rautmann R, Solonnikov VA (eds) (1990) The Navier–Stokes equations: theory and numerical methods. In: Proceedings of a conference held at Oberwolfach, FRG, 18–24, Sept 1988 (Lecture notes in mathematics), vol 1431. Springer, Berlin
Holland JH, Miller JH (1991) Artificial adaptive agents in economic theory. Am Econ Rev 81 (2):365–370
Izquierdo LR, Polhill JG (2006) Is your model susceptible to floating point errors? J Artif Soc Soc Simulat 9(4). http://jasss.soc.surrey.ac.uk/9/4/4.html
Kleijnen JPC (1995) Verification and validation of simulation models. Eur J Oper Res 82 (1):145–162
Kleindorfer GB, O'Neill L, Ganeshan R (1998) Validation in simulation: various positions in the philosophy of science. Manage Sci 44(8):1087–1099
Klemm K, Eguíluz V, Toral R, San Miguel M (2003a) Role of dimensionality in Axelrod's model for the dissemination of culture. Phys A 327:1–5

Klemm K, Eguíluz V, Toral R, San Miguel M (2003b) Global culture: a noise-induced transition in finite systems. Phys Rev E 67(4):045101

Klemm K, Eguíluz V, Toral R, San Miguel M (2003c) Nonequilibrium transitions in complex networks: a model of social interaction. Phys Rev E 67(2):026120

Klemm K, Eguíluz V, Toral R, San Miguel M (2005) Globalization, polarization and cultural drift. J Econ Dyn Control 29(1–2):321–334

Kluver J, Stoica C (2003) Simulations of group dynamics with different models. J Artif Soc Soc Simulat 6(4). http://jasss.soc.surrey.ac.uk/6/4/8.html

Leombruni R, Richiardi M (2005) Why are economists sceptical about agent-based simulations? Phys A 355:103–109

Moss S (2001) Game theory: limitations and an alternative. J Artif Soc Soc Simulat 4(2). http://jasss.soc.surrey.ac.uk/4/2/2.html

Moss S (2002) Agent based modelling for integrated assessment. Integr Assess 3(1):63–77

Moss S, Edmonds B, Wallis S (1997) Validation and verification of computational models with multiple cognitive agents (Report no. 97–25). Centre for Policy Modelling, Manchester. http://cfpm.org/cpmrep25.html

Ostrom T (1988) Computer simulation: the third symbol system. J Exp Soc Psychol 24(5):381–392

Parunak HVD, Savit R, Riolo RL (1998) Agent-based modeling vs. equation-based modeling: a case study and users' guide. In: Sichman JS, Conte R, Gilbert N (eds) Multi-agent systems and agent-based simulation, vol 1534, Lecture notes in artificial intelligence. Springer, Berlin, pp 10–25

Pavón J, Gómez-Sanz J (2003) Agent oriented software engineering with INGENIAS. In: Marik V, Müller J, Pechoucek M (eds) Multi-agent systems and applications III, 3rd international central and eastern European conference on multi-agent systems, CEEMAS 2003 (Lecture notes in artificial intelligence), vol 2691. Springer, Berlin, pp 394–403

Pignotti E, Edwards P, Preece A, Polhill JG, Gotts NM (2005) Semantic support for computational land-use modelling. In: Proceedings of the 5th international symposium on cluster computing and the grid (CCGRID 2005). IEEE Press, Piscataway, pp 840–847

Polhill JG, Gotts NM (2006) A new approach to modelling frameworks. In: Proceedings of the first world congress on social simulation, vol 1. Kyoto, 21–25 Aug 2006, pp 215–222

Polhill JG, Izquierdo LR (2005) Lessons learned from converting the artificial stock market to interval arithmetic. J Artif Soc Soc Simulat 8(2). http://jasss.soc.surrey.ac.uk/8/2/2.html

Polhill JG, Izquierdo LR, Gotts NM (2005) The ghost in the model (and other effects of floating point arithmetic). J Artif Soc Soc Simulat 8(1). http://jasss.soc.surrey.ac.uk/8/1/5.html

Polhill JG, Izquierdo LR, Gotts NM (2006) What every agent based modeller should know about floating point arithmetic. Environ Model Software 21(3):283–309

Riolo RL, Cohen MD, Axelrod RM (2001) Evolution of cooperation without reciprocity. Nature 411:441–443

Sakoda JM (1971) The checkerboard model of social interaction. J Math Sociol 1(1):119–132

Salvi R (2002) The Navier–Stokes equation: theory and numerical methods, Lecture notes in pure and applied mathematics. Marcel Dekker, New York

Sansores C, Pavón J (2005) Agent-based simulation replication: a model driven architecture approach. In: Gelbukh AF, de Albornoz A, Terashima-Marín H (eds) Proceedings of MICAI 2005: advances in artificial intelligence, 4th Mexican international conference on artificial intelligence, Monterrey, 14–18 Nov 2005. Lecture notes in computer science, vol 3789. Springer, Berlin, pp 244–253

Sansores C, Pavón J, Gómez-Sanz J (2006) Visual modeling for complex agent-based simulation systems. In: Sichman JS, Antunes L (eds) Multi-agent-based simulation VI, international workshop, MABS 2005, Utrecht, 25 July 2005, Revised and invited papers. Lecture notes in computer science, vol 3891. Springer, Berlin, pp 174–189

Sargent RG (2003) Verification and validation of simulation models. In: Chick S, Sánchez PJ, Ferrin D, Morrice DJ (eds) Proceedings of the 2003 winter simulation conference. IEEE, Piscataway, pp 37–48

Schelling TC (1971) Dynamic models of segregation. J Math Sociol 1(2):47–186
Schelling TC (1978) Micromotives and macrobehavior. Norton, New York
Schmeiser BW (2001) Some myths and common errors in simulation experiments. In: Peters BA, Smith JS, Medeiros DJ, Rohrer MW (eds) Proceedings of the winter simulation conference, vol 1. Arlington, 9–12 Dec 2001, pp 39–46
Takadama K, Suematsu YL, Sugimoto N, Nawa NE, Shimohara K (2003) Cross-element validation in multiagent-based simulation: switching learning mechanisms in agents. J Artif Soc Soc Simulat 6(4). http://jasss.soc.surrey.ac.uk/6/4/6.html
Taylor AJ (1983) The verification of dynamic simulation models. J Oper Res Soc 34(3):233–242
Xu J, Gao Y, Madey G (2003) A docking experiment: swarm and repast for social network modeling. In: seventh annual swarm researchers conference (SwarmFest 2003), 13–15 Apr 2003, Notre Dame
Yilmaz L (2006) Validation and verification of social processes within agent-based computational organization models. Comput Math Organ Theory 12(4):283–312

# Chapter 7
# Documenting Social Simulation Models: The ODD Protocol as a Standard

**Volker Grimm, Gary Polhill, and Julia Touza**

**Why Read This Chapter?** To learn about the importance of documenting your simulation model and discover a lightweight and appropriate framework to guide you in doing this.

**Abstract** The clear documentation of simulations is important for their communication, replication, and comprehension. It is thus helpful for such documentation to follow minimum standards. The "Overview, Design concepts and Details" document protocol (ODD) is specifically designed to guide the description of individual- and agent-based simulation models (ABMs) in journal articles. Popular among ecologists, it is also increasingly used in the social simulation community. Here, we describe the protocol and give an annotated example of its use, with a view to facilitating its wider adoption and encouraging higher standards in simulation description.

## 7.1 Introduction and History

A description protocol is a framework for guiding the description of something, in this case a social simulation model. It can be thought of as a check-list of things that need to be covered and rules that should be followed when specifying the details of a simulation (in a scholarly communication). Following such a protocol means that readers can become familiar with its form and that key elements are less likely to be

V. Grimm (✉)
UFZ, Helmholtz Centre of Environmental Research – UFZ, Department of Ecological Modelling, Permoserstr. 15, 04318 Leipzig, Germany
e-mail: volker.grimm@ufz.de

G. Polhill
The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, United Kingdom

J. Touza
Applied Economics Department, University of Vigo, Vigo C.P. 36.310, Spain

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_7,

forgotten. This chapter describes a particular documentation protocol, the ODD (pronounced: "odd", or "oh dee dee") protocol.

The ODD protocol (Grimm et al. 2006, 2010; Polhill et al. 2008; Polhill 2010) is a standard layout for describing individual- and agent-based simulation models (ABMs), especially for journal articles, conference papers, and other academic literature. It consists of seven elements which can be grouped into three blocks: Overview, Design concepts, Details (hence, "ODD"; see Table 7.1). The purpose of ODD is to facilitate writing and reading of model descriptions, to better enable replication of model-based research, and to establish a set of design concepts that should be taken into account while developing an ABM. It does this in a relatively lightweight way, avoiding over-formal approaches whilst ensuring that the essentials of a simulation are explicitly described in a flexible yet appropriate manner.

Originally, ODD was formulated by ecologists, where the proportion of ABMs described using ODD is increasing fast and might cross the 50 % margin in the near future. In social simulation, the acceptance of ODD has been slower. A first test, in which three existing descriptions of land-use models were re-formulated according to ODD, demonstrated the benefits of using ODD but also revealed that some refinements were needed to make it more suitable for social simulation (Polhill et al. 2008). In 2010, an update of ODD was released (Grimm et al. 2010), which is based on users' feedback and a review of more than 50 ODD-based model descriptions in the literature. In this update, ODD itself was only slightly modified but the explanation of its elements completely rewritten, with the specific intention of making it more suitable for social simulation.

Currently in social simulation, interest in ODD is also increasing (Polhill 2010). An indicator for this is the inclusion of ODD chapters in recent reference books (this volume; Heppenstall et al. 2012). Moreover, a recent textbook of agent-based modelling uses ODD consistently (Railsback and Grimm 2012), so that the next generation of agent-based modellers is more likely to be familiar with ODD, and hence to use it themselves.

## 7.2 The Purpose of ODD

Why is ODD (or a protocol very much like it) needed? There are a number of endeavours in agent-based social simulation that are facilitated through having a common approach to describing the models that is aimed at being readable and complete[1]:

[1] Many of these endeavours have been covered in submissions to the "model-to-model" series of workshops, organised by members of the social simulation community (Hales et al. 2003; Rouchier et al. 2008. The second workshop was held as a parallel session of the ESSA 2004 conference: see http://www.insisoc.org/ESSA04/M2M2.htm).

**Table 7.1** The seven elements of the ODD protocol. Descriptions of ABMs are compiled by answering the questions linked to each element

| | | | |
|---|---|---|---|
| ***O*verview** | 1. Purpose | | What is the purpose of the model? |
| | 2. Entities, state variables, scales | | What kind of entities are in the model? Do they represent managers, voters, landowners, firms or something else? By what state variables, or attributes, are these entities characterized? What are the temporal and spatial resolutions and extents of the model? |
| | 3. Process overview, scheduling | | What entity does what, in what order? Is the order imposed or dynamic? When are state variables updated? How is time modelled: as discrete steps or as a continuum over which both continuous processes and discrete events can occur? |
| ***D*esign concepts** | 4. Design concepts | Basic principles | Which general concepts, theories or hypotheses are included in the model's design? How were they taken into account? Are they used at the level of submodels or at the system level? |
| | | Emergence | What key results are emerging from the adaptive traits, or behaviours of individuals? What results vary in complex/unpredictable ways when particular characteristics change? Are there other results that are more tightly imposed by model rules and hence less dependent on what individuals do? |
| | | Adaptation | What adaptive traits do the individuals have? What rules do they have for making decisions or changing behaviour in response to changes in themselves or their environment? Do agents seek to increase some measure of success or do they reproduce observed behaviours that they perceive as successful? |
| | | Objectives | If agents (or groups) are explicitly programmed to meet some objective, what exactly is that and how is it measured? When individuals make decisions by ranking alternatives, what criteria do they use? Note that the objective of such agents as group members may not refer to themselves but the group |
| | | Learning | May individuals change their adaptive traits over time as a consequence of their experience? If so, how? |
| | | Prediction | Prediction can be part of decision-making; if an agent's learning procedures are based on estimating future consequences of decisions, how they do this? What internal models do agents use to estimate future conditions or consequences? What 'tacit' predictions are implied in these internal model's assumptions? |
| | | Sensing | What aspects are individuals assumed to sense and consider? What aspects of which other entities can an individual perceive (e.g. displayed 'signals')? Is sensing local, through networks or global? Is the structure of networks imposed or emergent? Are the mechanisms by which agents obtain information modelled explicitly in a process or is it simply 'known'? |

(continued)

**Table 7.1** (continued)

| | | | |
|---|---|---|---|
| | | Interaction | What kinds of interactions among agents are assumed? Are there direct interactions where individuals encounter and affect others, or are interactions indirect, e.g. via competition for a mediating resource? If the interactions involve communication, how are such communications represented? |
| | | Stochasticity | What processes are modelled by assuming they are random or partly random? Is stochasticity used, for example, to reproduce variability in processes for which it is unimportant to model the actual causes of the variability, or to cause model events or behaviours to occur with a specified frequency? |
| | | Collectives | Do the individuals form or belong to aggregations that affect, and are affected by, the individuals? Such collectives can be an important intermediate level of organization. How are collectives represented – as emergent properties of the individuals or as a separate kind of entity with its own state variables and traits? |
| | | Observation | What data are collected from the ABM for testing, understanding, and analyzing it, and how are they collected? Are all output data freely used, or are only certain data sampled and used, to imitate what can be observed in an empirical study? |
| ***D*etails** | 5. Initialization | | What is the initial state of the model world, i.e., at time $t = 0$? How many entities of what type are there initially, and what are the values of their state variables (or how were they set)? Is initialization always the same, or is it varied? Are the initial values chosen arbitrarily or based on available data? |
| | 6. Input data | | Does the model use input from external sources such as data files or other models to represent processes that change over time? |
| | 7. Submodels | | What are the submodels that represent the processes listed in 'Process overview and scheduling'? What are the model parameters, their dimensions, and reference values? How were submodels designed or chosen, tested, and parameterized? |

- *Communication* is the most basic aim of anyone trying to publish their results. For agent-based modellers, this can pose a particular challenge, as our models can be complicated, with many components and submodels. As a critical mass of papers using ODD develops, so readers of agent-based modelling papers will find themselves sufficiently more familiar with papers structured using ODD than those using an arbitrary layout devised by the authors that they will find the former easier to read and understand than the latter.
- *Replication*, as we discuss later in this chapter, is a pillar of the scientific endeavour. If our model descriptions are inadequate, our results are not repeatable, and the scientific value of our work commensurately reduced. ODD helps to encourage the adequacy of descriptions by saving authors having to 'reinvent the wheel' each time they describe a model, by providing a standard layout designed to ensure that all aspects of a model needed to replicate it are included in the account.
- *Comparing models* is likely to become increasingly important as work in agent-based modelling continues. If two or more research teams produce similar models with different outcomes, comparing the models will be essential to identifying the cause of the variance in behaviour. Such comparisons will be much easier if all teams have used the same protocol to describe the models. At a conceptual level, the design concepts also enable comparison of models with greater differences and application domains.
- *Dialogue among disciplines* can be encouraged through a standard that is used by both the ecological and social simulation communities. This is especially useful for those developing coupled socio-ecosystem models (Polhill et al. 2008), which is a rapidly growing area of research (Polhill et al. 2011).

In the following, we briefly describe the rationale of ODD and how it is used, provide an example model description, and finally discuss benefits of ODD, current challenges, and its potential future development.

## 7.3 The ODD Protocol

A core principle of ODD is that first an 'Overview' of a model's purpose, structure and processes should be provided, *before* 'Details' are presented. This allows readers to quickly get a comprehensive overview of what the model is, what it does, and for what purpose it was developed. This follows the journalistic 'inverted pyramid' style of writing, where a summary is provided in the first one or two paragraphs, and progressively further detail is added on the story the further on you read (see, e.g. Wheeler 2005). It allows the reader to easily access the information they are interested in at the level of detail they need. For experienced modellers, this overview part is sufficient to understand what the model is for, to relate it to other models in the field, and to assess the overall design and complexity.

Before presenting the 'Details', ODD requires a discussion of whether, and how, ten design concepts were taken into account while designing the model. This 'Design concepts' part of ODD does not describe the model itself but the principles and rationale underlying its design. 'Design concepts' is thus not needed for model replication but for making sure that important design decisions were made consciously and that readers are fully aware of these decisions. For example, it is important to be clear about what model output is designed to emerge from the behaviour the model's entities and their interactions, and what, in contrast, is imposed by fixed rules and parameters. Ideally, key behaviours in a model emerge, whereas other elements might be imposed. If modellers are not fully aware of this difference, which is surprisingly often the case, they might impose too much so that model output is more or less hard-wired into its design, or they might get lost in a too complex model because too much emergence makes it hard to understand anything. Likewise, the design concept 'stochasticity' requires that modellers explicitly say what model processes include a stochastic component, why stochasticity was used, and how it was implemented. Note that, in contrast to the seven elements of ODD, the sequence in which design concepts are described can be changed, if needed, and design concepts that are not relevant for the model can be omitted.

The 'Details' part of ODD includes all details that are needed to re-implement the model. This includes information about the values of all model entities' state variables and attributes at the begin of a simulation ('Initialisation'), the external models or data files that are possibly used as 'Input data' describing the dynamics of one or more driving contextual or environmental variables (e.g., rainfall, market price, disturbance events), and 'Details' where the submodels representing the processes listed in 'Process overview and scheduling' are presented. Here, it is recommended for every submodel to start with the factual description of what the submodel is and then explain its rationale.

Model parameters should be presented in a table, referred to in the 'Submodels' section of ODD, including parameter name, symbol, reference value, and – if the model refers to real systems – unit, range, and references or sources for choosing parameter values. Note that the simulation experiments that were carried out to analyse the model, characterized by parameter settings, number of repeated runs, the set of observation variables used, and the statistical analyses of model output, is not part of ODD but ideally should be presented in a section 'Simulation experiments' directly following the ODD-based model description.

## 7.4 How to Use ODD

To describe an ABM using ODD, the questions listed in Table 7.1 have to be answered. The identifiers of the three blocks of ODD elements – Overview, Design concepts, Details – are not used themselves in ODD descriptions (except for

'Design concepts', which is the only element of the corresponding block). Rather, the seven elements are used as headlines in ODD-based model descriptions. For experienced ODD users, the questions in Table 7.1 are sufficient. For beginners, however, it is recommended to read the more detailed description of ODD in Grimm et al. (2010) and to use the template, which provides additional questions and examples, and which is available via download.[2]

## 7.5 An Example

In the supplementary material of Grimm et al. (2010), publications are listed which use ODD in a clear, comprehensive, and recommendable way. Many further examples are provided in the textbook by Railsback and Grimm (2012). In Grimm and Railsback (2012), Schelling's segregation model, as implemented in the model library of the software platform NetLogo (Wilensky 1999), is used as an example. Here, we demonstrate the process of model documentation using ODD by describing a model developed by Deffuant et al. (2002), which explores the emergence of extreme opinions in a population. We choose this model because it is simple but interesting and opinion dynamics models are quite well-known in the social simulation community. It is also one of the introductory examples in Gilbert (2007). The ODD for the Deffuant et al. model is interspersed with comments on the information included, with a view to providing some guidelines for those applying ODD to their own model. Clearly this is a very simple example and many models would require more extensive description. The parts of ODD are set in italics and indented to distinguish them from comments. Normally the ODD description would simply form part of the text in the main body of a paper or in an appendix.[3]

### *7.5.1 Purpose*

> *The model's purpose is to study the evolution of the distribution of opinions in a population of interacting individuals, which is under the influence of extremists' views. Specifically, it aims to answer how marginal extreme opinions can manage to become the norm in large parts of a population. The central idea of the model is that people who have more extreme opinions are more confident than people with moderate views. More confident people are, however, assumed to more easily affect the opinion of others, who are less confident.*

**Comments**: The purpose section is deliberately brief. Even for more sophisticated models than this, we would not expect to see much more text here. This would

[2] E.g. http://www.ufz.de/index.php?de=10466.

[3] It is often the case that a substantial description needs to be included in the main text so readers can get an idea of what is being discussed, but maybe a more complete description might be added in an appendix.

otherwise repeat information in the rest of the paper. However, since the ODD, to some extent, needs to stand alone and be comprehensive, the summary of the purpose is included as here.

### 7.5.2 Entities, State Variables, and Scales

*The model includes only one type of entity: individuals. They are characterised by two continuous state variables, opinion x and uncertainty u. Opinions range from −1 to 1. Individuals with an opinion very close to x = −1 or +1 are referred to as "extremists", all other individuals are "moderates". Uncertainty u defines an interval around an individuals' opinion and determines whether two individuals interact and, if they do, on the relative agreement of those two individuals which then determines how much opinion and uncertainty change in the interaction. One time step of the model represents the time in which all individuals have randomly chosen another individual and possibly interacted with it. Simulations run until the distribution of opinions becomes stationary.*

**Comments**: For larger models, this section has the potential to get quite long if written in the same style as this example, which has only one type of entity, with two state variables. Other articles have taken the approach of using tables to express this information; one table per entity, with one row per state variable associated with that entity (see, e.g. Polhill et al. 2008). Other articles have used UML class diagrams (e.g., Bithel et al. 2009), as suggested in the original ODD article (Grimm et al. 2006); however, these do not provide a means for giving any description, however brief, of each state variable. Simply listing the entities and the data types of the state variables does not provide all the information that this element of ODD should provide. This, together with the fact that UML is focused on Object-Oriented Design (which is used to implement the majority of ABMs, but by no means all: NetLogo, for example, is not an object-oriented language, and many, particularly in agent-based social simulation, use declarative programming languages), meant that the recommendation to use UML was retracted in the recent ODD update (Grimm et al. 2010).

In declarative programming languages, the entities and their state variables may not be so explicitly represented in the program code as they are in object-oriented languages. For example, this information may be implicit in the arguments to rules. However, many declarative programs have a database of knowledge that the rules operate on. This database could be used to suggest entities and state variables. For example, a Prolog program might have a database containing the assertions person (volker) and nationality(volker, german). This suggests that 'person' is an entity, and 'nationality' a state variable. (It might be reasonable to suggest in general that assertions with one argument suggest entities, and those with two, state variables.)

### 7.5.3 Process Overview and Scheduling

*In each time step each individual chooses randomly one other individual to interact with, then the relative agreement between these two agents is evaluated, and the focal individual's opinion and uncertainty are immediately updated as a result of this* opinion interaction. *Updating of state variables is thus asynchronous. After all individuals have interacted, a* convergence index *is* calculated *which captures the level of convergence in the opinions of the population; additionally, and output is updated (*e.g.*: draw histogram of the population's opinions; write each individual's opinion to a file.)*

**Comments**: This section briefly outlines the processes (or submodels) that the model runs through in every time step (ignoring initialisation), and in what order. Notice how each process is given an emphasized label, which corresponds to subsection headings in the Submodels section. Whilst the ODD protocol does not make such precise stipulations as to formatting, there should be a clear one-to-one correspondence between the brief outlines of processes here, and the details provided on each in the Submodels section.

In describing larger models than Deffuant et al.'s, it may be appropriate to simply present the process overview as a list. Many models have a simple schedule structure consisting of a repeated sequence of actions; such a list would clearly show this schedule. However, others use more complicated scheduling arrangements (e.g. dynamic scheduling). In such cases, the rules determining when new events are added to the schedule would need to be described, as well as an (unordered) list of event types, each corresponding to a subsection of 'Submodels'.

The 'schedule' in a declarative model may be even less clear, as it will depend on how the inference engine decides which rules to fire. However, declarative programs are at least asked a query to start the model, and this section would be an appropriate place to mention that. Some declarative programs also have an implied ordering to rule firing. For example, in Prolog, the rule a :- x, y, z. will, in the event that the inference engine tries to prove a, try to prove x, then y, then z. Suppose the model is started with the query ?- a. In describing the model here, it might suffice simply to summarise how x, y and z change the state of the model. Any subrules called by the inference engine trying to prove these could be given attention in the Details section.

The declarative programmer may also use language elements (such as cuts in Prolog) to manage the order of execution. In deciding which rules to describe here, a declarative modeller might focus on those changing the value of a state variable over time. The key point is that the program will do *something* to change the values of state variables over time in the course of its execution. Insofar as that can be described in a brief overview, it belongs here.

### 7.5.4 Design Concepts

***Basic principles**. – This model extends earlier stylised models on opinion dynamics, which either used only binary opinions instead of a continuous range of opinions, or where*

*interactions only depended on whether opinion segments overlapped, but not on relative agreement (for references, see Deffuant* et al. 2002*).*
***Emergence**. – The distribution of opinions in the population emerges from interactions among the individuals.*
***Sensing**. – Individuals have complete information of their interaction partner's opinion and uncertainty.*
***Interaction**. – Pairs of individuals interact if their opinion segments, $[x - u, x + u]$, overlap.*
***Stochasticity**. – The interaction between individuals is a stochastic process because interaction partners are chosen randomly.*
***Observation**. – Two plots are used for observation: the histogram of opinions, and the trajectories of each individual's opinion. Additionally, a convergence index is calculated.*

**Comments**: Note that the design concepts are only briefly addressed. This would be expected in larger models too. Note also that several design concepts have been omitted because they are not appropriate to the model. Specifically, adaptation, objectives, learning, prediction, and collectives have been left out here: individuals change their opinion after interaction, but this change is not adaptive since it is not linked to any objective; there also no collectives since all individuals act on their own. Nevertheless, most models should be able to relate to some basic principles, emergence, interactions, and observation, and most often also stochasticity. Small models might use the option of concatenating the design concepts into a single paragraph to save space.

### *7.5.5 Initialization*

*Simulations are run with 1,000 individuals, of which a specified initial proportion, $p_e$, are extremists; $p_+$ denotes the proportion of 'positive' extremists, and $p_-$ are the proportion of 'negative' extremists. Each moderate individual's initial opinion is drawn from a random uniform distribution between $-1$ and $+1$ (not inclusive). Extremists have on opinion of either $-1$ or $+1$. Initially, individuals have a uniform uncertainty, which is larger for moderates than for extremists.*

**Comments**: This explains how the simulation is set up before the main schedule starts. In other models, this might include empirical data of various kinds from, for example, surveys. The key question to ask here, particularly given the potential for confusion with the next section ('input data'), is whether the data are used *only* to provide a value for a state variable before the schedule runs.

### *7.5.6 Input Data*

*The model does not include any input of external data.*

**Comments**: These are time-series data used to 'drive' the model. Some of these data *may* specify values for variables at time 0 (i.e. during initialisation); however, if a data series specifies values for any time step other than during initialisation, then it is input data rather than initialisation. It is also important not to confuse 'Input data' with parameter values.

### 7.5.7 Submodels

*All model parameters are listed in the following table.*

| Parameter | Description |
|---|---|
| N | Number of individuals in population |
| U | Initial uncertainty of moderate individuals |
| $\mu$ | Speed of opinion dynamics |
| $p_e$ | Initial proportion of extremists |
| $p_+$ | Initial proportion of positive extremists |
| $p_-$ | Initial proportion of negative extremists |
| $u_e$ | Initial uncertainty of extremists |

***Opinion interaction**. – This is run for an agent j, whose 'opinion segment' $s_j$ is defined in terms of its opinion $x_j$ and uncertainty $u_j$ as:*

$$s_j = \left[x_j - u_j, x_j + u_j\right]$$

*The length of the opinion segment is $2u_j$ and characterizes an individual's overall uncertainty.*
*In opinion interaction, agent j (the influenced, focal, or 'calling' individual) is paired with a randomly chosen agent, i, the influencing individual. The 'overlap' of their opinion segments, $h_{ij}$, is then computed as:*

$$h_{ij} = min\left(x_i + u_i, x_j + u_j\right) - max\left(x_i - u_i, x_j - u_j\right)$$

*This overlap determines whether an opinion interaction will take place or not: Agent j will change its opinion if $h_{ij} > u_i$, which means that overlap of opinions is higher than the uncertainty of the influencing agent (see Fig. 7.1).*

*For opinion interactions, the relative agreement of the two agents' opinions, RA, is calculated by dividing the overlap of their opinion segments ($h_{ij}$) minus the length of the non-overlapping part of influencing individual's opinion segment, ($2u_i$ – $h_{ij}$), and this difference divided by agent i's opinion segment length, $2u_i$ (Fig.7.1 depicts these terms graphically):*

$$RA = \left(h_{ij} - \left(2u_i - h_{ij}\right)\right)/2u_i = 2\left(h_{ij} - u_i\right)/2u_i = \left(h_{ij}/u_i\right) - 1$$

*The opinion and uncertainty of agent j are then updated as follows:*

$$x_j = x_j + \mu RA\left(x_i + x_j\right)$$

$$u_j = u_j + \mu RA\left(u_i + u_j\right)$$

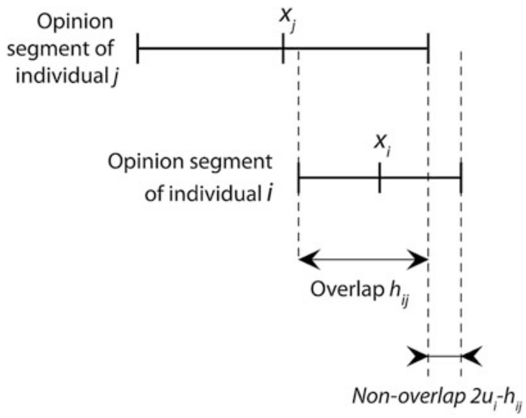



**Fig. 7.1** Visualisation of the individual's opinions, uncertainties, and overlap in opinions in the model of Deffuant et al. (2002)

*Thus, the new values are determined by the old values and the sum of the old values of both interacting individuals multiplied by the relative agreement, RA, and by parameter $\mu$, which determines how fast opinions change.*

*The main features of this interaction model are, according to Deffuant* et al. *(2002):*

- *Individuals not only influence each other's opinions but also each other's uncertainties.*
- *Confident agents, who have low uncertainty, are more influential. This reflects the common observation that confident people more easily convince more uncertain people than the other way round – under the conditions that their opinions are not too different at the beginning.*

***Calculate convergence index****. – This index, y, is used as a summary model output for sensitivity analysis and an exploration of the model's parameter space. It is defined as:*

$$y = q_+ + q_-$$

*where $q_+$ and $q_-$ are the proportions of initially moderate agents which become extremists in the positive extreme or negative extreme, respectively. If after reaching the steady state none of the initially moderate agents became extremist the index would take a value of zero. If half of them become positive extremists and the other half becomes negative extremists, the index would be 0.5. Finally, if all the initially moderate agents converge to only one extreme, the index would be one. Note that for calculating y, "positive" or "negative" extreme has to be defined* via *an interval close the extreme, with a width of, for example, 0.15.*

**Comments**: Here, details on the two processes described in Sect. 7.3 are provided, in sufficient depth to enable replication, i.e. *opinion interaction* and *calculate convergence index*. Note how these names match with those used in the process overview in Sect. 7.3.

Authors describing larger models may find journal editors protesting at the length of the ODD if all submodels are described in the detail required. There are various ways such constraints can be handled. One is to include the submodels in an appendix or supplementary material to the paper. Another is to provide them as a technical report accessible separately (e.g. on a website), and referred to in the text. If space is not too limited, a summary of each submodel could be provided in the main text, longer than the brief description in the process overview, but shorter than the full detail; the latter being provided separately. For very large models, or where space is highly constrained, there may be little room for much more than the three Overview sections in the journal article; again, making the full ODD available separately is a possible solution. Nevertheless, excluding the 'Submodels' element entirely from the main text should be avoided because this would mean to ask readers to accept, in the main text of the article, the model as a black box. Description of the most important processes should therefore be included also in the main text.

## 7.6 Discussion

Since the example model by Deffuant et al. (2002) is very simple, using ODD here comes with the cost of making the model description longer than the original one, through requiring the ODD labels. The original model is actually relatively clear and easy to replicate (which might partly explain this model's success). However, easy replication is much more the exception than the rule (Hales et al. 2003; Rouchier et al. 2008), and the more complex an ABM, the higher the risk that not all information is provided for unambiguous replication.

ODD facilitates writing comprehensive and clear documentations of ABMs. This does not only facilitate replication, it also makes writing and reading model documentations easier. Modellers no longer have to come up with their own format for describing their model, and readers know, once they are familiar with the structure of ODD, *exactly* where to look for what kind of information.

Whether or not to use ODD as a standard format for model descriptions might look like a rather technical question, but it has fundamental consequences, which go far beyond the issue of replication. Once ODD is used as a standard, it will be become much easier to compare different models addressing similar questions. Even now, ODD can be used to review models in a certain field, by rewriting existing model descriptions according to ODD (Grimm et al. 2010). Building blocks of existing models, in particular specific submodels, which seem to be useful in general, will be much easier to identify and re-use in new models.

Most importantly, however, using ODD affects the way we design and formulate ABMs in the first place. After having used ODD for documenting two or three models, you start formulating ABMs by answering the ODD questions: What 'things', or entities, do I need to represent in my model? What state variables and behavioural attributes do I need to characterize these entities? What processes do I want to represent explicitly, and how should they be scheduled? What are the spatial and temporal extent and resolution of my model, and why? What do I want to impose, and what to let emerge? What kind of interactions does the model include? For what purposes should I include stochasticity? How should the model world be initialized, what kinds of input data do I need, and how should I, in detail, formulate my submodels?

These questions do not impose any specific structure on simulation models, but they provide a clear checklist for both model developers and users. This helps avoiding “ad hoc-ery” in model design (Heine et al. 2005). Modellers can also more easily adopt designs of existing models and don’t have to start from scratch all the time, as in most current social simulation models.

Criticisms of ODD include Amouroux et al. (2010), who, acknowledging its merits, find the protocol ambiguous and insufficiently specified to enable replication. This article pertained to the Grimm et al. (2006) first description of ODD. The update in Grimm et al. (2010) endeavoured to address issues such as these. However, the success of the latter article in so doing, and indeed any future revisions of ODD, can only be measured by comparing replication efforts based on ODD descriptions with those not conforming to any protocol – the norm prior to 2006 when ODD was first published. As suggested above, the record for articles not using ODD has not been particularly good: Rouchier et al. (2008) observe in their editorial to a special section of JASSS on the third Model-2-Model workshop that several researchers attempting replications have to approach the authors of the original articles to disambiguate model specifications. If the models were adequately described in the original articles, this should not be necessary.

Polhill et al. (2008) also observed that those used to object-oriented designs for modelling will find the separation of what will for them effectively amount to instance variables and methods (state variables and processes respectively) counter-intuitive, if indeed not utterly opposed to encapsulation: one of the key principles of object orientation. For ODD, however, it is the reader who is important rather than programming principles intended to facilitate modularity and code reuse. It is also important that, as a documentation protocol, ODD does not tie itself to any particular ABM implementation environment. From the perspective of the human reader, it is illogical (to us at least) to discuss processes before being informed what it is the processes are operating on. Encapsulation is about hiding information; ODD has quite the opposite intention.

The main issue with ODD in social simulation circles as opposed to ecology, from which it originally grew, pertains to its use with declarative modelling environments. This matter has been raised in Polhill et al. (2008), and acknowledged in Grimm et al. (2010). Here we have tried to go further towards illustrating how a declarative modeller might prepare a description of their model that conforms to ODD. However, until researchers using declarative environments attempt to use ODD when writing an article, and feedback on their findings, this matter cannot be properly addressed.

Certainly, ODD is not the silver bullet regarding standards for documenting ABMs. Nevertheless, even at the current stage its benefits by far outweigh its limitations, and using it more widely is an important condition for further developments. Still, since ODD is a verbal format, not all ambiguities can be prevented. Whilst a more formal approach using, for example XML or UML (e.g. Triebig and Klügl 2010, and for ABMs of land use/cover change, the MRPOTATOHEAD framework – Livermore 2010; Parker et al. 2008) might address such ambiguities, we consider it important that written, natural language formulations of ABMs exist (Grimm and Railsback 2005). This is the only way to make modelling, as a scientific activity, independent of technical aspects of mark-up or programming languages and operating systems. Further, verbal descriptions force us to *think* about a model, to try to understand what it is, what it does, and why it was designed in that way and not another (J. Everaars, *pers. comm.*). We doubt that a 'technical' standard for documenting ABMs – one that can be read by compilers or interpreters, would ever initiate and require this critical thinking about a model.

Nevertheless, it is already straightforward to translate ODD model description to NetLogo programs because much of the way models are written in NetLogo corresponds to the structure of ODD: the declaration of 'Entities, state variables, and scales' is done via NetLogo's globals, turtles-own, and patches-own primitives, 'Initialization' is done via the setup procedure, 'Process overview and scheduling' corresponds to the go procedure, 'Details' are implemented as NetLogo procedures, and 'Design concepts' can be included, (as indeed can the entire ODD model description), on the 'Information' tab of NetLogo's user interface.

## 7.7 Conclusion

Clearly describing simulations well, so that other researchers can understand a simulation is important for the scientific development and use of complex simulations. It can help in: the assessment and comprehension of simulation results by readers; replicating simulations for checking and analysis by other researchers; transferring knowledge embedded within simulations from one domain to another; and allowing simulations to be better compared. It is thus an important factor for making the use of simulations more rigorous and useful. A protocol such as ODD is useful in standardising such descriptions and encouraging minimum standards. As the field of social simulation matures it is highly likely that the use of a protocol such as ODD will become standard practice.

The investment in learning and using ODD is minimal but the benefits, both for its user and the scientific community, can be huge. We therefore recommend learning and testing ODD by re-writing the model description of an existing, moderately complex ABM, and, in particular, using ODD to formulate and document the next ABM you are going to develop.

**Acknowledgements** We are grateful to Bruce Edmonds for inviting us to contribute this chapter, and for his helpful comments and suggested amendments to earlier drafts. Gary Polhill's contribution was funded by the Scottish Government.

## Further Reading

Railsback and Grimm (2012) is a textbook which introduces agent-based modelling with examples described using ODD. The OpenABM website (http://openabm.org) is a portal specifically designed to facilitate the dissemination of simulation code and descriptions of these using the ODD protocol. The original reference document for ODD is (Grimm et al. 2006) with the most recent update being (Grimm et al. 2010). Polhill (2010) is an overview of the 2010 update of ODD written specifically with the social simulation community in mind.

## References

Amouroux E, Gaudou B, Desvaux S, Drogoul A (2010) O.D.D.: a promising but incomplete formalism for individual-based model specification. In: Ho TB, Zuckerman DN, Kuonen P, Demaille A, Kutsche R-D (eds) 2010 IEEE-RIVF international conference on computing and communication technologies: research, innovation and vision for the future, Vietnam National University, Hanoi, 1–4 Nov 2010

Bithell M, Brasington J (2009) Coupling agent-based models of subsistence farming with individual-based forest models and dynamic models of water distribution. Environ Model Software 24:173–190

Deffuant G, Amblard F, Weisbuch G, Faure T (2002) How can extremism prevail? A study based on the relative agreement interaction model. J Artif Soc Soc Simulat 5(4):1. http://jasss.soc.surrey.ac.uk/5/4/1.html

Gilbert N (2007) Agent-based models. Sage, London

Grimm V, Railsback SF (2005) Individual-based modeling and ecology. Princeton University Press, Princeton

Grimm V, Railsback SF (2012) Designing, formulating, and communicating agent-based models. In: Heppenstall A, Crooks A, See LM, Batty M (eds) Agent-based models of geographical systems. Springer, Berlin, pp 361–377

Grimm V et al (2006) A standard protocol for describing individual-based and agent-based models. Ecol Model 198:115–126

Grimm V et al (2010) The ODD protocol: a review and first update. Ecol Model 221:2760–2768

Hales D, Rouchier J, Edmonds B (2003) Model-to-model analysis. J Artif Soc Soc Simulat 6(4):5. http://jasss.soc.surrey.ac.uk/6/4/5.html

Heine B-O, Meyer M, Strangfeld O (2005) Stylised facts and the contribution of simulation to the economic analysis of budgeting. J Artif Soc Soc Simulat 8(4):4. http://jasss.soc.surrey.ac.uk/8/4/4.html

Heppenstall A, Crooks A, See LM, Batty M (eds) (2012) Agent-based models of geographical systems. Springer, Berlin

Livermore M (2010) MR POTATOHEAD framework: a software tool for collaborative land-use change modeling. In: Swayne DA, Yang W, Voinov AA, Rizzoli A, Filatova T (eds) International Environmental Modelling And Software Society (iEMSs) 2010 international congress on environmental modelling and software: modelling for environment's Sake, Fifth Biennial Meeting, Ottawa. http://www.iemss.org/iemss2010/index.php?n=Main,Proceedings

Parker DC et al (2008) Case studies, cross-site comparisons, and the challenge of generalization: comparing agent-based models of land-use change in frontier regions. J Land Use Sci 3(1):41–72

Polhill JG (2010) ODD updated. J Artif Soc Soc Simulat 13(4):9. http://jasss.soc.surrey.ac.uk/13/4/9.html

Polhill JG, Parker D, Brown D, Grimm V (2008) Using the ODD protocol for describing three agent-based social simulation models of land use change. J Artif Soc Soc Simulat 11(2):3. http://jasss.soc.surrey.ac.uk/11/2/3.html

Polhill JG, Gimona A, Aspinall RJ (2011) Agent-based modelling of land use effects on ecosystem processes and services. J Land Use Sci 6(2–3):75–81

Railsback SF, Grimm V (2012) Agent-based and individual-based modeling: a practical introduction. Princeton University Press, Princeton

Rouchier J, Cioffi-Revilla C, Polhill JG, Takadama K (2008) Progress in model-to-model analysis. J Artif Soc Soc Simulat 11(2):8. http://jasss.soc.surrey.ac.uk/11/2/8.html

Triebig C, Klügl F (2010) Elements of a documentation framework for agent-based simulation. Cybern Syst 40(5):441–474

Wheeler S (2005) Beyond the inverted pyramid: developing news-writing skills. In: Keeble R (ed) Print journalism: a critical introduction. Routledge, Abingdon, pp 84–93

Wilensky U (1999) NetLogo. http://ccl.northwestern.edu/netlogo

# Chapter 8
# Validating Simulations

**Nuno David**

**Why Read This Chapter?** To help you decide how to check your simulation – both against its antecedent conceptual models (verification) and external standards such as data (validation) – and in this way help you to establish the credibility of your simulation. In order to do this the chapter will point out the nature of these processes, including the variety of ways in which people seek to achieve them.

**Abstract** Verification and validation are two important aspects of model building. Verification and validation compare models with observations and descriptions of the problem modelled, which may include other models that have been verified and validated to some level. However, the use of simulation for modelling social complexity is very diverse. Often, verification and validation do not refer to an explicit stage in the simulation development process, but to the modelling process itself, according to good practices and in a way that grants credibility to using the simulation for a specific purpose. One cannot consider verification and validation without considering the purpose of the simulation. This chapter deals with a comprehensive outline of methodological perspectives and practical uses of verification and validation. The problem of verification and validation is tackled in three main topics: (1) the meaning of the terms verification and validation in the context of simulating social complexity; (2) methods and techniques related to verification, including static and dynamic methods, good programming practices, defensive programming, and replication for model alignment; and (3) types and techniques of validation as well as their relationship to different modelling strategies.

N. David (✉)
ISCTE – Lisbon University Institute, Av. das Forças Armadas, Lisbon 1649-026, Portugal
e-mail: Nuno.David@iscte.pt

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_8,

## 8.1 Introduction

The terms verification and validation (V&V) are commonly used in science but their meaning is often controversial, both in the natural and the social sciences. The purpose of this chapter is not to describe any general theory of model V&V. A general theory of that kind does not exist.

Besides the epistemological underpinnings of the terms, their use in simulation has a pragmatic nature. In disciplines that make use of computerised models, the role of V&V is related to the need of evaluating models along the simulation development process. Basically, the very idea of V&V is comparing models with observations and descriptions of the problem modelled, and this may include other models that have been verified and validated to some level. This chapter describes methodological perspectives and practical uses of the terms as well as different strategies and techniques to verify and validate models of social complexity, mostly in social simulation.

The use of simulation for modelling social complexity is very diverse. Often, V&V do not refer to an explicit stage in the simulation development process, but to the modelling process itself according to good practices and in a way that grants credibility to using the simulation for a specific purpose. Normally, the purpose is dependent on different strategies or dimensions, along which simulations can be characterized, related to different kinds of claims intended by the modeller, such as theoretical claims, empirical claims or subjunctive theoretical claims. The term subjunctive is used when simulations are used for talking about scenarios in possible worlds, such as describing 'what *would* happen if something were the case'. There cannot be V&V without considering the purpose of the simulation.

In the next section of the chapter, I will deal with the meaning of the terms V&V in the context of the simulation development process. While the terms are often used with the same meanings, or even interchangeably, practical reasons exist for distinguishing between them. Whereas verification concerns the evaluation of the implementation of the model in terms of the researchers' intentions, validation refers to the evaluation of the credibility of the model as a representation of the subject modelled. In Sect. 8.3, methods and techniques related to verification are described. Subsequently, in Sect. 8.4, validation types and techniques are described, as well as their relationship to different modelling strategies.

## 8.2 The Simulation Development Process

Several chains of intermediate models are developed before obtaining a satisfactory verified and validated model. What does it mean to verify and validate a model in social simulation? Is there a fundamental difference between verifying and validating models? The purpose of this section is to define the role of V&V within the scope of the simulation development process.

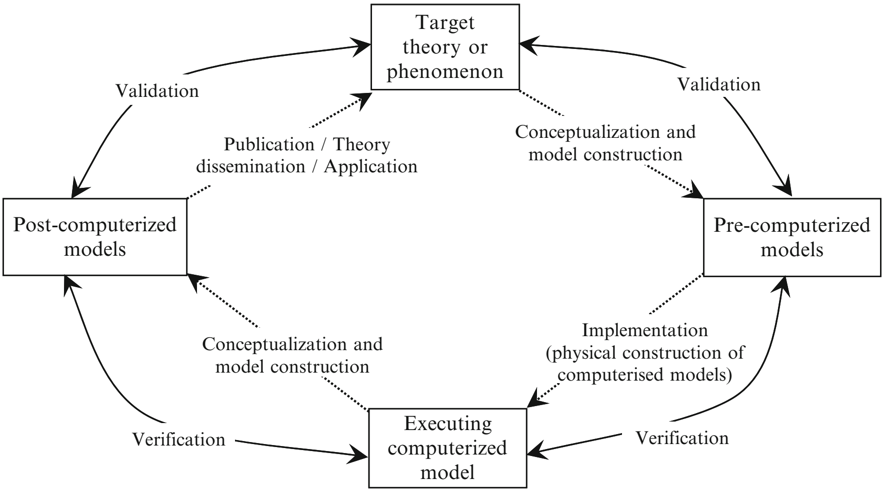



**Fig. 8.1** Verification and validation related to the model development process (David 2009)

The most common definitions of V&V are imported from computer science, as well as from technical and numerical simulation,[1] having intended distinct – although often overlapping – meanings. The reason for distinguishing the terms is the need to determine the suitability of certain models for representing two distinct subjects of inquiry. This is represented in Fig. 8.1, in which V&V are related to a simplified model development process. Two conceptual models mediate between two subjects of inquiry. The latter are (1) the target theory or phenomenon and (2) the executing computerised model. The conceptual model on the right, which is designated here as the *pre-computerised model*, is basically a representation in the minds and writing of the researchers, which presumably represents the target. This model must be implemented as a *computerised executable model*, by going through a number of intermediate models such as formal specification or textual programs written in high-level programming languages.

The analysis of the executing model gives rise to one or more conceptual models on the left, which are designated here as the *post-computerised* models. They are constructed based on the output of the computerised model, often with the aid of statistical packages, graphing and visualisation. The whole construction process results in categories of description that may not have been used for describing the pre-computerised model. This is the so-called idea of *emergence*, when interactions among objects specified through pre-computerised models at some level of description give rise to different categories of objects at macro levels of description observed in the executing model, which are later specified through post-computerised models.

[1] Numerical simulation refers to simulation for finding solutions to mathematical models, normally for cases in which mathematics does not provide analytical solutions. Technical simulation stands for simulation with numerical models in computational sciences and engineering.

As an example consider the culture dissemination model of Axelrod (1997a) which has a goal of analysing the phenomena of social influence. At a micro-level of description, a pre-computerised model defines the concept of *actors* distributed on a grid, the concept of *culture* of each actor as a set of five features and the *interaction mechanisms* specified with a bit-flipping schema, in which the probability of interaction between two actors is set proportionately to measure the similarity between two cultures.

The executing model is then explored and other categories of objects resulting from the interaction of individual cultures may be defined, associated with macro properties of interest and conditions in which they form, such as the concepts of *regions* and *zones* on the grid. A great deal of the simulation proposed by Axelrod concerns investigating properties of *regions* and *zones* in the context of the conceptual, post-computerised, model proposed, such as the relation between the size of a *region* formed and the number of features per *individual culture*. These concepts are interpreted in relation to the target social phenomena of social influence.

I will now situate the role of V&V in the modelling process of social simulation.

### *8.2.1 What Does It Mean to Verify a Computerised Model?*

Computerised model *verification* is defined as checking the adequacy among conceptual models and computerised models (see also Chap. 6 in this volume, Galán et al. 2013). Consider the lower quadrants of Fig. 8.1. They are concerned with ensuring that the pre-computerised model has been implemented *adequately* as an executable computerised model, according to the researchers' intentions in the parameter range considered, and also that the post-computerised model *adequately* represents the executing model in the parameter range considered.[2]

At this point you might question the meaning of *adequately*. There is no formal definition of adequately. Actually, it is a rather difficult epistemological problem, considered in the series of EPOS meetings about epistemological perspectives on simulation (see Frank and Troitzsch 2005; Squazzoni 2009; David et al. 2010). A practical and minimal definition could be the following: adequateness means that the relationship between inputs and outputs of the computerised model is consistent with the semantics of both the pre- and post-computerised models, in accordance with the researcher's intentions. However, the outcomes of computer programs in social simulation are often unintended or not known a priori and thus the verification process requires more than checking that the executing model does what it was planned to do. The goal of the whole exercise is to assess logical links within, as well as between, the pre- and the post-computerised models. This requires assessing whether the post-computerised model – while expressing concepts that the pre-computerised model does not express – is consistent with the latter. From a methodological point of view

[2] Verification in the left quadrant of Fig. 8.1 is sometimes known as "internal validation".

this is a complicated question, but from a practical perspective one might operationally define the verification problem with the following procedures:

(a) For some pre-computerised model definable as a set of input/output pairs *in a specified parameter range*, the corresponding executing model is *verified for the range considered* if the corresponding post-computerised model expresses the same set of inputs/outputs for the range considered.
(b) For some pre-computerised model defined according to the researcher and/or stakeholders' intentions *in a specified parameter range*, the corresponding executing model is *verified for the range considered* if the corresponding post-computerised model complies with the researchers and/or stakeholders' intentions for the range considered.

Note that both procedures limit the verification problem to a clearly defined parameter range. The first option is possible when extensive quantitative data is available from the target with which to test the executing model. This is normally not the case and the verification problem often amounts to the last option. This is possible since the aim of verification is to assess the appropriateness of the logical links that may be established between micro-levels of description specified in the pre-computerised model and macro-levels of description specified through post-computerised models, which should be amenable to evaluation by researchers and stakeholders. Nevertheless, as we will discuss further, a computerised model should only be qualified as verified with reasonable confidence if it has been successfully subjected to a procedure known in software engineering as *N-version programming*. A synonym commonly used in social simulation consists of *replicating implementations* for model alignment. This is described in Sect. 8.2.4.

It is important to observe that, unlike many areas of computer science and formal logics, the verification process should not be understood as a mere process of evaluating the correctness of formal inference steps. The coherence among conceptual and computerised models, in the light of the researchers' and stakeholders' intentions, is, to a large extent, narratively evaluated. Were we to envisage computers and programming languages as mere formal machines then computers and programming languages would have limited expressiveness for representing inferences upon semantically rich descriptions of social processes. After defining the meaning of validation in the following section, this will be a point where the meanings of verifying and validating a simulation model will overlap, and it may not be trivial to distinguish between the two.[3]

[3] From a technical point of view, in classical computer theory, verification amounts to ascertaining the validity of certain output as a function of given input, regardless of any interpretation given in terms of any theory or any phenomenon not strictly computational – it is pure inference in a closed world. But this would require us to assume that social processes are computational in a Church-Turing sense, which seems difficult to conceive. For an elaboration on this see (David et al. 2007).

### 8.2.2 What Does It Mean to Validate a Model?

Model *validation* is defined as ensuring that both conceptual and computerised models are adequate representations of the target. The term "adequate" in this sense may stand for a number of epistemological perspectives. From a practical point of view we could assess whether the outputs of the simulation are close enough to empirical data.

We could also assess various aspects of the simulation, such as if the mechanisms specified in the simulation are well accepted by stakeholders involved in a participative-based approach. In Sect. 8.4 we will describe the general idea of validation as the process that assesses whether the pre-computerised models – put forward as models of social complexity– can be demonstrated to represent aspects of social behaviour and interaction able to give rise to post-computerised models that are, at some given level, consistent with the subjacent theories or similar to real data.

Given the model development process described, is there any practical difference between verifying and validating simulations? Rather than being a sharp difference in kind it is a distinction that results from the computational method. Whereas *verification* concerns the assessment of the logical inferences that are established between micro and macro concepts with close reference to the computerised model, *validation* concerns the evaluation of such inferences and concepts with a closer reference to the target.

In paraphrasing Axelrod (1997b), at first sight, we could say that the problem is whether an unexpected result is a reflection of the computerised model, due to a mistake in the implementation of the pre-computerised model, or is a surprising consequence of the pre-computerised model itself. Unfortunately, the problem is more complicated than that. In many cases mistakes in the code may not be qualified simply as mistakes, but only as one interpretation among many other possible interpretations for implementing a conceptual model. Nevertheless, from a practical viewpoint there are still good reasons to make the distinction between V&V. A number of established practices exist for the corresponding quadrants of Fig. 8.1. We will address these in the following sections.

## 8.3 Verification Methods and Techniques

The need to transform conceptual models into computerised models and back to conceptual models is one of the substratums of simulation. This process is illustrated in Fig. 8.2. Several conceptual models are constructed and specified in order to obtain a proper implementation as a computerised model. The intended computerised model is specified as a computer program in a readable form with a high-level language. The compilation of the program into a low-level program, followed by its execution with a set of inputs, results in outputs described with data and visualization models. Given the use of pseudo-random generators, the simulation will be run many times with the

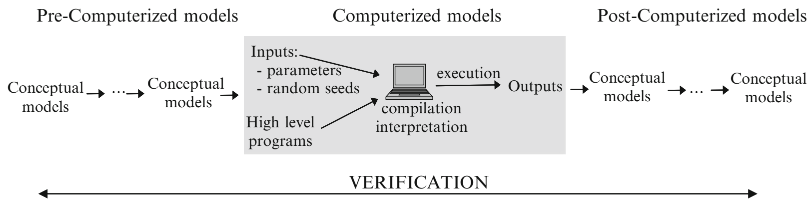



**Fig. 8.2** Implementation of pre-computerised models and construction of post-computerised models

same inputs. At any rate, the complexity of social theories and phenomena implies that the simulation must be run many times in order to observe its behaviour in a variety of conditions. A post-computerised model is built, i.e. a conceptual model describing the simulation behavior, possibly incorporating new concepts and knowledge not captured with the pre-computerised model.

### *8.3.1 Static and Dynamic Methods*

In simulation, verification techniques involve showing that the process of translation among conceptual and computerised models is carried out according to the researchers' intentions. Computer science, mostly in the field of software engineering, has developed numerous approaches and techniques for writing programs and verifying their execution in computers. Most of these are not particular to simulation and not all are well suited to social simulation. The techniques and terminology that we describe in this section are based on Sommerville (1995) and Sargent (1999), and are adapted to social simulation whenever appropriate. There are two basic approaches for verification: *static and dynamic methods*.

Dynamic methods involve exercising the computerised model with sets of inputs and the results obtained are used to determine whether the computerised model has been programmed appropriately. A possible *simulation failure* occurs when the simulation is executing and does not behave as expected. Whether the unexpected output is indeed a failure or a legitimate consequence of the pre-computerised model concerns the use of *static methods*. Static methods involve showing that the computerised model appropriately implements the pre-computerised models without *software faults*, i.e. without programming errors whereby the computerised model does not conform to the researchers' intentions.

Software faults are static and their existence may be inferred from simulation failures and further inspections to the high-level program code. A software fault causes a simulation failure when the faulty code is executed with a set of inputs that expose the simulation fault (cf. Sommerville 1995).

*We can never conclusively demonstrate that a simulation is fault free*, but we can increase our confidence in a program *within the range of the parameter space tested* by adopting good programming and testing practices in order to minimize faults.

Good practices common to any kind of programming include commenting your code and testing the program with parameter values whose outputs are known or with values that are at the extremes of the legal values.

Nevertheless, not all programming and testing techniques in computer science can be applied to social simulation. For the most part, simulation is used to explore concepts that are not anticipated during the specification of pre-computerised models. Virtually, the software notion of "functional requirement" does not exist in social simulation. If we are dealing with complex social models, it is virtually impossible to enumerate a priori an exhaustive list of requirements that a program should satisfy (David et al. 2003). To the extent that the researcher does not know what to expect, program testing is not enough to show that the program properly reflects the intended model.

In any case, the verifiability of a simulation is influenced by the process used to develop that simulation. The adoption of good programming practices for designing and implementing the model is fundamental. Defensive programming methodologies, like contract-based programming with use of assertions, are well suited for the explorative nature of simulation. Defensive programming consists of including extra, redundant, code to test for errors and to aid in general debugging. In addition, the need to produce several implementations of the same conceptual model is increasingly playing a role. These techniques are described below along four major topics: good programming practices, defensive programming, replication, and a very brief reference to participative-based methods.

### *8.3.2 Good Programming Practices*

Good programming in social simulation includes several approaches, techniques or mechanisms to improve verifiability in general and, specifically, to decrease the number of expected simulation faults, to ease debugging and to improve code readability as well as faster and more flexible development. Available techniques and mechanisms include modularity, encapsulation, high-level memory management and, more generally, software reuse. Most object-oriented programming languages include built-in mechanisms that simplify their use. The following deserve particular attention.

#### 8.3.2.1 Modularity and Encapsulation

Programs should be written and verified in modules or subprograms.[4] In the words of Kleijen (1995), this is a kind of "divide and conquer" process, where one verifies the whole simulation module by module. Modular programming simplifies the

[4] We consider "modules" and "sub-programs" as synonymous.

location of faults in the program. A module is any component that provides one or more services to other modules and which separates its public interface from its implementation details. The public interface of a module concerns the specification for how other modules may request services from it. The implementation concerns particular design decisions hidden within the module and not accessible to the other modules, which are more liable to faults or design changes along the simulation development process. If the implementation of a module is changed this should not affect how other modules request services from it. A typical kind of module is a *class* in object-oriented programming.

A class is a model for object instantiation. Objects are independent, loosely coupled entities in which the particular implementation of the object state (information representation) and of the services provided by the object (information processing) should not affect the public interface through which other objects request services from it. An object comprises a set of private attributes – which define its state – and a set of public operations that check or act on those attributes – which define its interface. Insofar as the public interface remains the same along the development process, the internal data representation of the class may be changed along the way without affecting the way other classes request services from it. The grouping of related classes within a single file is another kind of module, usually called a physical module. Yet another kind is the concept of *package*, defined as a group of related files, such as in the Java programming language. The public interface of a package is the set of all public services available in all files. The package design details most likely to change are hidden behind the interface, possibly implemented with the aid of other private classes in the package.

It is essentially up to the programmer to decide which data structures, classes and files should be grouped together. According to Sommerville (1995, pp. 218–219), the *cohesion* of a module is a measure of the closeness of the relationship between its components. Conversely, the *coupling* among modules measures the strength of interconnections among them. As a rule, modules are tightly coupled if they make use of shared variables or if they interchange many types of control information. A good design calls for *high cohesion within* a module and *low coupling among* modules. For instance, an agent implemented as a group of classes should have its public interface clearly defined and be as independent as possible from the group of classes that implement the interaction environment of the agents, such as a bi-dimensional torus grid. Whether the grid is implemented as a torus or as a network should not imply future changes to the implementation of the agent architecture, which in any case would request the same services from the grid, such as moving to the left or right. Modularity encourages the production of readable and testable simulations. Moreover, insofar as agent-based modelling can be easily mapped to the programming of interacting modules, the adoption of bottom-up design and testing approaches is a natural way to proceed. Small modules at the lower levels in the hierarchy are tested first and the other modules are worked up the hierarchy until the whole simulation is tested.

#### 8.3.2.2 High-Level Memory Management

The use of pointers and pointer arithmetic in a program are low-level constructs available in some high-level languages that manipulate variables referring directly to the machine memory. They are inappropriate in social simulation programs because they allow memory leaks and foment aliasing, which makes programs prone to bugs. A memory leak occurs when a program fails to release memory no longer needed, which may have the effect of consuming more memory than necessary and decreasing efficiency. We say there is aliasing when an object is changed through a reference variable with unwanted indirect effects on another variable. Both leaks and aliasing make programs harder to understand and faults harder to find. Whenever possible, high-level languages with pointer arithmetic should be avoided. In programming with object-oriented languages, the use of built-in mechanisms for automatic memory management, responsible for managing the low-level lifecycle of objects in memory, should be preferred. Whereas it is up to the programmer to determine the points where objects are created in the program, mechanisms of memory management free the programmer from the trouble of deleting the objects from memory at the points where they become useless. Fortunately, most programming languages used in current agent-based simulation toolkits are examples of high-level languages that have built-in mechanisms for automatic memory management.

#### 8.3.2.3 Software Reuse

We sometimes tend to prefer rewriting modules insofar as we believe that our modules are better programmed than others. However, software modules that are used and tested in a variety of different situations have fewer faults than modules developed for the purposes of a single simulation. A good programmer should not assume that all modules should be implemented especially for the simulation being developed. Reusable modules that have been subjected to previous use and verification should be preferred to those built from scratch. The more a simulation is based on reusable modules the fewer modules need to be specified, implemented and verified.

Software reuse is used in social simulation as more and more developing frameworks become available. Module reuse in simulation is basically a type of model embedding. The use of agent-based toolkits as special-purpose extensions to standard high-level languages based on Java, Objective C, or SmallTalk provides developing standards and faster development, making simulations more comparable to each other. Software reuse in computer science and engineering can be considered at a number of different levels (Sommerville 1995, p. 397). Likewise in social simulation we may identify different levels from the strongest to the weakest degrees of dependency on the specifics of simulation platforms and toolkits. In this sense, strong reuse means constraints on the simulation to existing models and weak reuse lets the researcher develop models more freely.

### Model Architecture Level

The control subsystem of a simulation, such as a discrete-time scheduling mechanism, may be reused. In most agent-based simulation toolkits, scheduling consists of setting up method calls on objects to occur at certain times. Scheduling mechanisms are thus part of a reusable agent skeleton template, extended according to the specifics of each agent model. Other examples include whole collections of reusable objects and mechanisms, such as interaction environments that act as agent containers and define the interactive relationship of agents relative to each other, such as a network model or a torus grid. Yet other examples may include entire canonical simulation models, which are used as components of another extended, more sophisticated, simulation model.

### Module or Object Level

Components representing a set of functions may be reused, such as class libraries for generating sequence graphs; histograms or plots for visualization models; statistical packages for data analysis; and widely tested pseudo-random number generators that aggregate a set of different random number distributions into a single class or package.

### Programming Language API Level

Standard libraries and Application Programming Interfaces (APIs) conventionally available in every implementation of high-level languages may and should be used. Typically, abstract data structures representing such things as lists, queues and trees, are coupled in a set of classes that implement reusable data structures, such as the Java Collections Framework.

### Generic-Type Level

The use of generic classes is a kind of reuse. Generics are a way of creating parameterized general purpose templates for class types and subroutines (see e.g. Sommerville 1995). If the same kind of abstract structure, such as a list, is used for many different types of elements, like a list of agents and a list of events, this avoids the need to develop a separate implementation of the list for each type of element. As a result, programmers are less likely to make mistakes, thus avoiding faults and improving confidence in the simulation. Generic programming facilities are available in several object-oriented languages, such as C++ and Java.

Routine Level

Independent components implementing a single procedure or function may be reused, for example, a mathematical function used in a previously tested and verified setting can be reused in another simulation.

### *8.3.3 Defensive Programming*

The focus of good programming practices is on good code readability, flexible and fast development as well as ease of debugging. Other focuses yet to be considered are the actual testing and location of faults in the program. A typical test consists of running test cases for which outputs are known and confirming these every time the program is changed. Defensive approaches require that the programmer should never assume that a module will work as expected but instead the programmer should handle the appropriate testing for every module, possibly embedding the test in the program itself. Rather than considering testing only after programming the model, some of the testing should be planned and embedded as redundant code in the program. In software engineering defensive programming stands for an approach to fault tolerance for failure recovery. In social simulation the goal is somewhat different and consists of including extra, redundant, code to test for program failures and assist in the location of program faults.

Two techniques seem relevant to social simulation: contract-based programming and unit testing. Although the latter is often referred to in the literature, it is not a common practice in social simulation, probably because it requires a considerable amount of redundant code to test programs.

The former stands for a methodology for checking the violation of specified conditions in order to verify programs written with procedural languages on the basis of declarative statements, called *assertions*, which are appropriate to the kind of explorative programming used in simulation.

Testing and contract-based programming can be combined with the use of assertions and *exceptions*. Assertions are predicates placed in the program to test whether a specified condition is satisfied during testing. If the predicate evaluates to false, the program should halt for further debugging. Exceptions are messages that indicate the occurrence of exceptional conditions that were anticipated by the programmer, which usually halt the program as well.

Contract-based programming is a methodology for designing class specifications with checkable conditions at run time. The attributes of a class define the state space for the objects of that class. The set of states considered valid act as constraints on the range of values that objects of that class can assume. Constraints are limits defined over the ranges of the attributes of a class and may define dependencies among those attributes. If the valid state space of a class is violated in execution time then an exceptional condition is launched and the program execution halts, indicating information about the location and the type of violation occurred.

Constraints can also be specified at the routine level by specifying the range of the parameter values and the dependencies among parameters. In fact, the specification of a class may be understood as defining a *contract* between two programmers, the *client* and the *supplier* of that class, resulting in two types of constraints:

1. The methods' pre-conditions specified by the supplier, to which the client should abide;
2. The class invariant and the methods' post-conditions to which the supplier commits.

If the pre-conditions are not satisfied, then an exceptional condition occurs, which must be dealt with by the client programmer who violated the contract of that class. In contrast, if the class invariant or the post-conditions are not satisfied, then the fault is due to the particular implementation of that class, resulting in an assertion failure. This means that the supplier may not have programmed the class appropriately.

As an example, consider a partial definition of a class named Agent that specifies an agent in a culture dissemination model. Agents are distributed on a grid and the culture of each actor is defined as a nonempty set of features with a positive number of traits per feature. Each agent is aware of its position on the grid, which cannot be outside the grid bounds. The attributes of that class include a list of features, the number of traits per feature and the agent's position. The invariant of such a class would be defined as follows:

$$\text{features} \neq \{\} \ \textit{and} \ \text{number\_traits} > 0 \ \textit{and} \ \text{position is legal}$$

Suppose that agent interaction is specified with a bit-flipping kind of schema, which occurs only between contiguous neighbors and if the agents have different traits in at least one feature. Moreover, the agents must share the same trait in at least one feature after the interaction. In order to guarantee that every state change in objects of type Agent is valid, the invariant must be checked (1) right after the calling of that method; (2) right before it returns to the caller; and (3) right before the end of the class constructors. That is, whereas the bit-flipping pre-conditions should be checked *before* the interaction takes place, the post-conditions should be checked *after* the interaction. A partial definition of a Java class could be the following (with redundant code in boldface):

```
public class Agent {

    private SpaceGrid grid; // the agent´s interaction
    environment
    private int [] features;
    private int number_of_traits;
    private Position position; // the agent's current
    position in the grid
```

```
// class constructor
public Agent(final SpaceGrid grid, final int [] features,
               final int number_of_traits, final Position
               position) {

    this.grid = grid;
    this.features = features;
    this.number_of_traits = number_of_traits;
    this.position = position;

    // check the class invariant after initialisation
    assert checkInvariant();
}

// other methods here

// the bit-flipping operation
public void bitFlip(final Agent theOtherAgent) {

    // check the invariant before every method that
    change the agent's state
    assert checkInvariant();

    // test pre-conditions for bit-flipping
    if (theOtherAgent == null)
       throw new NullPointerException();
    if (!grid.areNeigbourhs(this, theOtherAgent)) //
    are the agents neighbours?
       throw
       new IllegalArgumentException("Agents are not
       neighbours, cannot bit-flip");
    // agents must have a different trait in at least one
    feature
    if (Arrays.dontMatch(features, theOtherAgent.
    getFeatures()) < 1)
       throw
       new IllegalArgumentException("Insufficient
       number of features to bit-flip");

    // make bit-flipping

    int[] other_features = theOtherAgent.getFeatures();

    int rand = randomFeature(features);
    while (features[rand] == other_features[rand])
      rand = randomFeature(features);
    other_features[rand] = features[rand]; // bit-flip
```

```
        // test post-conditions (agents must share at least
        one feature after bit-flip)
        assert Arrays.match(features, theOtherAgent.
        getFeatures()) >= 1;

        // check the invariant after every method that
        changed the agent's state
        assert checkInvariant() ;
    }

    private boolean checkInvariant() {

        if (grid == null || features == null || position ==
        null ||
            number_of_traits < 1 || grid.outOfBounds
            (position))
            return false;

        return true;
    }
} // end of class Agent
```

In the program, the invariant is checked at the end of the constructor, right after initializing the agent's attributes, and both before and after any state change in the methods. Since the invariant concerns the testing on the part of the programmer of that class (the supplier), rather than the user of that class (the client), the method checkInvariant is set to private. Since the method bitFlip changes the agent's state, the method pre-conditions and post-conditions should be checked. If the pre-conditions are not satisfied, an exception is thrown. This means that the error is on the caller of that method. If the class invariant or the post-conditions are not satisfied, then an assertion failure is launched, suggesting an error in the implementation of that class.

Contract-based programming helps programmers reason about the conceptual model and its programs. Too often it is difficult to understand the details of an implementation. Insofar as the contracts of a module are fully documented and are regarded as a precise specification of the computerised model – akin to a declarative programming language style – additional benefits are increased readability and the fostering of replicability. During testing the programmer will typically run the program with assertions enabled. For reasons of efficiency, the assertion mechanism can either be set to on or off.

### *8.3.4 Replication for Model Alignment*

Program testing is an essential activity for building simulations. However, testing alone cannot distinguish between faults in the implementation of the pre-computerised model and a surprising consequence of the model itself. Static methods are thus unavoidable and the examination of the source code is important in all stages of the

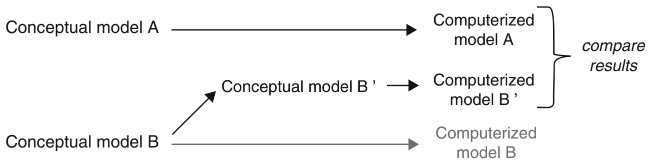



**Fig. 8.3** Model alignment

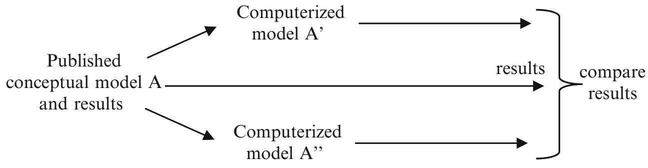



**Fig. 8.4** Model replication

development process. In practice, static and dynamic methods are used together. Most integrated development environments provide debugging tools and allow the programmer to trace the execution of a program and observe the values of variables statement by statement.

A slightly different technique, mostly used in numerical and technical simulation, is called "structured walk-throughs." This consists of having more than one person reading and debugging a program. All members of the development team are given a copy of a particular module to be debugged and the module developer goes through the code but does not proceed from one statement to another until everyone is convinced that a statement is correct (Law and Kelton 1991).

A different technique described by Sargent (1999) consists of reprogramming critical modules to determine if the same results are obtained. If two members are given the same specification for a module and they make two distinct implementations and obtain the same results, then the confidence in the implementation is increased. A more general problem is the extent to which models can be related to others so that their consequences and results are consistent with each other. In its most general form, this concerns both to V&V. After Axtell et al. (1996) it became known as the process of *model alignment*, which is used for determining whether different published models describing the same class of social phenomena produce the same results. Usually the alignment of two models A and B requires modifying certain features of model B – for instance by turning off certain features – in order to become equivalent to model A. This is represented in Fig. 8.3.

The term "model alignment" is frequently used synonymously for *model replication*. This assesses the extent to which building computerised models that draw on the same conceptual, usually published, model give results compatible with the ones reported in the published model. If the new results are similar to the published results, then the confidence in the correspondence between the computerised and the conceptual models is increased. Replication is represented in Fig. 8.4.

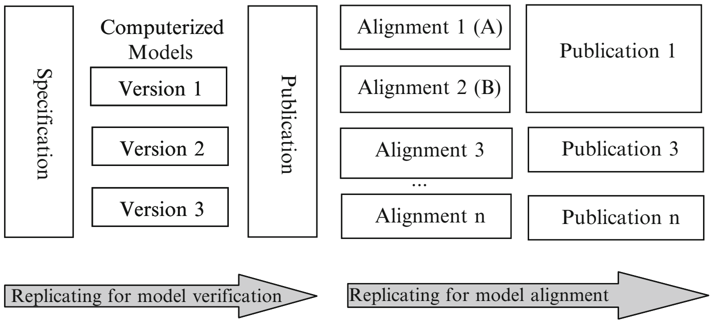



**Fig. 8.5** N-version programming and replication for model alignment

In the vast majority of cases, replication processes are mostly reactive and not proactive, in other words, the creator of the model publishes it and in the future someone else replicates it. However, we should not take replication lightly and see it as something that will possibly be done in the future. A verification technique, not so frequently used in social simulation, is called *N-version programming*. This is similar to reprogramming critical modules. In contrast to the practice of having other teams replicating models after they have already been reviewed, accepted and published, the effort of replication becomes focused on producing multiple computerised versions of a conceptual model *before* submitting it for peer reviewing.

Both N-version programming and replication are depicted in Fig. 8.5. The right-hand side of the diagram illustrates a perspective where a published conceptual model is replicated. The left-hand side illustrates a perspective where a conceptual model gives origin to multiple computerised versions of the model, implemented by different persons before it is actually published. In any case, if all versions lead to the same results, then there are reasonable grounds for trusting in the results obtained, as well as in the correspondence between the conceptual and computerised models.

#### 8.3.4.1 Types of Model Equivalence

The work of Axtell et al. (1996) is arguably the most-cited attempt to align two distinct but similar models. Rather than re-implementing Axelrod's culture dissemination model, Axtell and colleagues focused on the general case of aligning two models that reflected slightly distinctive mechanisms. For this purpose, Epstein and Axtell's Sugarscape model was progressively simplified in order to align with the results obtained by Axelrod's culture dissemination model.

Model alignment has been further investigated in a series of meetings called model-to-model (M2M) workshops (see Rouchier et al. 2008). The M2M workshops attract researchers interested in understanding and promoting the transferability of knowledge between model users. The replication of Edmonds and Hales (2003) is particularly informative on the problem of verification. They suggested that one should not trust an unreplicated simulation, since its results are almost certainly wrong in the sense that the computerised model differs from what was intended or assumed in the conceptual model. In order to align with a published conceptual model they had to rely on a double replication process insofar as the first replication did not seem to result in alignment. It was only after implementing a second computerised version of the published model that the authors were sure that the results reported in the published model did not align with their own. They concluded that the original implementation of the published model was producing results that could be misleading. This was due to different interpretations of the correct implementation of a mechanism of tournament selection for reproducing agents. Subtle differences in the mechanism implied different conclusions about the functioning of the model.

But how do we determine whether or not two models produce equivalent results? Axtell et al. (1996) defined three kinds of equivalence:

- Numerical identity: shows that the two models reproduce the results exactly.
- Relational equivalence: shows that the two models produce the same internal relationship among their results, for instance that a particular variable is a quadratic function of another.
- Distributional equivalence: shows that the two models produce distributions of results that cannot be distinguished statistically.

Numerical identity is hardly attainable in social complexity, except in the circumstances where models converge to some kind of final equilibrium, such as in Axelrod's culture dissemination model. Most simulations display several kinds of global behaviours that do not converge to equilibrium, are in the medium term self-reinforcing and make numerical identity unattainable. Among the three kinds of equivalence, distributional equivalence is the most demanding: it is achieved when the distributions of results cannot be distinguished statistically. What this shows is that at conventional confidence probabilities the statistics from different implementations *may* come from the same distribution, but it does *not prove* that this is actually the case. In other words, it does not prove that two implementations are algorithmically equivalent, but it allows us to disconfirm that they are. If this test survives repeatedly we somehow increase our confidence in the equivalence but only in the parameter range within which it has been tested.

#### 8.3.4.2 Programming for Replication

An important aspect when programming a simulation is to guarantee that it may be replicated by other researchers. Designing and programming for replicability involves a number of aspects that should be considered. Simulations are often a mix of conceptual descriptions and hard technical choices about implementation. The author who reports a model should assume that an alignment may later be tried and thus should be careful about providing detailed information for future use:

- Provide effective documentation about the conceptual and the computerised models; provide information about those technical options where the translation from the conceptual to the computerised model is neither straightforward nor consensual. Even if the difference between two computerised models may seem minor or irrelevant, small differences can affect the level of equivalence, even if the overall character of the simulation does not seem to change significantly.

- Outline in pseudo-code the parts of the program that may be liable to ambiguities; use modelling languages to represent implementation options, like UML. This is a formalism used to describe not only the static structure of the relations between classes but also different aspects of its dynamic behaviour, for instance, with activity graphs and sequence diagrams.
- Make the source code available online and ready to install. If possible, use a simulation platform to implement the model, hence fostering software reuse in order to make simulations reliable and more comparable to each other. If possible, make the simulation available to be run online, for instance, by using such technologies as Applets, a technology that allows the embedding of an execution model in Web pages, making it possible to be loaded and executed remotely. Making the computerised model available is crucial for others to be able to run the model with parameters settings that were not reported in your papers. Whereas for a certain set of parameter settings two simulations may match, this may not happen with other settings, suggesting the two models are not equivalent.
- Provide a detailed description about the results, statistical methods used, distributional information and qualitative measures. Make the bulk outputs available online or in appendices.

While programming for replicability is something to be considered on the part of the team that programs the model, a number of aspects should be considered by the team that decides to replicate a model. Often apparently irrelevant differences in two implementations can be the cause of different statistics that do not match (Edmonds and Hales 2003). This is particularly relevant if the description of the original model is not detailed sufficiently, making it difficult to assess whether the computerised model is implemented correctly and according to the conceptual model. When the original implementation is available, high degrees of confidence in the new implementation requires matching results with runs based on different parameters from those reported in the original model. The experiment of Edmonds and Hales provides a good deal of informative techniques and tips on this kind of model alignment:

- If the simulation shows very dynamic self-reinforcing effects in the long run, check the alignment of simulations in the first few cycles of their runs, averaged over several runs, in order to test whether they are initialized in the same way.
- Use different parameter settings and for each setting make several runs of both implementations over long-term periods. For each implementation collect several statistics from each time period and average them over the number of runs. For each such pair of sets of averages use statistical tests, such as the Kolmogorov-Smirnov test for the goodness of fit of cumulative distribution functions.
- Use modularity to test features separately. If two simulations do not align, turn off certain features of the implementations until they do so, and find the source of errors in the different modules. Reprogramming whole critical modules may also apply here.
- Use different toolkits or programming languages to re-implement simulations and if possible have this done by different people.

### 8.3.5 Participative-Based Methods

Replication is feasible when models are simple. When the goal is modelling a specific target domain, full of context, with significant amounts of rich detail, and use of empirical data and stakeholder participation, such as with the Companion Modelling approach, replication may not be feasible for verifying the computerised model. As in any other simulation, good programming practices and defensive programming are thus fundamental. In addition, insofar as some results may be due to software faults and be unexpected for stakeholders, participative-based methods are also a form of verification. This fact stresses the importance of involving domain experts and stakeholders as much as possible in all stages of the development and implementation process. Here, documentation and visualisation techniques can play a crucial role in bridging between the stakeholders' opinions and the intention of the programmer of the simulation. This is discussed in more detail in Chap. 10 in this volume (Barreteau et al. 2013).

## 8.4 Validation Approaches

We offered a conceptual definition of validation in Sect. 8.2.2. Had we given an operational definition, things would have become somewhat problematical. Models of social complexity are diverse and there is no definitive and guaranteed criterion of validity. As Amblard et al. (2007) remarked, "validation suggests a reflection on the intended use of the model in order to be valid, and the interpretation of the results should be done in relation to that specific context."

A specific use may be associated with different methodological perspectives for building the model, with different strategies, types of validity tests, and techniques - Fig. 8.6. Consider the kind of subjunctive, metaphorical, models such as Schelling's. In these models there is no salient validation step during the simulation development process. Design and validation walk together, and the intended use is not to show that the simulation is plausible against a specific context of social reality but to propose abstract or schematic mechanisms as broad representations of classes of social phenomena. In other different cases, the goal may be modelling a specific target domain, full of context, with use of empirical data and significant amounts of rich detail. Whereas in the former case a good practice could be modelling with the greatest parsimony possible so as to have a computational model sanctionable by human beings and comparable to other models, parsimony can be in opposition to the goal of descriptive richness and thus inappropriate to the latter case.

There are also different methodological motivations behind the use of a model, such as those conceived to predict or explain and those merely conceived to describe. Regardless of what method is used, the reproduction of characteristics of the object domain is important, but this can be assessed through rather different approaches during the model development process. If it is prediction you are

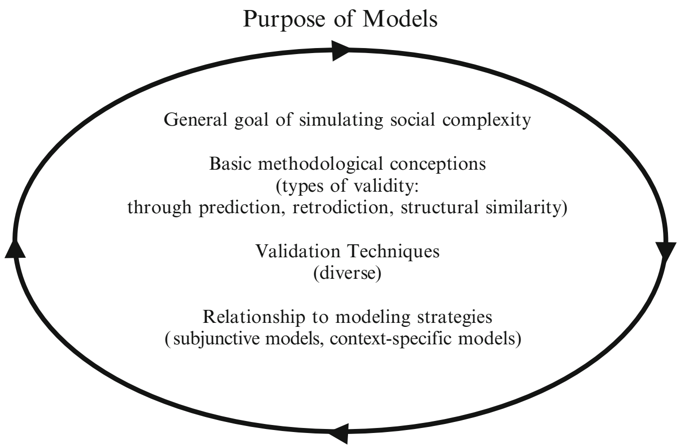



**Fig. 8.6** Validation implies considering the purpose of the model

seeking, validation consists of confronting simulated behaviour with the future behaviour of the target system (however, attempting to establish numerical prediction is not a normal goal in simulation). If it is explanation, validation consists of building plausible mechanisms that are able to reproduce simulated behaviour similar to real behaviour. If the goal is the more general aim of descriptiveness, explanation may probably be a goal as well, and a creative integration of ways for assessing the structure and results of the model, from quantitative to qualitative and participatory approaches, will be applied.

In conclusion, one should bear in mind that there is no one special method for validating a model. However, it is important to assess whether the simulation is subjected to good practices during its conception, whether it fits the intended use of the model builder and whether it is able to reproduce characteristics of the object domain. Assessing whether the goals of the modellers are well stated and the models themselves are well described in order to be understood and sanctioned by other model builders are *sine qua non* conditions for good simulation modelling.

In the remainder of this section, we revise the issue of the purpose of validating simulations along four dimensions: the general goal of simulation in social complexity, three basic methodological conceptions of validity types, a set of usual techniques applied in social simulation, and finally the relationship of validation to different modelling strategies with respect to the level of descriptive detail embedded in a simulation.

### *8.4.1 The Goal of Validation: Goodness of Description*

If one is using a predictive model, then the purpose of the model is to predict either past or future states of the target system. On the other hand, one may strive for a model that is able to describe the target system with satisfactory accuracy in order to become more knowledgeable about the functioning of the system, to exercise future and past scenarios, and to explore alternative designs or inform policies.

The objective in this section is to define the purpose of validation in terms of the purpose of simulating social complexity, which we will define as being of good description. This position entails that there is no single method or technique for validating a simulation. A diversity of methods for validating models is generally applied.

In the rest of this chapter we adopt the multi-agent paradigm for modelling. A conceptual understanding of validation, similar but more general than Moss and Edmonds' (2005), will be used:

> The purpose of validation is to assess whether the design of micro-level mechanisms, put forward as theories of social complexity validated to arbitrary levels, can be demonstrated to represent aspects of social behaviour and interaction that are able to produce macro-level effects either (i) broadly consistent with the subjacent theories; and/or (ii) qualitatively or quantitatively similar to real data.

By broad consistency we mean the plausibility of both micro specification and macro effects accounted as general representations of the target social reality. In its most extreme expression, plausibility may be evaluated on a metaphorical basis. By qualitative similarity to real data we mean a comparison with the model in terms of categorical outcomes, accounted as qualitative features, such as the shape of the outcomes, general stylized facts, or dynamical regimes. As for quantitative similarity we mean the very unlikely case in which the identification of formal numerical relationships between aggregate variables in the model and in the target – such as enquiring as to whether both series may draw from the same statistical distribution – proves to be possible.

Notice that this definition is general enough to consider both the micro-level mechanisms and macro-level effects assessed on a participatory basis. It is also general enough to consider two methodological practices – not necessarily incompatible – related to the extent to which models in social science simulation ought to be constructed on the basis of formal theories or ought to be based on techniques and approaches on the basis of the intuition of the model builders and stakeholders – an issue that we will come back to later. These are omnipresent methodological questions in the social simulation literature and are by no means irrelevant to the purpose of simulation models.

Suppose that on the basis of a very abstract model, such as the Schelling model, you were to evaluate the similarity of its outputs with empirical data. Then you will probably not take issue with the fact that the goal of predicting future states of the target would be out of the scope of simulation research for that kind of modelling. However, despite the belief that other sorts of validation are needed, this does not imply excluding the role of prediction, but emphasises the importance of description as the goal of simulating social complexity. In truth, what could be more contentious in assessing the Schelling model is the extreme simplicity used to describe the domain of social segregation. The descriptive power of multi-agent models makes them suited to model social complexity. Computational modelling corresponds to a process of abstraction, in that it selects some aspects of a subject being modelled, like entities, relations between entities and change of state, while ignoring those that may be considered less relevant to the questions that are of interest to the model builder. The expressiveness of multi-agent models allows the researcher to play with intuitive representations of different aspects of the target, such as defining societies with different kinds of agents, organizations, networks and environments, which interact with each other and represent social heterogeneity. By selecting certain aspects of social reality into a model, this process of demarcation makes multi-agent modelling suited to represent sociality as perceived by researchers and often by the stakeholders themselves.

The descriptive power of simulation is on par with the diversity of ways used for informing the construction and validation of models, from theoretic approaches to the use of empirical data or stakeholder involvement. At any rate, measuring the goodness of fit between the model and real data expressed with data series is neither the unique nor a typical criterion for sanctioning a model. The very idea of using a diversity of formal and informal methods is to assess the credibility of the mechanisms of the model as good descriptions of social behaviour and interaction, which must be shown to be resilient in the face of multiple tests and methods, in order to provide robust knowledge claims and allow the model to be open to scrutiny.

## *8.4.2 Broad Types of Validity*

When we speak about types of validity we mean three general methodological perspectives for assessing whether a model is able to reproduce expected characteristics of an object domain: validation through prediction, validation through retrodiction and validation through structural similarity. Prediction refers to validating a model by comparing the states of a model with future observations of the target system; retrodiction compares the states of the model with past observations of the target system; and structural similarity refers to assessing the realism of the structure of the model in terms of empirical and/or theoretical knowledge of the target system. In practice, all three approaches are interdependent and no single approach is used alone.

### 8.4.2.1 Validation Through Prediction

Validation through prediction requires matching the model with aspects of the target system before they were observed. The logic of predictive validity is the following: If one is using a *predictive model* – in which the purpose of the model is to predict future states of the target system – and the predictions prove satisfactory in repeated tested events, it may be reasonable to expect the model outcomes to stay reliable under similar conditions (Gross and Strand 2000). The purpose of prediction is somewhat problematic in social simulation:

- Models of social complexity usually show nonlinear effects in which the global behaviour of the model can become path-dependent and self-reinforcing, producing high sensitivity to initial conditions, which limits the use of predictive approaches.
- Many social systems show high volatility with unpredictable events, such as turning points of macroeconomic trade cycles or of financial markets that are in practice (and possibly in principle) impossible to predict; see (Moss and Edmonds 2005) for a discussion on this.
- Many social systems are not amenable to direct observation, change too slowly, and/or do not provide enough data to be able to compare model outcomes. Most involve human beings and are too valuable to allow repeated intervention, which hinders the acquisition of knowledge about its future behaviour. Policies based on false predictions could have serious consequences, thus making the purpose of prediction unusable (Gross and Strand 2000).

While quantitative prediction of the target system behaviour is rare or simply unattainable, prediction in general is not able to validate per se the *mechanisms* of the model as good representations of the target system. In the words of Troitzsch (2004), “What simulations are useful to predict is only how a target system might behave in the future qualitatively”. But a different model using different mechanisms that could lead to the same qualitative prediction may always exist, thus providing a different explanation for the same prediction. More often, the role of predicting future states of the target system becomes the exploration of new patterns of behaviour that were not identified before in the target system, whereby simulation acquires a speculative character useful as a heuristic and learning tool. What we are predicting is really new concepts that we had not realized as being relevant just for looking into the target.

#### 8.4.2.2 Validation Through Retrodiction

The difference from retrodiction to prediction is that in the former the intention is to reproduce *already* observed aspects of the target system. Given the existence of a historical record of facts from the target system, the rationale of *retrodictive* validity for a *predictive* model is the following: If the model is able to reproduce a historical record consistently and correctly, then the model may also be trusted for the future (Gross and Strand 2000). However, as we have mentioned, predictive models of social complexity are uncommon in simulation. Explanation rather than prediction is the usual motive for retrodiction. The logic of retrodictive validity is the following: If a model is able to consistently reproduce a record of past behaviours of the target system, then the mechanisms that constitute the model are *eligible candidates* for explaining the functioning of the target system. Nevertheless, retrodiction alone is not sufficient to assess the validity of the candidate explanations:

- Underdetermination: Given a model able to explain a certain record of behaviours or historical data, there will always be a different model yielding a different explanation for the same record.
- Insufficient quality of data: In many cases it is impossible to obtain long historical series of social facts in the target system. In the social sciences the very notion of social facts or data is controversial, can be subjective, and is not dissociable from effects introduced by the measurement process. Moreover, even when data is available it may not be in a form suitable to be matched to the bulk of data generated by simulation models.

Underdetermination and insufficient data suggest the crucial importance of domain experts for validating the *mechanisms* specified in the model. A model is only valid provided that *both* the generated outcomes and the mechanisms that constitute the model are sanctioned by experts in the relevant domain. The importance of validating the mechanisms themselves leads us to the *structural* validity of the model, which neither predictive nor retrodictive validity is able to assess alone.

#### 8.4.2.3 Validation Through Structural Similarity

In practice, the evaluation of a simulation includes some kind of prediction and retrodiction, based on expertise and experience. Given the implementation of micro-level mechanisms in the simulation, classes of behaviour at the macroscopic scale are identified in the model and compared to classes of behaviour identified in the target. Similarly, known classes of behaviour in the target system are checked for existence in the simulation. The former case is generally what we call the "surprising" character of simulations in which models show something beyond what we expect them to. However, only an assessment of the model at various points of view, including its structure and properties on different grains and levels, will truly determine whether the system reflects the way in which the target system operates. For instance, do agents' behaviour, the constituent parts and the structural evolution of the model match the conception we have about the target system with satisfactory accuracy? These are examples of the elements of realism between the model and the system that the researcher strives to find, which requires expertise in the domain on the part of the person who builds and/or validates the model.

### *8.4.3 Validation Techniques*

In this section we describe validation techniques used in social simulation. Some are used as common practices in the literature, and most of the terminology has been inhered from simulation in engineering and computer science, particularly from the reviews of validation and verification in engineering by Sargent (1999). All techniques that we describe can be found in the literature, but it would be rare to find a model in which only one technique was used, consistent with the fact that the validation process should be diverse. Also, there are no standard names in the literature and some techniques overlap with others.

#### 8.4.3.1 Face Validity

Face validity is a general kind of test used both before and after the model is put to use. During the model development process, the various intermediate models are presented to persons who are knowledgeable about, or are relevant to the problem in order to assess whether it is compatible with their knowledge and experience and reasonable for its purpose (Sargent 1999). Face validity may be used for evaluating the conceptual model, the components thereof, and the behaviour of the computerised models in terms of categorical outcomes or direct input/output relationships. This can be accomplished via documentation, graphing visualisation models, and animation of the model as it moves through time. Insofar as this is a general kind of test, it is used in several iterations of the model.

#### 8.4.3.2 Turing Tests

People who are knowledgeable about the behaviour of the target system are asked if they can discriminate between system and model outputs (Sargent 1999). The logic of Turing tests is the following: If the outputs of a computerised model are qualitatively or quantitatively indistinguishable from the observation of the target system, a substantial level of validation has been achieved.

Note that the behaviour of the target system does not need to be observed directly in the cases where a computerised representation is available. For example, suppose that videos of car traffic are transformed into three-dimensional scenes, whereby each object in the scene represents a car following the observed trajectory. If an independent investigator is not able to distinguish the computerised reproduction from an agent-based simulation of car traffic, then a substantial level of validation has been obtained for the set of behaviours represented in the simulation model.

#### 8.4.3.3 Historical Validity

Historical validity is a kind of retrodiction where the results of the model are compared with the results of previously collected data. If only a portion of the available historical data is used to design the model then a related concept is called *out-of-sample tests* in which the remaining data are used to test the predicative capacity of the model.

#### 8.4.3.4 Event Validity

Event validity compares the occurrence of particular events in the model with the occurrence of events in the source data. This can be assessed at the level of individual trajectories of agents or at any aggregate level. Events are situations that should occur according to pre-specified conditions, although not necessarily predictable. Some events may occur at unpredictable points in time or circumstances. For instance, if the target system data shows arbitrary periods of stable behaviours interwoven with periods of volatility with unpredictable turning points, the simulation should produce similar kinds of unpredictable turning events.

#### 8.4.3.5 Extreme Condition Tests

Extreme conditions are used for both verifying and validating the validation. The experimenter uses unlikely combinations of factors in the system, usually very high or low values for the inputs and parameters in order to test whether the simulation continues to make sense at the margins. For instance, if interaction among agents is nearly suppressed the modeller should be surprised if such activities as trade or culture dissemination continues in a population.

#### 8.4.3.6 Sensitivity Analysis

As a precautious rule one should consider that a model is only valid for the range of parameters that have been tested. Sensitivity analysis stands for tests in which parameters or even the inter-relations of model components are systematically varied in order to determine the effect on the behaviour of the model. It aims at three sorts of related considerations:

- Understanding the basic conditions under which the model behaves as expected;
- Finding the conditions that maximize the agreement of the model behaviour with the target system behaviour;
- Identifying the conditions that are sensitive, for instance, when changes in the input yield outputs that do not remain within known intervals, even when changes are carried out in a very controlled way.

Parameters that are sensitive should be made sufficiently accurate prior to using the model. If the output remains unpredictable even with controlled changes, the modeller should be concerned about making claims about the model.

Executing sensitivity tests is not a trivial task, and there are no special methods. If one imagines sweeping three parameters with 10 distinct values each then 720 configurations of input sets will be defined. If for each configuration we carry out five runs so as to obtain meaningful statistical figures, we can imagine 3,600 runs of the model. Since it is likely that some of the parameters will interact they should be swept in combinations – a fact which makes sensitivity tests already intractable for only a relatively small number of parameters.

Experience, the availability of empirical data, and the use of sampling techniques are the usual solution. A possible approach is constraining the range of values according to empirical data by ignoring ranges of values that are known from the start to be implausible in the target system. Yet, this might not be possible. The correspondence between parameter ranges in the model and in the target must be somehow known a priori, which requires that the model be subjected to some kind of testing anyway.

Sampling the parameter space is the usual solution. Related techniques in social simulation include learning algorithms for searching the parameter space, such as the Active Nonlinear Test (Miller 1998). Genetic algorithms for exploring the parameters of the model more efficiently are often used.

Besides sweeping parameters, any changes in the conditions of the model should be tested. Two architectural levels in the model must be considered:

1. The conceptual level, which involves changing the internal mechanisms or sub-models that constitute the larger model, such as changing the decision processes of the agents, their learning mechanisms or their interaction topology.
2. The system level, which involves low-level elements of the model, such as the effect of changing the agent activation regimes (e.g. uniform activation or random activation).

If changing elements at the system level determines different behaviours of the model that cannot be adequately interpreted, then the validity of the model can be compromised. The case of changing elements at conceptual levels is more subtle, and the validity of the results must be assessed by the researcher with reference to the validity of the composing elements of the model. This is basically a kind of cross-model or cross-element validation, as described below.

#### 8.4.3.7 Cross-Sectional Validity

Cross-sectional validity refers to the examination of the fit of congruence of social data to the results that simulation models produce in a specific point in time. This may be accomplished by comparing empirical data, such as a cross-sectional survey, with output generated by a model at a single time period. For example, simulation models of withdrawal behaviours in organisations, based on fuzzy set theory, have been used by Munson and Hulin (2000) for comparing correlations among frequencies of withdrawal behaviours in a cross-section survey with correlations among simulated behaviours after a certain number of iterations. The model that generated the correlation matrix that fitted better with the empirical data (evaluated by calculating root mean squared discrepancies) gained support as the most useful model of the organisational context that was being analysed. Notwithstanding, quantitative assessments between cross-sectional and simulated data are rare. Moreover, the benefits of simulation for representing time suggest an obvious disadvantage of cross-sectional approaches: they are unable to assess results from a longitudinal point of view.

#### 8.4.3.8 Comparison to Other Models

There are strong reasons to compare a model with other models. One reason is the unavailability, or insufficient quality, of data. Another is that models are often specified at a level of abstraction not compatible with available data. Moreover, even if data were available, the goodness of fit between real and simulated data, albeit reflecting evidence about the validity of the model as a data-generating process, does not provide evidence on how it operates. The most important reason for comparing models is intrinsic to the scientific practice. The very idea of validation is comparing models with other descriptions of the problem modelled, and this may include other simulation models that have been validated to some level.

If two models of the same puzzle lead to different conclusions, then they motivate us to check the validity of the models. Or, as we mentioned in Sect. 8.2, indicate that the computerised models have not been appropriately verified. The bricolage of methods for relating models in social simulation has become more mature after the aligning experience of Axtell et al. (1996), and is now a highly cited area. The goal of relating models can have different origins. In Sect. 8.2.4 the meaning of *model alignment* was described as well as the different methods for replicating (computerised) models. These focused on the *verification* of algorithmic equivalence of models through comparison of data. There are a number of approaches for relating models focused on the *validation* perspective:

- Extending models or composing models in a larger model, where different concepts, mechanisms or results are abstracted as components of the larger model; software reuse, as described in Sect. 8.2.2, is a kind of model composition.
- Docking data produced by a model to a set of data produced by another model; this may require changing and calibrating the model, for instance, by turning off features of the former in order to align with the latter.
- Varying systematically and controllably the internal structure of a model; in other words, playing with models within models in order to assess the overall validity of the larger model with reference to the validity of each one of the composing models; this is a kind of sensitivity analysis that resembles *cross-element validation*, which is mentioned below.
- Relating different computerised models that are based on different paradigms and provide different scales and perspectives for the same class of target; for instance, agent-based models and equation-based modelling.

#### 8.4.3.9 Cross-Element Validity

Cross-element validation, rather than comparing whole models, compares the results of a model whose architecture of the agents differs only in a few elements. The goal is to assess the extent to which changing elements of the model architecture produces results compatible with the expected results of the (larger) model; basically an exercise of composing different models within a larger model, which resembles structural sensitivity analysis. For instance, one may study the effects of using a model with agents in a bargaining game employing either evolutionary learning or reinforcement learning strategies, and assess which one of the strategies produces results compatible with theoretical analysis in game theory (Takadama et al. 2003).

A difficult issue faced with cross-element validation is that different results depend on the different elements used. The results obtained may only be a result of sensitivity to the elements. How can different tendencies be compared resulting from using different elements in the model? Arai and Watanabe (2008) have introduced a promising approach for use with time-series data. A quantitative comparison method based on the Fourier transform is used for measuring the distance between the results of two models that use different elements (e.g. different learning mechanisms).

Another problem with abstract simulations, in which the context of the empirical referent is vague, refers to the lack of real data for comparing which of the simulation outcomes are more realistic. As a result, the outcomes may only be assessed against a reference theoretical model. In contrast, when enough data from the target system are available, cross-element validation becomes a particular case of validation through retrodiction. For example, the cross-sectional validity tests employed by Munson and Hulin (2000) are a kind of cross-element validation through retrodiction, in which different theoretical models of individual withdrawal behaviours were tested within the same organisational model, with post-model data fit evaluation. Data generated for each theoretical model was compared with data from structured interviews in order to determine which theoretical model provided the best fit to the empirical data.

#### 8.4.3.10 Participatory Approaches for Validation

Participatory approaches refer to the involvement of stakeholders both in the design and the validation of a model. Such an approach, also known as Companion Modelling (Barreteau et al. 2001), assumes that model development must be itself considered in the process of social intervention, where dialogue among stakeholders, including both informal and theoretical knowledge, is embedded in the model development process. Rather than just considering the final shape of the model, both the process and the model become instruments for negotiation and decision making. It is particularly suited for policy or strategy development. This topic is discussed in Chap. 10 “Participatory Approaches” (Barreteau et al. 2013).

### 8.4.4 Relationship to Modelling Strategies

Regarding the diversity of methodological conceptions for modelling, different strategies may be adopted relating to the level of descriptive richness embedded in the simulations.

Several taxonomies of modelling strategies have been described in the literature (see David et al., 2004; Boero et al., 2005; Gibert 2008, pp. 42–44). Normally, the adoption of these strategies does not depend on the class of the target to be modeled, but in different ways to address it as the problem domain. For example, if a simulation is intended to model a system for the purpose of designing policies, this will imply representing more detail than a simulation intended for modeling certain social mechanisms of the system in a metaphorical way. But this also means that there is a trade-off between the effort that a modeler puts into verifying the simulation and puts into validating it. As more context and richness are embedded in a model, the more difficult it will be to verify it. Conversely, as one increases the descriptive richness of a simulation, more ways will be available to assess its validity. A tension that contrasts the tendency for constraining simulations by formal-theoretical constructs – normally easier to verify – and constraining simulations by theoretical-empirical descriptions – more amenable to validation by empirical and participative-based methods. In the remaining sections of this chapter, two contrasting modelling strategies are described.

#### 8.4.4.1 Subjunctive Agent-Based Models

A popular strategy in social simulation consists of using models as a means for expressing subjunctive moods to talk about possible worlds using what-if scenarios, like 'what would happen if something were the case'. The goal is building artificial societies for modelling possible worlds that represent classes of social mechanisms, while striving for maximal simplicity and strong generalisation power of the representations used. Reasons for striving for simplicity include the computational tractability of the model and to keep the data analysis as simple as possible.

Simplicity and generalization power are often seen as elements of elegance in a model. However, making the model simpler in the social sciences does not necessarily make the model more general. More often than not this kind of modelling only makes it metaphorically general, or simply counterfactual (with false assumptions). For example, 'What would happen if world geography is regarded as a two-dimensional space arranged on a $10 \times 10$ grid, where agents are thought of as independent political units, such as nations, which have specific behaviours of interaction according to simple rules?' To assume that world geography is one-dimensional, as Axelrod does in his Tribute Model, is clearly a false assumption. Often these models are associated with a design slogan coined by Axelrod, called the KISS approach – "Keep it Simple Stupid". Despite their simplicity, these models prove useful for concept formation and theoretical abstraction. The emergence of macro regularities from micro-levels of interaction becomes the source of concept formation and hypothesis illustration, with the power of suggesting novel theoretical debates.

Given the preference for simplification and abstraction, mechanisms used in these models are normally described in a formalized or mathematical way. Axelrod's models, such as the culture dissemination model, or Schelling's residential segregation model, are canonical examples. Their simplicity and elegance have been factors for popularity and dissemination that span numerous disciplines and ease replication and verification.

However, whereas simplicity eases verification, the use of metaphorical models also brings disadvantages. Consider a word composed of several attributes that represents an agent's culture, such as in Axelrod's culture dissemination model. The attributes do not have any specific meaning and are only distinguishable by their relative position in the word and so can be interpreted according to a relatively arbitrary number of situations or social contexts. However, such a representation may also be considered too simplified to mean anything relevant for such a complex concept as a cultural attribute. As a consequence, verification is hardly distinguishable from validation, insofar as the model does not represent a specific context of social reality. In such a sense, the researcher is essentially verifying experimentally whether the conceptions that he has in his mind are met by an operationalisation that is computationally expressed (David et al. 2007). Nevertheless, given their simplicity, subjunctive models can be easily linked and compared to other models, extended with additional mechanisms, as well as modified for model alignment, docking, or replication. Cross-element validation is a widely used technique.

At any rate, the fact that these models are easily replicable and comparable – but hardly falsifiable by empirically acquired characteristics of social reality – stresses their strong characteristic: when models based on strategies of maximal simplicity become accepted by a scientific community, their influence seems to reach several other disciplines and contexts. Perhaps for this reason, these kind of models are the most popular in social simulation, and some models are able to reach a considerable impact in many strains of social science.

#### 8.4.4.2 Context-Specific Agent-Based Models

It would be simplistic to say that models in social simulation can be characterized according to well-defined categories of validation strategies. Even so, the capacity to describe social complexity, whether through simplicity or through rich detail and context, is a determining factor for a catalogue of modelling strategies.

We cannot hope to model general social mechanisms that are valid in all contexts. There are many models that are not designed to be markedly general or metaphorically general, but to stress accurateness, diversity, and richness of description. Instead of possible worlds representing very arbitrary contexts, models are explicitly bounded to specific contexts. Constraints imposed on these models can vary from models investigating properties of social mechanisms in a large band of situations which share common characteristics, to models with the only ambition of representing a single history, like Dean's retrodiction of the patterns of settlement of a specific population in the southwestern United States, household by household (see Dean et al. 2000).

Constructing and validating a model of this kind requires the use of empirical knowledge. They are, for this reason, often associated with the idea of "Empirical Validation of Agent-Based Models."

What is the meaning of empirical in this sense? If the goal is to discuss empirical claims, then models should attempt to capture empirically enquired characteristics of the target domain. Specifying the context of descriptions in the model will typically provide more ways for enquiring quantitative and qualitative data in the target, as well as using experimental and participative methods with stakeholders. In this sense, empirical may be understood as a stronger link between the model and a context-specific, well-circumscribed, problem domain.

The model of Dean et al. (2000), which attempted to retrodict the patterns of settlement of the Anasazi in the Southwestern United States, household by household, is a well-known and oft-cited example of a highly contextualized model built on the basis of numerous sources, from archaeological data to anthropological, agricultural and ethnographic analyses, in a multidisciplinary context.

Given the higher specificity of the target domain, the higher diversity of ways for enriching the model as well as the increased semantic specificity of the outputs produced by the model, context-specific models may be more susceptible to be compared with empirical results of other methods of social research. On the other hand, comparison with other simulation models is complex and these models are more difficult to replicate and verify.

#### 8.4.4.3 Modus Operandi: Formal and Informal Approaches

The tension between simplicity and descriptive richness expresses two different ways for approaching the construction and validation of a model. One can start with a rich, complex, realistic description and only simplify it where this turns out to be possible and irrelevant to the target system – known as the KIDS approach (Edmonds and Moss 2005). Or one starts from the outset with the simplest possible description and complexifies it only when it turns out to be necessary to make the model more realistic, nevertheless keeping the model as simple as possible – known as the KISS approach (Axelrod 1997b).

In practice, both trends are used for balancing trades-offs between the model's descriptive accuracy and the practicality of modelling, according to the purpose and the context of the model. This raises yet another methodological question: the extent to which models ought to be designed on the basis of formal theories, or ought to be constrained by techniques and approaches just on the basis of the intuition of the model builders and stakeholders. As we have seen, strong, subjunctive, agent-based models with metaphorical purposes tend to adopt the simplicity motto with extensive use of formal constructs, making the models more elegant from a mathematical point of view, easier to verify, but less liable to validation methods. Game theoretical models, with all their formal and theoretical apparatus, are a canonical example. Results from these models are strongly constrained by the formal theoretical framework used.

A similar problem is found when agent-based models make use of cognitive architectures strongly constrained by logic-based formalisms, such as the kind of formalisms used to specify BDI-type architectures. If the cognitive machinery of the agents relies on heuristic approaches that have been claimed valid, many researchers in the literature claim that cognitive agent-based models can be validated in the empirical sense of context-specific models. Cited examples of this kind usually point to agent-based models based on the Soar architecture.

At any rate, context-specific models are normally more eclectic and make use of both formal and informal knowledge, often including stakeholder evidence in order to build and validate the models. Model design tends to be less constrained a priori by formal constructs. In principle, one starts with all aspects of the target domain that are assumed to be relevant and then explores the behaviour of the model in order to find out whether there are aspects that do not prove relevant for the purpose of the model. The typical approach for modelling and validation can be summarized in a cycle with the following iterative and overlapping steps:

A. *Building and validating pre-computerised and computerised models*: Several descriptions and specifications are used to build a model, eventually in the form of a computer program, which are micro-validated against a theoretical framework and/or empirical knowledge, usually qualitatively. This may include the individual agents' interaction mechanisms (rules of behaviour for agents or organisations of agents), their internal mechanisms (e.g. their cognitive machinery), the kind of interaction topology or environment, and the passive entities with which the agents interact. The model used should be as accurate as possible for the context in consideration as well as flexible for testing how parameters vary in particular circumstances. Empirical data – if available – should be used to help configure the parameters. Both the descriptions of the model and the parameters used should be validated for the specific context of the model. For example, suppose empirical data are available for specifying the consumer demand of products. If the demand varies from sector to sector, one may use data to inform the distribution upon which the parameter could be based for each specific sector.
B. *Specifying expected behaviours of the computerised model*: Micro and macro characteristics that the model is designed to reproduce are established from the outset based on theoretical and/or empirical knowledge. Any property, from quantitative to qualitative measures, such as emergent key facts the model should reproduce (stylized facts), the statistical characteristic or shape of time-data series (statistical signatures) and individual agents' behaviour along the simulation (individual trajectories), can be assessed. This may be carried out in innumerable ways, according to different levels of description or grain, and be more or less general depending on the context of the model and the kind of empirical

knowledge available. For instance, in some systems it may be enough to predict just a "weak" or "positive" measure on some particular output, such as a positive and weak autocorrelation. Or we might look for the emergence of unpredictable events, such as stable regimes interleaved with periods of strong volatility, and check their statistical properties for various levels of granularity. Or the emergence of different structures or patterns associated with particular kinds of agents, such as groups of political agents with "extremist" or "moderate" agents.

C. *Testing the computerised model and building and validating post-computerised models*: The computerised model is executed. Both individual and aggregate characteristics are computed and tested for sensitivity analysis. These are micro-validated and macro-validated against the expected characteristics of the model established in step B according to a variety of validation techniques, as described in Sects. 8.4.2 and 8.4.3. A whole process of building post-computerised models takes place, possibly leading to the discovery of unexpected characteristics in the behaviour of the computerised model which should be assessed with further theoretical or empirical knowledge about the problem domain.

## Further Reading

Good introductions to validation and verification of simulation models in general are Sargent (1999) and Troitzsch (2004), the latter with a focus on social simulation. Validation of agent-based models in particular is addressed by Amblard and colleagues (Amblard et al. 2007).

For readers more interested in single aspects of V&V with regard to agent-based models in the context of social simulation, the following papers provide highly accessible starting points:

- Edmonds and Hales (2003) demonstrate the importance of model replication (or model alignment) by means of a clear example.
- Boero and Squazzoni (2005) examine the use of empirical data for model calibration and validation and argue that "the characteristics of the empirical target" influence the choice of validation strategies.
- Moss and Edmonds (2005) discuss an approach for cross-validation that combines the involvement of stakeholders to validate the model qualitatively on the micro level with the application of statistical measures to numerical outputs to validate the model quantitatively on the macro level.

Finally, for a more in-depth epistemological perspective on verification and validation I would refer the inclined reader to a revised version of my EPOS 2006 paper (David 2009).

## References

Amblard F, Bommel P, Rouchier J (2007) Assessment and validation of multi-agent models. In: Phan D, Amblard F (eds) Agent-based modelling and simulation in the social and human sciences. The Bardwell Press, Oxford, pp 93–114

Arai R, Watanabe S (2008) A quantitative comparison method for multi-agent based simulation in the feature space. In: David N, Sichman JS (eds) Multi-agent-based simulation IX, MABS 2008, revised selected papers (Lecture notes in artificial intelligence), vol 5269. Springer, Berlin, pp 154–166

Axelrod R (1997a) The dissemination of culture: a model with local convergence and global polarization. J Conf Resolut 41(2):203–226

Axelrod R (1997b) Advancing the art of simulation in the social sciences. In: Conte R, Hegselmann R, Terna P (eds) Simulating social phenomena. Springer, Berlin, pp 21–40

Axtell R, Axelrod R, Epstein J, Cohen M (1996) Aligning simulation models: a case study and results. Comp Math Organiz Theory 1(2):123–141

Barreteau O, Bousquet F, Attonaty JM (2001) Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems. J Artif Soc Soc Simul 4(2). http://jasss.soc.surrey.ac.uk/4/2/5.html

Barreteau O et al (2013) Participatory approaches. Chapter 10 in this volume

Boero R, Squazzoni F (2005) Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science. J Artif Soc Soc Simul 8(4). http://jasss.soc.surrey.ac.uk/8/4/6.html

David N (2009) Validation and verification in social simulation: patterns and clarification of terminology. In: Squazzoni F (ed) Epistemological aspects of computer simulation in the social sciences, second international workshop, EPOS 2006, Brescia, 5–6 Oct 2006, revised selected and invited papers (Lecture notes in computer science), vol 5466. Springer, Berlin, pp 117–129

David N, Sichman J, Coelho H (2003) Towards an emergence-driven software process for agent-based-simulation. In: Sichman JS, Bousquet F, Davidsson P (eds) Multi-agent-based simulation II, third international workshop, MABS 2002, Bologna, 15–16 July 2002, revised papers (Lecture notes in artificial intelligence), vol 2581. Springer, Berlin, pp 89–104

David N, Marietto M, Sichman J, Coelho H (2004) The structure and logic of interdisciplinary research in agent-based social simulation. J Artif Soc Soc Simul 7(3). http://jasss.soc.surrey.ac.uk/7/3/4.html

David N, Sichman J, Coelho H (2007) Simulation as formal and generative social science: the very idea. In: Gershenson C, Aerts D, Edmonds B (eds) Worldviews, science and us: philosophy and complexity. World Scientific, Singapore, pp 266–284

David N, Caldas JC, Coelho H (2010) Epistemological perspectives on simulation III: selected papers of EPOS 2008. J Artif Soc Soc Simul 13(1), special section. http://jasss.soc.surrey.ac.uk/13/1/

Dean S et al (2000) Understanding Anasazi culture change through agent-based modeling. In: Kohler T, Gumerman G (eds) Dynamics in human and primate societies: agent-based modeling of social and spatial processes. Oxford University Press, New York/Oxford, pp 179–205

Edmonds B, Hales D (2003) Replication, replication and replication: some hard lessons from model alignment. J Artif Soc Soc Simul 6(4). http://jasss.soc.surrey.ac.uk/6/4/11.html

Edmonds B, Moss S (2005) From KISS to KIDS: an 'anti-simplistic' modelling approach. In: Davidsson P, Logan B, Takadama K (eds) Multi-agent and multi-agent-based simulation (Lecture notes in artificial intelligence), vol 3415. Springer, Berlin, pp 130–144

Frank U, Troitzsch KG (2005) Epistemological perspectives on simulation. J Artificial Soc Soc Simul 8(4). http://jasss.soc.surrey.ac.uk/8/4/7.html

Galán JM et al (2013) Checking simulations: detecting and avoiding errors and artefacts. Chapter 6 in this volume

Gilbert N (2008) Agent-based models, vol 153, Quantitative applications in the social sciences. Sage, London

Gross D, Strand R (2000) Can agent-based models assist decisions on large-scale practical problems? A philosophical analysis. Complexity 5(6):26–33

Kleijnen JPC (1995) Verification and validation of simulation models. Euro J Operat Res 82: 145–162

Law A, Kelton D (1991) Simulation modelling and analysis, 2nd edn. McGraw Hill, New York

Miller JH (1998) Active nonlinear tests (ANTs) of complex simulation models. Manage Sci 44(6): 820–830

Moss S, Edmonds B (2005) Sociology and simulation: statistical and qualitative cross-validation. Am J Sociol 110(4):1095–1131

Munson L, Hulin C (2000) Examining the fit between empirical data and theoretical simulations. In: Ilgen DR, Hulin CL (eds) Computational modeling of behaviour in organisations: the third scientific discipline. American Psychological Association, Washington, DC, pp 69–84

Rouchier J, Cioffi-Revilla C, Polhill JG, Takadama K (2008) Progress in model-to-model analysis. J Artif Soc Soc Simul 11(2). http://jasss.soc.surrey.ac.uk/11/2/8.html

Sargent R (1999) Verification and validation of simulation models. In: Farrington PA, Nembhard HB, Sturrock DT, Evans GW (eds) Proceedings of the 1999 winter simulation conference, Squaw Peak, Phoenix, 5–8 Dec 1999, pp 39–48. http://www.informs-sim.org/wsc99papers/005.PDF

Sommerville I (1995) Software engineering, 5th edn. Addison Wesley, Boston

Squazzoni F (2009) Epistemological aspects of computer simulation in the social sciences, second international workshop, EPOS 2006, Brescia, Revised selected and invited papers (Lecture notes in computer science), vol 5466. Springer, Berlin, 5–6 Oct 2006

Takadama K, Suematsu Y, Sugimoto N, Nawa N, Shimohara K (2003) Cross-element validation in multi-agent-based simulation: switching learning mechanisms in agents. J Artif Soc Soc Simul 6(4). http://jasss.soc.surrey.ac.uk/6/4/6.html

Troitzsch K (2004) Validating simulation models. In: Horton G (ed) Proceedings of the 18th European simulation multiconference, ESM 2004, 13–16 June 2004, Magdeburg. SCS, Erlangen, pp 265–270

# Chapter 9
# Understanding Simulation Results

**Andrew Evans, Alison Heppenstall, and Mark Birkin**

**Why Read This Chapter?** Overall, this chapter aims to help you understand the results that a simulation model produces, by suggesting some ways to analyse and visualise them. The chapter concentrates on the internal dynamics of the model rather than its relationship to the outside world.

**Abstract** Simulation modelling is concerned with the abstract representation of entities within systems and their inter-relationships; understanding and visualising these results is often a significant challenge for the researcher. Within this chapter we examine particular issues such as finding "important" patterns and interpreting what they mean in terms of causality. We also discuss some of the problems with using model results to enhance our understanding of the underlying social systems which they represent, and we will assert that this is in large degree a problem of isolating causal mechanisms within the model architecture. In particular, we highlight the issues of identifiability and equifinality – that the same behaviour may be induced within a simulation from a variety of different model representations or parameter sets – and present recommendations for dealing with this problem. The chapter ends with a discussion of avenues of future research.

## 9.1 Introduction

Simulation models may be constructed for a variety of purposes. Classically these purposes tend to centre on either the capture of a set of knowledge or making predictions. Knowledge capture has its own set of issues that are concerned with structuring and verifying knowledge in the presence of contradiction and uncertainty. The problems of prediction, closely associated with calibration and

A. Evans (✉) • A. Heppenstall • M. Birkin
Centre for Spatial Analysis and Policy, University of Leeds, Leeds, Britain, LS2 9JT
e-mail: a.j.evans@leeds.ac.uk

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_9,

validation, centre around comparisons with real data, for which the methods covered in Chap. 8 (David 2013) are appropriate. In this chapter, however, we look at what our models tell us through their internal workings and logic; how we might understand/interpret simulation results as results about an attempted simulation of the real world, rather than as results we expect to compare directly with the world. Here then, we tackle the third purpose of modelling: the *exploration* of abstracted systems through simulation. In a sense, this is a purpose predicated only on the limitations of the human mind. By common definition, simulation modelling is concerned with abstract representations of entities within systems and their interrelationships, and with the exploration of the ramifications of these abstracted behaviours at different temporal and geographical scales. In a world in which we had larger brains, models would not be required to reveal anything – we would instantly see the ramifications of abstracted behaviours in our heads. To a degree, therefore, models may be seen as replacing the hard joined-up thinking that is required to make statements about the way the world works. This chapter looks at what this simplifying process tells us about the systems we are trying to replicate.

In part, the complications of simulation modelling are a product of the dimensionality of the systems with which we are dealing. Let us imagine that we are tackling a system of some spatio-temporal complexity, for example, the prices in a retail market selling items A, B, and C. Neighbouring retailers adjust their prices based on local competition, but the price of raw materials keeps the price surface out of equilibrium. In addition, customers will only buy one of the products at a time, creating a link between the prices of the three items. Here, then, we have three interdependent variables, each of which varies spatio-temporally, with strong auto- and cross-correlations in both time and space. What kinds of techniques can be used to tease apart such complex systems? In Sect. 9.2 of this chapter we will discuss some of the available methodologies broken down by the dimensionality of the system in question and the demands of the analysis. Since the range of such techniques is extremely sizable, we shall detail a few traditional techniques that we believe might be helpful in simplifying model data that shows the traits of complexity, and some of the newer techniques of promise.

Until recently, most social science models represented social systems using mathematical aggregations. We have over 2,500 years' worth of techniques to call upon that are founded on the notion that we need to simplify systems as rapidly as we can to the point at which the abstractions can be manipulated within a single human head. As is clear, not least from earlier contributions in this volume, it is becoming increasingly accepted that social scientists might reveal more about systems by representing them in a less aggregate manner. More specifically, the main difference between mathematics and the new modelling paradigm is that we now aspire to work at a scale at which the components under consideration can be represented as having their own discrete histories; mathematics actually works in a very similar fashion to modern models, but at all the other scales. Naturally there are knock-ons from this in terms of the more explicit representation of objects, states and events, but these issues are less important than the additional simulation and analytical power that having a history for each component of a system gives us. Of course, such a "history" may just be the discrete position of an object at a single historical moment, and plainly at this level of complication the boundary between such models and, for example, Markov models is somewhat diffuse, however, as the history of components becomes more involved, so the power of modern modelling paradigms comes to the fore. What is lacking, however, are the techniques that are predicated on these new architectures. Whilst models which are specified at the level of individual entities or 'agents' may also be analysed using conventional mathematical techniques, in Sect. 9.3 of the chapter we will discuss some more novel approaches which are moving the direction of understanding the outputs of these new, unaggregated, models on their own terms.

One of the reasons that simulation models are such a powerful methodology for understanding complex systems is their ability to display aggregate behaviour which goes beyond the simple extrapolation of the behaviour of the individual component parts. In mathematical analysis, such as dynamical systems theory, this behaviour tends to be linked to notions of equilibrium, oscillation, and catastrophe or bifurcation. Individual and agent-based modelling approaches have veered more strongly towards the notion of emergence, which can be defined as "an unforeseen occurrence; a state of things unexpectedly arising" (OED 2010). The concept of emergence is essentially a sign of our ignorance of the causal pathways within a system. Nevertheless, emergence is our clearest hope for developing an understanding of systems using models. We hope that emergence will give us a perceptual shortcut to the most significant elements of a system's behaviour. When it comes to applications, however, emergence is a rather double-edged blade: emergence happily allows us to see the consequence of behaviours without us having to follow the logic ourselves, however it is problematic in relying upon us to filter out which of the ramifications are important to us. As emergence is essentially a sign of incomplete understanding, and therefore weakly relative, there is no objective definition of what is "important". One might imagine one day a classification of the kinds of patterns that relate to different types of causal history, but there is no objective manner of recognising a pattern as "important" as such. These two problems: finding "important" patterns (in the absence of any objective way of defining "important") and then interpreting what they mean in terms of causality are the issues standing between the researcher and perfect knowledge of a modelled system. In the fourth section of this chapter, we will discuss some of the problems with using model results to enhance our understanding of the underlying social systems which they represent, and we will assert that this is in large degree a problem of isolating causal mechanisms within the model architecture. In particular, we highlight the issues of equifinality and identifiability – that the same behaviour may be induced within a simulation from a variety of different model representations or parameter sets – and present recommendations for dealing with this problem. Since recognising emergence and combating the problems of identifiability and equifinality are amongst the most urgent challenges to effective modelling of complex systems, this leads naturally to a discussion of future directions in the final section of the chapter.

## 9.2 Aggregate Patterns and Conventional Representations of Model Dynamics

Whether a model is based on deductive premises or inferred behaviours, any new understanding of a given modelled system tends to be developed inductively. Modellers examine model outputs, simplify them, and then try to work out the cause utilising a combination of hypothesis dismissal, refinement and experimentation. For example, a modeller of a crowd of people might take all the responses of each person over time and generate a single simple mean statistic for each person; these might then be correlated against other model variables. If the correlation represents a real causal connection, then varying the variables should vary the statistic. Proving such causal relationships is not something we often have the ability to do in the real world. During such an analysis the simplification process is key: it is this that reveals the patterns in our data. The questions are: how do we decide what needs simplifying and, indeed, how simple to make it?

We can classify model results by the dimensionality of the outputs. A general classification for social systems would be:

- Single statistical aggregations (1D)
- Time series of variables (2D)
- The spatial distributions of invariants (2D) or variables (3D)
- Spatio-temporal locations of invariants (3D) or variables (4D)
- Other behaviours in multi-dimensional variable space (*n*D)

For simplicity, this assumes that geographical spaces are essentially two-dimensional (while recognising that physical space might also be represented along linear features such as a high street, across networks, or within a three dimensional topographical space for landforms or buildings). It should also be plain that in the time dimension, models do not necessarily produce just a stream of data, but that the data can have complex patternation. By their very nature, individual-level models, predicated as they are on a life-cycle, will never stabilise in the way a mathematical model might (Grimm 1999); instead models may run away or oscillate, either periodically or chaotically.

Methods for aiding pattern recognition in data break down, again, by the dimensionality of the data, but also by the dimensionality of their outputs. It is quite possible to generate a one-number statistic for a 4D spatio-temporal distribution. In some cases, the reduction of dimensionality is the explicit purpose of the technique, and the aim is that patterns in one set of dimensions should be represented as closely as possible in a smaller set of dimensions so they are easier to understand. Table 9.1 below presents a suite of techniques that cross this range (this is not meant to be an exhaustive list; after all, pattern recognition is a research discipline of its own with a whole body of literature including several dedicated journals).

To begin with, let us consider some examples which produce outputs in a single dimension. In other words, techniques for generating global and regional statistics describing the distribution of variables across space, either a physical space, or a

**Table 9.1** Pattern recognition techniques for different input and output data dimensions

| | 1D output | 2D output | 3D output | 4D output | *N*D |
|---|---|---|---|---|---|
| 1D input | | | | | |
| 2D input | Exploratory statistics | Cluster locating<br>Fourier/wavelet transforms | | | |
| 3D input | Entropy statistics | Phase diagrams<br>Fourier/wavelet transforms | | | |
| 4D input | Diffusion statistics | Time slices | Recurrence plots | | |
| *n*D | Network statistics | Eigenvector analysis | Sammon mapping | Animations | Heuristic techniques |

variable space. Such statistics generally tend to be single time-slice, but can be generated for multiple time-slices to gauge overall changes in the system dynamics.

Plainly, standard aggregating statistics used to compare two distributions, such as the variable variance, will lose much of interest, both spatially and temporally. If we wish to capture the distribution of invariants, basic statistics like Nearest-Neighbour (Clark and Evans 1954) or the more complex patch shape, fragmentation and connectivity indices of modern ecology (for a review and software see McGarigal 2002) provide a good starting point. Networks can be described using a wide variety of statistics covering everything from shortest paths across a network, to the quantity of connections at nodes (for a review of the various statistics and techniques associated with networks, see Boccaletti et al. 2006; Evans 2010). However, we normally wish to assess the distribution of a variable across a surface – for example a price surface or a surface of predicted retail profitability. One good set of global measures for such distributions are entropy statistics. Suppose we have a situation in which a model is trying to predict the number of individuals that buy product A in one of four regions. The model is driven by a parameter, beta. In two simulations we get the following results: *simulation one* (low beta): 480, 550, 520, 450; s*imulation two*: (high beta) 300, 700, 500, 400. Intuitively the first simulation has less dispersal or variability than the second simulation. An appropriate way to measure this variability would be through the use of entropy statistics. The concept of entropy originates in thermodynamics, where gases in a high entropy state contain dispersed molecules. Thus high entropy equates to high levels of variability. Entropy statistics are closely related to information statistics where a high entropy state corresponds to a high information state. In the example above, *simulation two* is said to contain more 'information' than *simulation one*, because if we approximate the outcome using no information we would have a flat average – 500, 500, 500, 500 – and this is closer to *simulation one* than *simulation two*. Examples of entropy and information statistics include Kolmogorov-Chaitin, mutual information statistics and the Shannon information statistic. Most applications in the literature use customised code for the computation of entropy statistics, although the computation of a limited range of Generalised Entropy indices is possible within Stata.[1] Entropy statistics can also be used to describe flows across networks. In this sense they provide a valuable addition to networks statistics: most network statistics concentrate on structure rather than the variable values across them. Unless they are looking specifically at the formation of networks over time, or the relationship between some other variable and network structure, modellers are relatively bereft of techniques to look at variation *on* a network.

In the case where variability is caused and constrained by neighbourhood effects, we would expect the variation to be smoother across a region. We generally expect objects in space under neighbourhood effects to obey Tobler's first law of geography (1970) that everything is related, but closer things are related more. This leads to spatial auto- or cross-correlation, in which the values of variables at a point reflect those of their neighbours. Statistics for quantifying such spatial auto- or cross-correlation at the global level, or for smaller regions, such as Moran's I and Geary's C, are well-established in the geography literature (e.g. Haining 1990); a useful summary can be found in Getis (2007).

Such global statistics can be improved on by giving some notion of the direction of change of the auto- or cross-correlation. Classically this is achieved through semi-variograms, which map out the intensity of correlation in each direction traversed across a surface (for details, see Isaaks and Srivastava 1990). In the case where it is believed that local relationships hold between variables, local linear correlations can be determined, for example using Geographically Weighted Regression (GWR: for details, see Fotheringham et al. 2002). GWR is a technique which allows the mapping of $R^2$s calculated within moving windows across a multivariate surface and, indeed, mapping of the regression parameter weights. For example, it would be possible in our retail results to produce a map of the varying relationship between the amount of A purchased by customers and the population density, if we believed these were related. GWR would not just allow a global relationship to be determined, but also how this relationship changed across a country. One important but somewhat overlooked capability of GWR is its ability to assess how the strength of correlations varies with scale by varying the window size. This can be used to calculate the key scales at which there is sufficient overlap between the geography of variables to generate strong relationships (though some care is needed in interpreting such correlations, as correlation strength generally increases with scale: Robinson 1950; Gehlke and Biehl 1934). Plainly, identifying the key scale at which the correlations between variables improve gives us some

[1] Confusingly, 'generalised entropy' methods are also widely used in econometrics for the estimation of missing data. Routines which provide this capability, e.g. in SAS, are not helpful in the description of simulation model outputs!

ability to recognise key distance-scales at which causality plays out. In our example, we may be able to see that the scale at which there is a strong relationship between sales of A and the local population density increases as the population density decreases, suggesting rural consumers have to travel further and a concomitant non-linearity in the model components directing competition.

If, on the other hand, we believe the relationships do not vary smoothly across a modelled surface, we instead need to find unusual clusters of activity. The ability to represent spatial clustering is of fundamental importance, for example, within Schelling's well-known model of segregation in the housing market (Schelling 1969). However clustering is often not so easy to demonstrate within both real data and complex simulation outputs. The most recent techniques use, for example, wavelets to represent the regional surfaces, and these can then be interpreted for cluster-like properties. However, for socio-economic work amongst the best software for cluster detection is the Geographical Analysis Machine (GAM), which not only assesses clustering across multiple scales, but, also allows assessment of clustering in the face of variations in the density of the population at risk. For example, it could tell us where transport network nodes were causing an increase in sales of A, by removing regions with high sales caused by high population density (the population "at risk" of buying A). Clusters can be mapped and their significance assessed (Openshaw et al. 1988).

Often, simulations will be concerned with variations in the behaviour of systems, or their constituent agents, over time. In common with physical systems, social and economic systems are often characterised by periodic behaviour, in which similar states recur, although typically this recurrence is much less regular than in many physical systems. For example, economic markets appear to be characterised by irregular cycles of prosperity and depression. Teasing apart a model can provide non-intuitive insights into such cycles. For example, Heppenstall et al. (2006) considered a regional network of petrol stations and showed within an agent simulation how asymmetric cyclical variations in pricing (fast rises and slow falls), previously thought to be entirely due to a desire across the industry to maintain artificially high profits, could in fact be generated from more competitive profit-maximisation in combination with local monitoring of network activity. While it is, of course, not certain these simpler processes cause the pattern in real life, the model exploration does give researchers a new explanation for the cycles and one that can be investigated in real petrol stations.

In trying to detect periodic behaviour, wavelets are rapidly growing in popularity (Graps 2004). In general, one would assume that the state of the simulation can be represented as a single variable which varies over time (let's say the average price of A). A wavelet analysis of either observed or model data would decompose this trend into chunks of time at varying intervals, and in each interval the technique identifies both a long-term trend and a short-term fluctuation. Wavelets are therefore particularly suitable for identifying cycles within data. They are also useful as filters for the removal of noise from data, and so may be particularly helpful in trying to compare the results from a stylised simulation model with observed data which would typically be messy, incomplete or subject to random bias. It has been argued that such decompositions are fundamentally helpful in establishing a basis for forecasting (Ramsey 2002). A software environment for wavelet analysis is provided by Wavelab, using procedures derived from the Matlab statistical package.

Wavelets are equally applicable in both two and three dimensions. For example, they may be useful in determining the diffusion of waves across a two-dimensional space and over time, and can be used to analyse, for example, the relationship between wave amplitude and propagation distance. Viboud et al. (2006) provide a particularly nice example of such a use, looking at the strength of the propagation of influenza epidemics as influenced by city size and average human travel distances in the USA. Other, more traditional statistics, such as the Rayleigh statistic (Fisher et al. 1987; Korie et al. 1998) can also be used to assess the significance of diffusion from point sources.

In addition to global and regional aggregate statistics of single variables or cross-correlations, it may be that there is simply too great a dimensionality to recognise patterns in outputs and relate them to model inputs. At this point it is necessary to engage in multi-dimensional scaling. If an individual has more than four characteristics, then multi-dimensional scaling methods can be used to represent the individuals in two or three dimensions. In essence, the problem is to represent the relation between individuals such that those which are most similar in *n*-dimensions still appear to be closest in a lower dimensional space which can be visualised more easily. The most popular technique is Sammon mapping. This method relies on the ability to optimise an error function which relates original values in high dimensional space to the transformed values. This can be achieved using standard optimisation methods within packages such as Matlab. Multidimensional scaling can be useful in visualising the relative position of different individuals within a search space, for exploring variations in a multi-criteria objective function within a parameter space, or for comparing individual search paths within different simulations (Pohlheim 2006).

Eigenvector methods are another form of multi-dimensional scaling. Any multi-dimensional representation of data in *n*-dimensional space can be transformed into an equivalent space governed by *n* orthogonal eigenvectors. The main significance of this observation is that the principal eigenvector constitutes the most efficient way to represent a multi-dimensional space within a single value. For example, Moon, Schneider and Carley (2006) use the concept of 'eigenvector centrality' within a social network to compute a univariate measure of relative position based on a number of constituent factors.

Clustering techniques collapse multi-dimensional data so that individual cases are members of a single group or cluster, classified on the basis of a similarity metric. The method may therefore be appropriate if the modeller wishes to understand the distribution of an output variable in relation to the combination of several input variables. For example, Schreinemachers and Berger (2006) present a model of land-use in relation to farm characteristics such as soil quality, agricultural intensity, market orientation, water use, social capital and migration. In order to understand land use outcomes (for example, preference for crops versus livestock) farms were clustered into nine groups based on their shared characteristics. Cluster analysis is easy to implement in all the major statistics packages (SAS, SPSS, BMDP). The technique is likely to be most useful in empirical applications with a relatively large number of agent characteristics (i.e. six or more) than in idealised simulations with simple agent rules. One advantage of this technique over multidimensional scaling is that it is possible to represent statistical variation within the cluster space.

## 9.3 Individual Patterns, Novel Approaches and Visualisation

Plainly aggregate statistics like those above are a useful way of simplifying individual-level data, both in terms of complexity and dimensionality. However, they are the result of over 2,500 years of mathematical development in a research environment unsuited to the mass of detail associated with individual-level data. Now, computers place us in the position of being able to cope with populations of individual-level data at a much smaller scale. We still tend to place our own understanding at the end of an analytical trail, constraining the trail to pass through some kind of simplification and higher level of aggregation for the purposes of model analysis. Despite this, it is increasingly true that individual-level data is dealt with at the individual-level for the body of the analysis and this is especially true in the case of individual-level modelling, in which experimentation is almost always enacted at the individual-level. Whether it is really necessary to simplify for human understanding at the end of an analysis is not especially clear. It may well be that better techniques might be developed to do this than those built on an assumption of the necessity of aggregation.

At the individual-level, we are interested in recognising patterns in space and time, seeing how patterns at different scales affect each other, and then using this to say something about the behaviour of the system/individuals. Patterns are often indicators of the attractors to which individuals are drawn in any given system, and present a shortcut to understanding the mass of system interactions. However, it is almost as problematic to go through this process to understand a model as it is, for example, to derive individual-level behaviours from real large-size spatio-temporal datasets of socio-economic attributes. The one advantage we have in understanding a model is that we do have some grip on the foundation rules at the individual-scale. Nonetheless, understanding a rule, and determining how it plays out in a system of multiple interactions are very different things. Table 9.2 outlines some of the problems.

Our chief tool for individual-level understanding without aggregation is, and always has been, the human ability to recognise patterns in masses of data. Visualisation, for all its subjectivity and faults, remains a key element of the research process. The standard process is to present one or more attributes of the individuals in a map in physical or variable space. Such spaces can then be evolved in movies or sliced in either time or space (Table 9.3 shows some examples). In general, we cannot test the significance of a pattern without first recognising it exists, and to that extent significance testing is tainted by the requirement that it test

**Table 9.2** Issues related to understanding a model at different levels of complexity

| Complexity | Issues |
|---|---|
| Spatial | What is the impact of space (with whom do individuals initiate transactions and to what degree)? |
| Temporal | How does the system evolve? |
| Individuals | How do we recognise which individual behaviours are playing out in the morass of interactions? |
| Relationships | How do we recognise and track relationships? |
| Scale | How can we reveal the manner in which individual actions affect the large scale system and vice versa? |

our competency in recognising the correct pattern as much as that the proposed pattern represents a real feature of the distribution of our data. Visualisation is also a vital tool in communicating results within the scientific community and to the wider public. The former is not just important for the transmission of knowledge, but because it allows others to validate the work. Indeed, the encapsulation of good visualisation techniques within a model framework allows others to gain deeper understanding of one's model, and to experiment at the limits of the model – what Grimm (2002) calls "Visual Debugging". Good model design starts like the design of any good application, with an outline of what can be done to make it easy to use, trustworthy, and simple to understand. Traditionally, user interface design and visualisation have been low on the academic agenda, to the considerable detriment of both the science and the engagement of taxpayers. Fortunately, in the years since the turn of the millennium there has been an increasing realisation that good design engages the public, and that there is a good deal of social-science research that can be built on that engagement. Orford et al. (1999) identify computer graphics, multimedia, the World Wide Web, and virtual reality as four visualisation technologies that have recently seen a considerable evolution within the social sciences. There is an ever increasing array of visualisation techniques at our disposal: Table 9.3 presents a classification scheme of commonly used and more novel visualisation methods based on the dimensionality and type of data that is being explored.

Another classification scheme of these techniques that is potentially very useful comes from Andrienko et al. (2003). This classification categorises techniques based on their applicability to different types of data.

- "Universal" techniques that can be applied whatever the data, e.g. querying and animation.
- Techniques revealing existential change, e.g. time labels, colouring by age, event-lists and space-time cubes.
- Techniques about moving objects, e.g. trajectories, space-time cubes, snapshots in time.
- Techniques centered on thematic/numeric change, e.g. change maps, time-series and aggregations of attribute values.

**Table 9.3** Classification of visualisation methods according to dimensionality and type of data

| | Method | Pro | Con |
|---|---|---|---|
| Spatial 1D/2D | Map: overlay; animated trajectory representation (e.g. arrows); snapshots | View of whole trajectory of an object | Cannot analyse trajectory of movement. If several objects cross paths, cannot tell whether objects met at crossing point or visited points at different times |
| | Spatial distribution: e.g. choropleth maps | Gives a snapshot of an area | Cannot see how a system evolves through time. Aggregate view of area. Only represents one variable; hard to distinguish relationships |
| Temporal 1D | Time-series graphs/ linear and cyclical graphs | Show how the system (or parameters) change over time | No spatial element. Hard to correlate relationships between multivariate variables |
| | Rank clocks (e.g. Batty 2006) | Good for visualising change over time in ranked order of any set of objects | No spatial element |
| | Rose diagrams (e.g. Parry 2005) | Good for representation of circular data e.g. wind speed and direction | No spatial element |
| | Phase diagram | Excellent for examining system behaviour over time for one or two variables | No spatial element. Gets confusing quickly with more than two variables |
| Spatio-temporal 3D/4D | Map animation (e.g. Patel and Hudson-Smith 2012) | Can see system evolving spatially and temporally | Hard to quantify or see impacts of individual behaviour, i.e. isolated effects |
| | Space-time cube (Andrienko et al. 2003) | Can contain space-time paths for individuals | Potentially difficult to interpret |
| | Recurrence plot | Reveals hidden structures over time and in space | Computationally intensive. Methods difficult to apply. Have to generate multiple snapshots and run as an animation |
| | Vector plotting/ Contour slicing (Ross and Vosper 2003) | Ability to visualise 2D or 3D data and multiple dimensional data set | Hard to quantify individual effects |

For information on other visualisation schemes, see Cleveland (1983), Hinneburg et al. (1999) and Gahegan (2001).

In each case, the techniques aim to exploit the ease with which humans recognise patterns (Muller and Schumann 2003). Pattern recognition is, at its heart, a human attribute, and one which we utilise to understand models, no matter how we process the data. The fact that most model understanding is founded on a human recognition of a "significant" pattern is somewhat unfortunate, as we will bring our own biases to the process. At worst we only pay attention to those patterns that confirm our current prejudices: what Wyszomirski et al. (1999) call the *WYWIWYG* – What You Want is What You Get – fallacy. At best, we will only recognise those patterns that match the wiring of the human visual system and our cultural experiences. The existence of visualisation techniques generally points up the fact that humans are better at perceiving some patterns than others, and in some media than others – it is easier to see an event as a movie and not a binary representation of the movie file displayed as text. However, in addition to standard physiological and psychological restrictions on pattern recognition consistent to all people, it is also increasing apparent there are cultural differences in perceptions. Whether there is some underlying biological norm for the perception of time and space is still moot (Nisbett and Masuda 2003; Boroditsky 2001) but it is clear that some elements of pattern recognition vary by either culture or genetics (Nisbett and Masuda 2003; Chua et al. 2005). Even when one looks at the representation of patterns and elements like colour, there are clear arguments for a social influence on the interpretation of even very basic stimuli into perceptions (Roberson et al. 2004). Indeed, while there is a clear and early ability of humans to perceive moving objects in a scene as associated in a pattern (e.g. Baird et al. 2002), there are cultural traits associated with the age at which even relatively universal patterns are appreciated (Clement et al. 1970). The more we can objectify the process, therefore, the less our biases will impinge on our understanding. In many respects it is easier to remove human agents from data comparison and knowledge development than pattern hunting, as patterns are not something machines deal with easily. The unsupervised recognition of even static patterns repeated in different contexts is far from computationally solved (Bouvrie and Sinha 2007). Most pattern-hunting algorithms try to replicate the process found in humans and in that sense one suspects we would do better to skip the pattern hunting and concentrate on data consistency and the comparison of full datasets directly. At best we might say that an automated "pattern" hunter that wasn't trying to reproduce the human ability would instead seek to identify attractors within the data.

Figure 9.1 presents several visualisation methods that are commonly found in the literature, ranging from 1D time-series representation Fig. 9.1a to contour plots Fig. 9.1d that could be potentially used for 4D representation.

Visualisations are plainly extremely useful. Here we'll look at a couple of techniques that are of use in deciphering individual-level data: phase maps and recurrence plots. Both techniques focus on the representation of individual level states and the relationships between stated individuals.

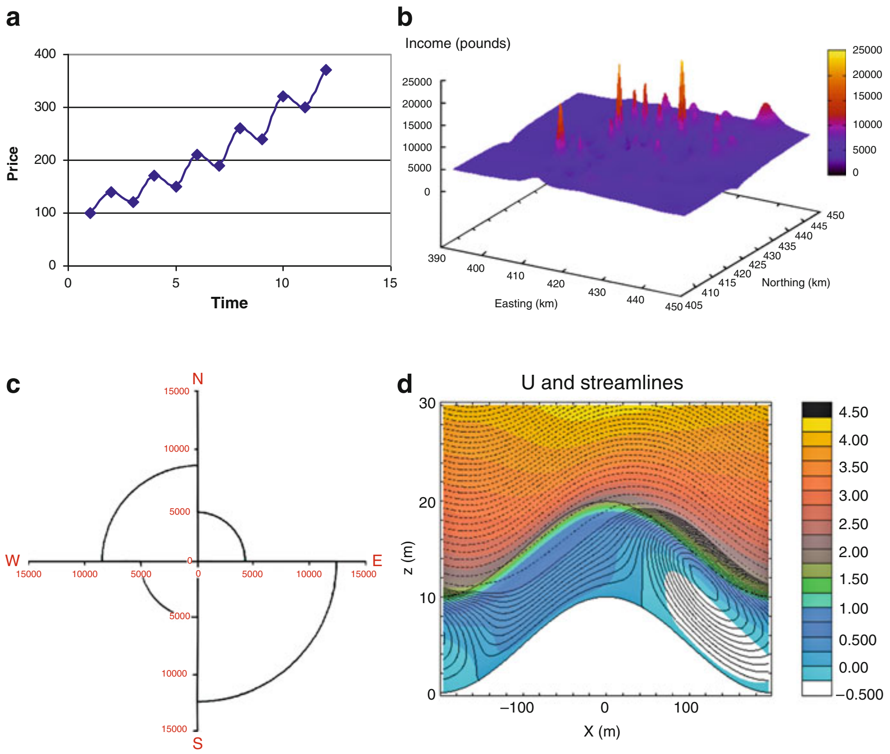



**Fig. 9.1** Examples of different visualisation methods. (**a**) 1D time-series graph (idealised data). (**b**) 3D interpolated map (idealised data). (**c**) Rose diagram. (**d**) Contour plot

### *9.3.1 Phase Maps*

Phase space maps are commonly used by physicists to study the behaviour of physical systems. In any graphical representation, a phase space map represents an abstract view of the behaviour of one or more of the system components. These can be particularly useful to us as we can plot the behaviour of our system over time. This allows us to understand how the system is evolving and whether it is chaotic, random, cyclical or stable (Fig. 9.2).

Each of the graphs produced in Fig. 9.2 are a representation of the coincident developments in two real neighbouring city centre petrol stations in Leeds (UK) over a 30 day period (sampled every other day). Figure 9.2a represents a stable system. Here, neither of the stations is changing in price and thus, a fixed point is produced. However, this behaviour could easily change if one or both of the stations alters it price. This behaviour is seen in Fig. 9.2b. Both stations are changing their prices each day (from 75.1p to 75.2p to 75.1p), this creates a looping effect; the stations are cycling through a pattern of behaviour before returning to their starting

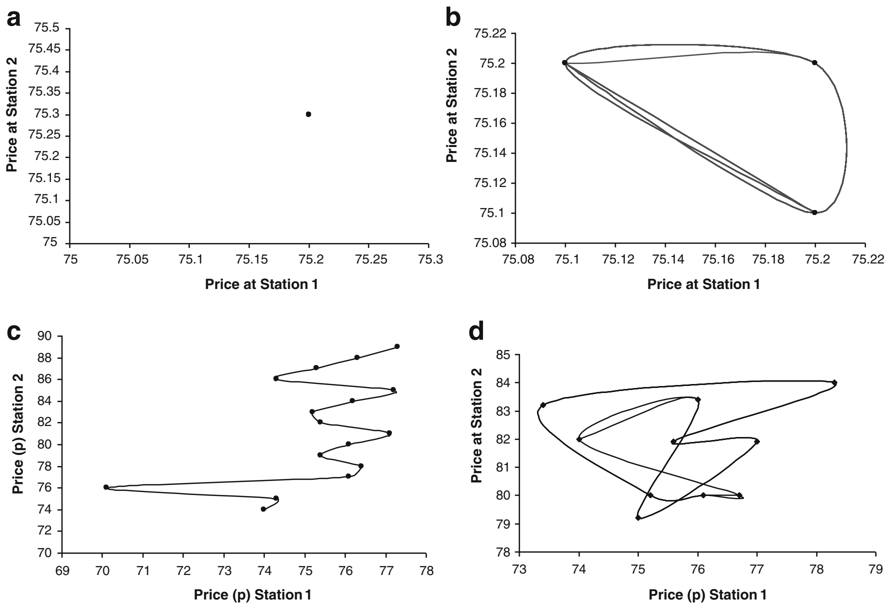



**Fig. 9.2** Examples of different types of behaviour found in urban petrol stations (Leeds). (**a**) Stable (**b**) Looping (**c**) Two types of behaviour (**d**) Chaotic

point. Note that the graph appears to reveal a causative link between the two stations as they are never simultaneously low. Figure 9.2c and d show a more varied pattern of behaviour between the stations. In Fig. 9.2c, one point is rising in price whilst the other is oscillating. In Fig. 9.2d, there is no apparent pattern in the displayed behaviour. Simply knowing about these relationships is valuable information and allows us a greater understanding of this system, its behaviour and its structure. For example, it may be that the only difference between the graphs is one of distance between the stations, but we would never see this unless the graphs allowed us to compare at a detailed level the behaviours of stations that potentially influence each other.

### *9.3.2 Recurrence Plots*

Recurrence plots (RPs) are a relatively new technique for the analysis of time series data that allows both visualisation and quantification of structures hidden within data or exploration of the trajectory of a dynamical system in phase space (Eckmann et al. 1987). They are particularly useful for graphically detecting hidden patterns and structural changes in data as well as examining similarities in patterns across a time series data set (where there are multiple readings at one point). RPs

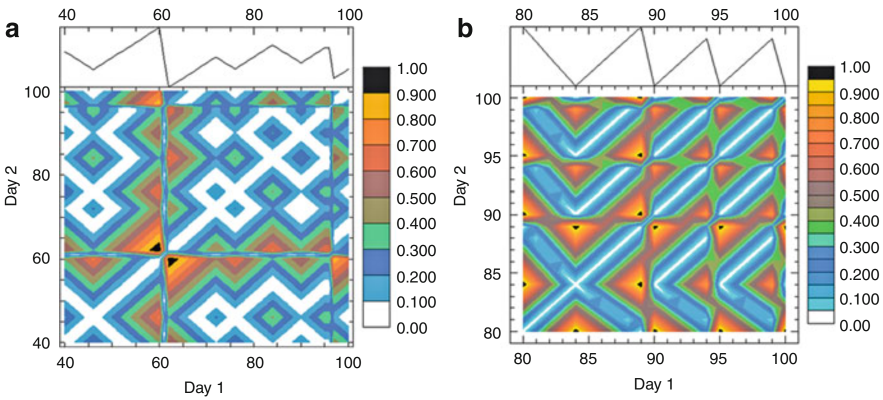



**Fig. 9.3** Example of recurrence plots

can be also used to study the nonstationarity of a time series as well as to indicate its degree of aperiodicity (Casadagli 1997; Kantz and Schreiber 1997). These features make RPs a very valuable technique for characterising complex dynamics in the time domain (Vasconcelos et al. 2006), a factor reflected in the variety of applications that RPs can now found in, ranging from climate variation (Marwan and Kruths 2002), and music (Foote and Cooper 2001) to heart rate variability (Marwan et al. 2002).

Essentially a RP is constructed via a matrix where values at a pair of time-steps are compared against each other. If the system at the two snapshots is completely different, the result is 1.0 (black), while completely similar periods are attributed the value 0.0 (represented as white). Through this, a picture of the structure of the data is built up. Figure 9.3a shows the RP of the change in price at a retail outlet over 100 days. Above the RP is a time-series graph diagrammatically representing the change in price. Changes in price, either increases, decreases, or oscillations, can be clearly seen in the RP. Figure 9.3b illustrates how oscillations in the change in the price data are represented in the RP.

Early work on this area has shown that there is considerable potential in the development and adaptation of this technique. Current research is focused on the development of cross-reference RPs (consideration of the phase space trajectories of two different systems in the same phase space) and spatial recurrence plots.

## 9.4 Explanation, Understanding and Causality

Once patterns are recognised, "understanding" our models involves finding explanations highlighting the mechanisms within the models which give rise to these patterns. The process of explanation may be driven with reference to current theory or developing new theory. This is usually achieved through:

1. Correlating patterns visually or statistically with other parts of the model, such as different geographical locations, or with simulations with different starting values;
2. Experimentally adjusting the model inputs to see what happens to the outputs;
3. Tracking the causal processes through the model.

It may seem obvious, and yet it is worth pointing out that model outputs can only causally relate to model inputs, not additional data in the real world. Plainly insights into the *system* can come from comparison with external data that is correlated or mis-correlated with model outputs, but this is not the same as understanding your model and the way it currently represents the system. One would imagine that this means that understanding of a model cannot be facilitated by comparing it with other, external, data, and yet it can often be worth:

4. Comparing model results with real world data,

because the relationships between real data and both model inputs and model outputs may be clearer than the relationships between these two things within the model.

Let's imagine, for example, a model that predicts the location of burglaries across a day in a city region where police corruption is rife. The model inputs are known offenders' homes, potential target locations and attractiveness, the position of the owners of these targets and the police, who prefer to serve the wealthy. We may be able to recognise a pattern of burglaries that moves, over the course of the day, from the suburbs to the city centre. Although we have built into our model the fact that police respond faster to richer people, we may find, using (1) that our model doesn't show less burglaries in rich areas, because the rich areas are so spatially distributed that the police response times are stretched between them. We can then alter the weighting of the bias away from the wealthy (2) to see if it actually reduces the burglary rate in the rich areas by placing police nearer these neighbourhoods as an ancillary effect of responding to poor people more. We may be able to fully understand this aspect of the model and how it arises (3), but still have a higher than expected burglary rate in wealthy areas. Finally, it may turn out (4) that there is a strong relationships between these burglaries and real data on petrol sales, for no other reason than both are high at transition times in this social system, when the police would be most stretched between regions – suggesting in turn that the change in police locations over time is as important as their positions at any one time.

Let us look at each of these methodologies for developing understanding in turn.

**Correlation**: Most social scientists will be familiar with linear regression as a means for describing data or testing for a relationship between two variables; there is a long scientific tradition of correlating data between models and external variables, and this tradition is equally applicable to intra-model comparisons. Correlating datasets is one of the areas where automation can be applied. As an exploratory tool, regression modelling has its attractions, not least its simplicity in both concept and execution. Simple regressions can be achieved in desktop applications like Microsoft Excel, as well as all the major statistical packages

(SAS, SPSS, BMDP, GLIM etc.). Standard methodologies are well known for cross-correlation of both continuous normal data and time series. However even for simple analyses with a single input and single output variable, linear regression is not always an appropriate technique. For example, logistic regression models will be more appropriate for binary response data, Poisson models will be superior when values in the dependent tend to be highly clustered, while binomial models may be the most effective when observations are highly dispersed around the mean. An interesting example is Fleming and Sorenson (2002) in which binomial estimates of technological innovation are compared to the complexity of the invention measured by both the number of components and the interdependence between those components. In behavioural space, methodologies such as Association Rule Making (e.g. Hipp et al. 2002) allow the Bayesian association of behavioural attributes. It is worth noting that where models involve a distribution in physical space this can introduce problems, in particular where the model includes neighbourhood-based behaviours and therefore the potential to develop spatial auto and cross-correlations. These alter the sampling strategies necessary to prove relationships – a full review of the issues and methodologies to deal with them can be found in (Wagner and Fortin 2005).

**Experimentation**: In terms of experimentation, we can make the rather artificial distinction between sensitivity testing and "what if?" analyses – the distinction is more one of intent than anything. In sensitivity analysis one perturbs model inputs slightly to determine the stability of the outputs, under the presumption that models should match the real world in being insensitive to minor changes (a presumption not always well founded). In "what if?" analyses, one alters the model inputs to see what would happen under different scenarios. In addition to looking at the output values at a particular time-slice, the stability or otherwise of the model, and the conditions under which this varies, also give information about the system (Grimm 1999).

**Tracking Causality**: Since individual-based models are a relatively recent development, there is far less literature dealing with the tracking of causality through models. It helps a little that the causality we deal with in models, which is essentially a mechanistic one, is far more concrete than the causality perceived by humans, which is largely a matter of the repeated co-incidence of events. Nevertheless, backtracking through a model to mark a causality path is extremely hard, primarily for two reasons. The first is what we might call the "find the lady problem" – that the sheer number of interactions involved in social processes tends to be so large we don't have the facilities to do the tracking. The second issue, which we might call the "drop in the ocean problem", is more fundamental as it relates to a flaw in the mathematical representation of objects, that is, that numbers represent aggregated quantities, not individuals. When transacted objects in a system are represented with numbers greater than one, it is instantly impossible to reliably determine the path taken by a specific object through that system. For objects representing concepts, either numerical (for example, money) or non-numerical (for example, a meme), this isn't a problem (one dollar is much like any other; there is only one YouTube to know). However, for most objects such aggregations place ambiguous nodes between what would otherwise be discrete causal pathways. Fortunately, we tend to use numbers in agent-models as a methodology to cope with our ignorance (for example, in the case of calibrated parameters), or the lack of the computing power we'd need to deal with individual objects and their transactional histories (for example, in the case of a variable like "number of children"). As it happens, every day brings improvements to both. It addition, the last 10 years or so has seen considerable theoretical advances in the determination of the probabilities of causation (for example, Pearl and Verma 1991; Greenland and Pearl 2006). For now, however, the tracking of causality is much easier if the models build in appropriate structures from the start. While they are in their infancy, techniques like Process Calculi (Worboys 2005) and Petri Nets show the potential of this area.

The inability to track causality leads to the perennial problem of identifiability, that is, that a single model outcome may have more than one history of model parameters that leads to it. Identifiability is part of a larger set of issues with confirming that the model in the computer accurately reflects the system in the real world – the so-called equifinality issue. These are issues that play out strongly during model construction from real data and when validating a model against real data, and a review of techniques to examine these problems, including using model variation to determine the suitability of variables and parameters, can be found in Evans (2012). At the model stage we are interested in, however, we at least have the advantage that there is only one potential model that may have created the output – the one running. Nevertheless, the identifiability of the parameters in a running model still makes it hard to definitively say when model behaviour is reasonable. For those modelling for prediction, this is of little consequence – as long as the model gives consistently good predictions it may as well be a black-box. However, if we wish to tease the model apart and look at how results have emerged, these issues become more problematic.

The mechanisms for dealing with these problems are pragmatic.

1. Examine the stability of the calibration process and/or the state of internal variables that weren't inputs or outputs across multiple runs.
2. Validate internal variables that weren't inputs or outputs used in any calibration against real data.
3. Run the model in a predictive mode with as many different datasets as possible – the more the system can replicate reality at output, the more likely it is to replicate reality internally. If necessary engage in inverse modelling: initialize parameters randomly then adjust them over multiple runs until they match all known outputs.

Of these, by far the easiest, but the least engaged with, is checking the stability of the model in parameter space (see Evans 2012 for a review). Various AI techniques have been applied to the problem of optimizing parameters to fit model output distributions to some pre-determined pattern (such as a 'real world' distribution). However, the stability of these parameterizations and the paths AIs take to generate them are rarely used to examine the degree to which the model fluctuates between different states, let alone to reflect on the nature of the system. The assumption of equifinality is that the more parameterized a model, the more likely it is a set of parameter values can be derived which fit the data but don't represent the true values. However, in practice the limits on the range of parameter values within any given model allows us an alternative viewpoint: that the more parameterized rules in a model, the more the system is constrained by the potential range of the elements in its structure and the interaction of these ranges. For example, a simple model $a = b$ has no constraints, but $a = b/c$, *where* $c =$ *distance between a and b*, adds an additional constraint even though there are more parameters. As such rules build up in complex systems, it is possible that parameter values become highly constrained, even though, taken individually, any given element of the model seems reasonably free. This may mean that if a system is well modelled, exploration of the model's parameter space by an AI might reveal the limits of parameters within the constraints of the real complex system. For example, Heppenstall et al. (2007) use a Genetic Algorithm to explore the parameterisation of a petrol retail model/market, and find that while some GA-derived parameters have a wide range, others consistently fall around specific values that match those derived from expert knowledge of the real system.

The same issues as hold for causality hold for data uncertainty and error. We have little in the way of techniques for coping with the propagation of either through models (see Evans 2012 for a review). It is plain that most real systems can be perturbed slightly and maintain the same outcomes, and this gives us some hope that errors at least can be suppressed, however we still remain very ignorant as to how such homeostatic forces work in real systems, and how we might recognise or replicate them in our models. Data and model errors can breed patterns in our model outputs. An important component of understanding a model is understanding when this is the case. If we are to use a model to understand the dynamics of a real system and its emergent properties, then we need to be able to recognise novelty in the system. Patterns that result from errors may appear to be novel (if we are lucky), but as yet there is little in the way of toolkits to separate out such patterns from truly interesting and new patterns produced intrinsically.

Currently our best option for understanding model artifacts is model-to-model comparisons. These can be achieved by varying one of the following contexts while holding the others the same: the model code (the model, libraries, and platform), the computer the model runs on, or the data it runs with (including internal random number sequences). Varying the model code (for instance from Java to C++, or from an object-orientated architecture to a procedural one) is a useful step in that it ensures the underlying theory is not erroneously dependent on its representation. Varying the computer indicates the level of errors associated with issues like rounding and number storage mechanisms, while varying the data shows the degree to which model and theory are robust to changes in the input conditions. In each case, a version of the model that can be transferred between users, translated onto other platforms and run on different data warehouses would be useful. Unfortunately, however, there is no universally recognised mechanism for representing models abstracted from programming languages. Mathematics, UML, and natural languages can obviously fill this gap to a degree, but not in a manner that allows for complete automatic translation. Even the automatic translation of computer languages is far from satisfactory when there is a requirement that the results be understood by humans so errors in knowledge representation can be checked. In addition, many such translations work by producing the same binary executable. We also need standard ways of comparing the results of models, and these are no more forthcoming. Practitioners are only really at the stage where we can start to talk about model results in the same way (see, for example, Grimm et al. 2006). Consistency in comparison is still a long way off, in part because statistics for model outputs and validity are still evolving, and in part because we still don't have much idea which statistics are best applied and when (for one example bucking this trend see Knudsen and Fotheringham 1986).

## 9.5 Future Directions

Recognising patterns in our modelled data allows us to:

1. Compare it with reality for validation.
2. Discover new information about the emergent properties of the system.
3. Make predictions.

Of these, discovering new information about the system is undoubtedly the hardest, as it is much easier to spot patterns you are expecting. Despite the above advances, there are key areas where current techniques do not match our requirements. In particular, these include:

1. Mechanisms to determine when we do not have all the variables we need to model a system and which variables to use.
2. Mechanisms to determine which minor variables may be important in making emergent patterns through non-linearities.
3. The tracking of emergent properties through models.
4. The ability to recognise all but the most basic patterns in space over time.
5. The ability to recognise action across distant spaces over space and time.
6. The tracking of errors, error acceleration, and homeostatic forces in models.

While we have components of some of these areas, what we have is but a drop in the ocean of techniques we need. In addition, the vast majority of our techniques are built on the 2,500 years of mathematics that resolved to simplify systems that were collections of individuals because we lacked the ability (either processing power or memory) to cope with the individuals *as* individuals. Modern computers have given us this power for the first time and, as of yet, the ways we describe such systems have not caught up, even if we accept that some reduction in dimensionality and detail is necessary for a human to understand our models. Indeed in the long-run it might be questioned whether the whole process of model understanding and interpretation might be divorced from humans, and delegated instead to an artificially intelligent computational agency that can better cope with the complexities directly.

## Further Reading

Statistical techniques for spatial data are reviewed by McGarigal (2002) while for network statistics good starting points are (Newman 2003) and (Boccaletti et al. 2006), with more recent work reviewed by Evans (2010). For information on coping with auto/cross-correlation in spatial data, see (Wagner and Fortin 2005). Patel and Hudson-Smith (2012) provide an overview of the types of simulation tool (virtual worlds and virtual reality) available for visualising the outputs of spatially explicit agent-based models. Evans (2012) provides a review of techniques for analysing error and uncertainty in models, including both environmental/climate models and what they can bring to the agent-based field. He also reviews techniques for identifying the appropriate model form and parameter sets.

## References


Andrienko N, Andrienko G, Gatalsky P (2003) Exploratory spatio-temporal visualisation: an analytical review. J Vis Lang Comput 14(6):503–541

Baird AA et al (2002) Frontal lobe activation during object permanence: data from near-infrared spectroscopy. Neuroimage 16:1120–1126

Batty, M (2006) Rank Clocks. Nature, 444: 592–596

Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks: structure and dynamics. Phys Rep 424(4–5):175–308

Boroditsky L (2001) Does language shape thought? Mandarin and English Speakers' conceptions of time. Cogn Psychol 43:1–22

Bouvrie JV, Sinha P (2007) Visual object concept discovery: observations in congenitally blind children, and a computational approach. Neurocomputing 70(13–15):2218–2233

Casdagli M (1997) Recurrence plots revisited. Phys D 108(1–2):12–44

Chua HF, Boland JE, Nisbett RE (2005) Cultural variation in eye movements during scene perception. Proc Natl Acad Sci USA 102(35):12629–12633

Clark PJ, Evans FC (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 35(4):445–453

Clement DE, Sistrunk F, Guenther ZC (1970) Pattern perception among Brazilians as a function of pattern uncertainty and age. J Cross Cult Psychol 1(4):305–313

Cleveland WS (1983) Visualising data. Hobart Press, Summit

David N (2013) Validating simulations. Chapter 8 in this volume

Eckmann JP, Kamphorst SO, Reulle D (1987) Recurrence plots of dynamical systems. Europhys Lett 4(9):973–977

Evans AJ (2010) Complex spatial networks in application. Complexity 16(2):11–19

Evans AJ (2012) Uncertainty and error. In: Heppenstall AJ, Crooks AT, See LM, Batty M (eds) Agent-based models of geographical systems. Springer, Berlin, chapter 15

Fisher N, Lewis T, Embleton B (1987) Statistical analysis of spherical data. Cambridge University Press, Cambridge

Fleming L, Sorenson O (2001) Technology as a complex adaptive system: evidence from patent data. Research Policy 30: 1019–1039

Foote J, Cooper M (2001) Visualising music structure and rhythm via self-similarity. In: Proceedings of the international computer music conference, ICMC'01, Havana. ICMA, San Francisco, pp 419–422

Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, Chichester

Gahegan M (2001) Visual exploration in geography: analysis with light. In: Miller HJ, Han J (eds) Geographic data mining and knowledge discovery. Taylor & Francis, London, pp 260–287

Gehlke CE, Biehl H (1934) Certain effects of grouping upon the size of correlation coefficients in census tract material. J Am Stat Assoc 29(2):169–170

Getis A (2007) Reflections on spatial autocorrelation. Reg Sci Urban Econ 37(4):491–496

Graps A (2004) Amara's wavelet page. Access date 8 Jun 2011. http://web.archive.org/web/20110608005544/http://www.amara.com/current/wavelet.html

Greenland S, Pearl J (2006) Causal diagrams (Technical report, R-332). UCLA Cognitive Systems Laboratory, Los Angeles. Access date 8 June 2011 http://ftp.cs.ucla.edu/pub/stat_ser/r332.pdf

Grimm V (1999) Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? Ecol Model 115(2):129–148

Grimm V (2002) Visual debugging: a way of analyzing, understanding, and communicating bottom-up simulation models in ecology. Nat Resour Model 15:23–38

Grimm V et al (2006) A standard protocol for describing individual-based and agent-based models. Ecol Model 198(1–2):115–126

Haining R (1990) Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge

Heppenstall AJ, Evans AJ, Birkin MH (2006) Using hybrid agent-based systems to model spatially-influenced retail markets. J Artif Soc Soc Simul 9(3). http://jasss.soc.surrey.ac.uk/9/3/2.html

Heppenstall AJ, Evans AJ, Birkin MH (2007) Genetic algorithm optimisation of a multi-agent system for simulating a retail market. Environ Plann B 34(6):1051–1070

Hinneburg A, Keim DA, Wawryniuk M (1999) HD-eye: visual mining of high-dimensional data. IEEE Comput Graph Appl 19(5):22–31

Hipp J, Güntzer U, Nakhaeizadeh G (2002) Data mining of association rules and the process of knowledge discovery in databases. In: Perner P (ed) Advances in data mining, vol 2394, Lecture notes in computer science. Springer, Berlin, pp 207–226

Isaaks EH, Srivastava RM (1990) Applied geostatistics. Oxford University Press, Cary

Kantz H, Schreiber T (1997) Non-linear time series analysis. Cambridge University Press, Cambridge

Knudsen DC, Fotheringham AS (1986) Matrix comparison, goodness-of-fit, and spatial interaction modelling. Int Reg Sci Rev 10:127–147

Korie S et al (1998) Analysing maps of dispersal around a single focus. Environ Ecol Stat 5(4):317–344

Marwan N, Kruths J (2002) Nonlinear analysis of bivariate data with cross recurrence plots. Phys Lett A 302(5–6):299–307

Marwan N, Wessel N, Meyerfeldt U, Schirdewan A, Kurths J (2002) Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. Phys Rev E 66(2):026702

McGarigal K (2002) Landscape pattern metrics. In: El-Shaarawi AH, Piegorsch WW (eds) Encyclopedia of environmentrics, vol 2. Wiley, Chichester, pp 1135–1142

Moon I-C, Schneider M, Carley K (2006) Evolution of player skill in the America's army game. Simulation 82(11):703–718

Müller W, Schumann HS (2003) Visualisation methods for time-dependent data: an overview. In: Chick S, Sánchez PJ, Ferrin D, Morrice DJ (eds) Proceedings of winter simulation 2003, New Orleans, 7–10 Dec 2003. http://informs-sim.org/wsc03papers/090.pdf

Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167–256

Nisbett RE, Masuda T (2003) Culture and point of view. Proc Natl Acad Sci USA 100(19):11163–11170

OED (2010) Oxford English dictionary. Access date 21 Dec 2010 http://www.oed.com/

Openshaw S, Craft AW, Charlton M, Birch JM (1988) Investigation of leukaemia clusters by use of a geographical analysis machine. Lancet 331(8580):272–273

Orford S, Harris R, Dorling D (1999) Geography: information visualisation in the social sciences. Soc Sci Comput Rev 17(3):289–304

Parry H (2005) Effects of Land Management upon Species Population Dynamics: A Spatially Explicit, Individual-based Model. Unpublished PhD thesis, University of Leeds. pp. 281

Patel A, Hudson-Smith A (2012) Agent tools, techniques and methods for macro and microscopic simulation. In: Heppenstall AJ, Crooks AT, See LM, Batty M (eds) Agent-based models of geographical systems. Springer, Berlin, chapter 18

Pearl J, Verma TS (1991) A theory of inferred causation. In: Allen JA, Fikes R, Sandewall E (eds) Proceedings of the 2nd international conference on principles of knowledge representation and reasoning (KR'91), Cambridge, MA, 22–25 Apr 1991. Morgan Kaufmann, San Mateo, pp 441–452

Pohlheim H (2006) Multidimensional scaling for evolutionary algorithms: visualisation of the path through search space and solution space using Sammon mapping. Artif Life 12(2):203–209

Ramsey JB (2002) Wavelets in economics and finance: past and future. Stud Nonlinear Dynam Econ 6(3):1–27

Roberson D, Davidoff J, Davies IRL, Shapiro LR (2004) The development of color categories in two languages: a longitudinal study. J Exp Psychol Gen 133(4):554–571

Robinson WS (1950) Ecological correlations and the behaviour of individuals. Am Sociol Rev 15: 351–357

Ross AN, Vosper SB (2003) Numerical simulations of stably stratified flow through a mountain pass. Q J R Meteor Soc 129:97–115

Schelling TC (1969) Models of segregation. Am Econ Rev 59(2):88–493

Schreinemachers P, Berger T (2006) Land-use decisions in developing countries and their representation in multi-agent systems. J Land Use Sci 1(1):29–44

Tobler WR (1970) A computer model simulation of urban growth in the Detroit region. Econ Geogr 46(2):234–240

Vasconcelos DB, Lopes SR, Kurths J, Viana RL (2006) Spatial recurrence plots. Phys Rev E 73(5): 056207

Viboud C et al (2006) Synchrony, waves, and spatial hierarchies in the spread of influenza. Science 312(5772):447–451

Wagner HH, Fortin M-J (2005) Spatial analysis of landscapes: concepts and statistics. Ecology 86:1975–1987

Worboys MF (2005) Event-oriented approaches to geographic phenomena. Int J Geogr Info Sci 19(1):1–28

Wyszomirski T, Wyszomirska I, Jarzyna I (1999) Simple mechanisms of size distribution dynamics in crowded and uncrowded virtual monocultures. Ecol Model 115(2–3):253–273

# Chapter 10
# Participatory Approaches

**Olivier Barreteau, Pieter Bots, Katherine Daniell, Michel Etienne, Pascal Perez, Cécile Barnaud, Didier Bazile, Nicolas Becu, Jean-Christophe Castella, William's Daré, and Guy Trebuil**

O. Barreteau (✉)
IRSTEA, UMR G-EAU, 361 rue Jean-François Breton, BP 5095, Montpellier 34196, France
e-mail: olivier.barreteau@irstea.fr

P. Bots
Delft University of Technology, Faculty of Technology, Policy and Management, Delft 50152600 GA, The Netherlands
e-mail: p.w.g.bots@tudelft.nl

K. Daniell
Centre for Policy Innovation, The Australian National University, Canberra ACT 0200, Australia
e-mail: katherine.daniell@anu.edu.au

M. Etienne
Ecodevelopment UnitDomaine St-Paul, site Agroparc, Avignon cedex 9 84914, France
e-mail: etienne@avignon.inra.fr

P. Perez
SMART, University of Wollongong, Northfields Ave, Wollongong NSW 2522, Australia
e-mail: pascal_perez@uow.edu.au

C. Barnaud
INRA, Centre INRA de Toulouse UMR Dynafor Chemin de Borde Rouge, Castanet Tolosan Cedex BP 5262731326, France
e-mail: cecile.barnaud@toulouse.inra.fr

D. Bazile • W. Daré • G. Trebuil
Cirad GREEN, TA C-47/F. Campus international de Baillarguet, Montpellier Cedex 5 34398, France
e-mail: didier.bazile@cirad.fr; dare@cirad.fr; guy.trebuil@cirad.fr

N. Becu
CNRS, Laboratoire de géographie PRODIG 2, rue Valette, Paris 75005, France
e-mail: nicolas.becu@univ-paris1.fr

J.-C. Castella
Institut de Recherche pour le Développement (IRD), UMR 220 GRED (IRD UPV Montpellier 3), PO Box 5992, Vientiane, Laos
e-mail: j.castella@ird.fr

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_10,

**Why Read This Chapter?** To help you understand how one might involve stakeholders in all stages of the modelling process. This approach allows for including stakeholders' expertise as well as giving them more control over the process.

**Abstract** This chapter aims to describe the diversity of participatory approaches in relation to social simulations, with a focus on the interactions between the tools and participants. We consider potential interactions at all stages of the modelling process: conceptual design; implementation; use; and simulation outcome analysis. After reviewing and classifying existing approaches and techniques, we describe two case studies with a focus on the integration of various techniques. The first case study deals with fire hazard prevention in southern France, and the second one with groundwater management on the Atoll of Kiribati. The chapter concludes with a discussion of the advantages and limitations of participatory approaches.

## 10.1 Introduction

In this chapter, social simulation is cross-examined with a currently very active trend in policy making: participation or stakeholder involvement. This cross-examination has two main outputs: the development of tools and methods to improve or facilitate participation; and the development of more grounded simulation models through participatory modelling. Technological development provides new devices to facilitate interaction around simulation models: from the phase of conceptual design to that of practical use. In many fields there is a growing requirement from stakeholders and the public to become more actively involved in policy making and to be aware of probable changing trends due to global policy decisions. New tools and methods related to social simulation have started to be made available for this purpose such as many Group Decision Support Systems which use computer simulation, including potentially social items components, to facilitate communication to formulate and solve problems collectively (DeSanctis and Gallupe 1987; Shakun 1996; Whitworth et al. 2000). In addition, simulation of social complexity occurs in models whose validation and suitability depend on their close fit to society, as well as on their acceptability by it. These issues are tackled through the use of participatory modelling, such as group model building (Vennix 1996) or participatory agent based simulations (Bousquet et al. 1999; Guyot and Honiden 2006; Moss et al. 2000; Pahl-Wostl and Hare 2004; Ramanath and Gilbert 2004). The topic is also related to participatory design as it is a mean of involving end-users of computer systems in their design, including social simulations focussed ones (Schuler and Namioka 1993).

Group Decision Support as well as Participatory Modelling stem from the interactions between simulation models and participants. There is a diversity of ways though which these interactions might take place. They are related to the diversity of approaches to simulate society or to organise participation. It is important to make the choices made for these interactions explicit: for distinction between approaches to be possible; to provide the opportunity for stakeholders to discuss the process; and for them to be prepared to be involved in. There is a need to go further than the development of tools as they are liable to create filters that reshape the understanding of social complexity. Description of the mechanisms behind interactions is a way to qualify the potential effects of these interactions.

This chapter aims to describe the diversity of participatory approaches in relation to social simulations, with a focus on the interactions between the tools and participants. This overview is limited to simulation models. Model is considered here as a representation of shared knowledge, which means the gathering of pieces of knowledge and assumptions about a system, written altogether in a model so that they might play or work together. We limit this scope further to simulation model, hence models including the representation of dynamics. We consider here potential interactions among participatory and modelling processes at all stages of the modelling process: conceptual design; implementation; use; and simulation outcome analysis.

The first section of this chapter outlines a number of factors which have paved the way for development of the association between social simulation and participation. There is a large body of literature in which authors have developed their own participatory modelling approaches, justified by some specific expectations on participation for modelling or vice-versa. This first section makes a synthesis of these expectations and draws out some principles on which various participatory modelling settings should be assessed. The second section describes some existing techniques and approaches. The third section proposes a classification of these participatory approaches according to three dimensions: the level of involvement in the process; the timeliness of involvement; and the heterogeneity of population involved. The fourth section describes two case studies with a focus on the integration of various techniques. We discuss the advantages of these approaches but also some limits, according to the expectations and in comparison with more traditional techniques in the fifth section.

## 10.2 Expectations of Using Participatory Approaches with Simulation of Social Complexity

Joint use of participatory approaches with social simulations is based upon three categories of expectations. They vary according to the target of the expected benefits of the association:

1. Quality of the simulation model per se;
2. Suitability of the simulation model for a given use; and
3. Participation support.

These three targets are linked to three different components of a modelling process. Target one is linked to the output, target three to the source system, and target two to the relation between both the output and source system. In this section we further develop these three categories.

## *10.2.1 Increasing Quality of Simulation Models of Social Complexity*

The objective here is to produce a good quality model to simulate social complexity. Participation is then pragmatically assumed to be a means for improving this quality. There is no normative belief which would value participation by itself in this category of expectations.

Quality of the simulation model is understood here rather classically with the following indicators:

– Realism: is the simulation model able to tackle key features of the social complexity it aims to represent?
– Efficiency: is the simulation model representing its target system with a minimum of assumptions and minimal simulation run-times?

Quality of the representation according to its use is another classical indicator of a simulation model's quality. It is specifically tackled in the following subsection.

#### 10.2.1.1 Taking Social Diversity and Capacity to Evolve into Account

One of the key features to be taken into account when representing a social system is to deal with its diversity. This diversity is related not only to individual characteristics, but also to viewpoints, expectations towards the system, and positions in the decision making processes. Dealing with diversity in simulation of social complexity involves embracing it as well as profiting by its existence.

Classically, dealing with diversity is a process of aggregation or selection. Aggregation consists of the identification of classes of individuals and representatives for them. Selection consists of choosing a few cases with all of their characteristics. This may lead to very simple simulation models with a generic diversity. Aggregation is rather greedy on data and modelling time and is still dependent on the viewpoint of the observers who provide the information leading to the categorisation. Selection is weak to cope with relations among various sources of diversity.

Involvement of stakeholders in the modelling process allows them to bring their own diversity. Concerns over representation are then transferred onto the constitution of the sample of participants. Fischer and colleagues have shown through development of situations to support creativity in various fields, such as art, open source development and urban planning, that diversity, as well as complexity, is important to enhance creativity (Fischer et al. 2005). This creativity is expected to pave the way for surprises in the simulation model.

Involvement of stakeholders in the modelling process is a way to externalise part of this diversity outside the model towards a group of stakeholders. The issue is then to work on the relation between the model and a number of stakeholders to allow a transfer of knowledge and ideas.

Social systems are open and evolving. Their definition depends on the viewpoint of the analyst. As far as simulation is concerned, this means depending on the viewpoint of the model designer(s). This choice means framing: cutting a number of links around the boundaries of the system studied, as well as around the interpretation which might occur based on the simulation outcomes (Dewulf et al. 2006). Firstly, participation provides the opportunity to consider problem boundaries which would be plurally defined, increasing the potential coherence of the model. However, it is still an operation of cutting links out of the real world situation, even though these chosen cuttings are more grounded and discussed. Secondly, interactive use of a simulation model is a means to keep some of these links open and active, with participants as driving belts. Stakeholders are embedded in social networks which cross the boundaries into the physical and environmental networks. They make the links come alive, which allows them to function and be updated.

There is thus a need to question the boundaries set in the interactive setting: actors in the neighbourhood; concerns of actors connected to those tackled by the (simulation) model; and how these relations are to be mobilised in the interaction.

#### 10.2.1.2 Distribution of Control

A key characteristic of social systems which is to be addressed through social simulation is their complexity. This complexity leads to various consequences, such as the emergence of phenomena, delay effects or discontinuities in some trends, which are present in social systems as in any complex systems. These are usually the effects which one likes to discover or better understand when experimenting with social simulations. From the internal point of view of simulations, Schelling has shown experimentally that reproducing settings with multiple decision centres improves the quality of representation of complexity (Schelling 1961). He could generate complexity through experimental games because of the presence of independent decision centres, the players. This result has also been shown with simulations used for forecasting (Green 2002). Green compared the capacity of forecasting the outcome of past social conflicts with: a role playing game with students; game theorists; and a group of experts. He compared the simulated outcomes with those from the real negotiations and found that the role playing game setting produced the best results. This was the one with the main distribution of decisions among autonomous centres.

The purpose of associating participatory processes and social simulation here is then to increase the complexity through interactive use or implementation of a social model. Unless computational agents are effectively used, which is rare (Drogoul et al. 2003), formal theories of complex systems that are completely embedded in a simulation model do not simulate complex patterns but implement an explanation of a complex pattern. In other words, they should be implemented in a distributed setting with autonomous entities. Participatory approaches provide such settings. There is then an issue of a deep connection between a simulation model and participants in a participatory modelling setting.

### *10.2.2 Improving Suitability of Simulation Model's Use*

Quality of a model is also assessed according to its suitability for its intended use. In this subsection, two cases of use are considered: knowledge increase; and policy making. In both cases, it is expected that involvement of stakeholders at any stage of a modelling process will aid better tuning of the model with its intended use: either through interactions with people represented in the model, or with potential users. Both cases have a major concern with making viewpoints explicit.

#### 10.2.2.1 Case of Increasing Knowledge

The case of use for knowledge increase builds upon the previous subsection. The key element treated here deals with the uncertainty of social systems. The involvement of stakeholders represented in the simulation model is a way to improve its validation or calibration. Participants may bring their knowledge to reduce or better qualify some uncertainties. The simulation model is then expected to give back to the participants simulation outputs based on the interactions between their pieces of knowledge. On the other hand, this feedback is sometimes difficult to validate (Manson 2002). Its presentation and discussion with stakeholders represented in the simulation model is a way to cope with this issue. This approach has been explored by Barreteau and colleagues to improve the validation of an Agent Based Model of irrigated systems in Senegal River valley (Barreteau and Bousquet 1999). The format of this feedback, information provided and medium of communication, might make the model really open to discussion.

This joins another expectation which is probably the most common in work that has so far implemented such participatory approaches with a social simulation model: making each participant's assumptions explicit, included the modellers (Fischer et al. 2005; Moss et al. 2000; Pahl-Wostl and Hare 2004). This is a requirement from the simulation modelling community: making stakeholders' beliefs, points of view and tacit knowledge explicit (Barreteau et al. 2001; Cockes and Ive 1996; D'Aquino et al. 2003; McKinnon 2005). Moreover, so that participants might become part of the model, the assumptions behind the model should be made explicit in order to be discussed, as should the outputs of the simulations so that they can also be discussed, transferred and translated in new knowledge. This is to overcome one major pitfall identified with the development of models which is the under-use of decision support models because of their opacity (Loucks et al. 1985; Reitsma et al. 1996). This concern of making explicit assumptions in the modelling process is also at the heart of the participatory approach community. One aim of gathering people together and making them collectively discuss their situation in a participatory setting is to make them aware of others' viewpoints and interests. This process involves and stimulates some explanation of tacit positions.

This means that the interactive setting should allow a bi-directional transfer of knowledge between stakeholders and the simulation model: knowledge elicitation in one direction and validation and explanation of simulation outputs in the other direction.

#### 10.2.2.2 Case of Policy Making

In the case of simulation focusing on policy issues, there is a pragmatic, moral, and now sometimes legal need to involve stakeholders, which may lead to open the black box of models of social complexity used in policy making. Post-normal approaches aim at making the decision process and its tools explicit so that stakeholders can better discuss it and appropriate its outcomes. When this decision process involves the use of decision support tools, which might include social simulation models, this means that the models themselves should be opened to stakeholders (Funtowicz et al. 1999). A simulation model is then expected to be explicit enough so that stakeholders who might be concerned by the implementation of the policy at stake could discuss it. This legitimisation is socially based, while validation, as mentioned with the previous case of use, is scientifically based (Landry et al. 1996). Even though validation is still required in this case of use, because it is the mode of evaluation for some participants, it is rather the legitimisation of the model by the stakeholders which is to be worked out.

Participatory approaches may be a means for opening these models to stakeholders, provided that formats of communication of models' assumptions and structure can be genuinely discussed. Involvement of stakeholders is expected to raise their awareness of the assumptions of the model and potentially able to discuss these and modify them. This includes the evolution of underlying values and choices made in the design of model.

### *10.2.3 Simulation as a Means to Support Participation*

Social simulation might also benefit to participation. While the previous subsection was dedicated to appropriateness between the model and its use as a group decision support tool, we focus here on participation which might be a component of a decision making process.

Social simulation is seen here as an opportunity to foster participation and cope with some of its pitfalls (Eversole 2003). Use of simulation models may lead to some outcomes such as community building or social learning.

#### 10.2.3.1 Dynamics and Uncertainties

Social systems have to deal with uncertainties just as social simulation models do. This might hamper participatory processes: in wicked problems (Rittel and Webber 1973), encountered in many situations where participatory processes are organised, stakeholders always maintain the opportunity related to these uncertainties to challenge others' viewpoints or observations. As an example: origin, flow and consequences of non point source pollution are uncertain. This leads some farmers to challenge the accusation, made by domestic water companies downstream of their fields, that they are polluting their sources. Sometimes, disparate viewpoints do not conflict. The gathering of these disparate pieces of knowledge is a way to reduce uncertainty and allow the group of stakeholders involved in a participatory process to progress; provided that they can work together.

Another characteristic of any social system which might hamper participation is its dynamicity. Socio-ecological systems exhibit a range of dynamics; not only social, but also natural, which evolve at various paces. In the application developed by Etienne and colleagues in Causse Mejan, pine tree diffusion has a typical time step of 20 years which is long according to the typical time steps of land use choices and assessment (Étienne et al. 2003). In a participatory process it might be difficult to put these dynamics on the agenda. Simulation models are known to be good tools to deal with dynamic systems.

Simulation models are therefore a means to gather distributed pieces of knowledge among stakeholders and to cope with scenarios in the face of uncertainties. They can also help make the participants aware of potential changes or regime shifts generated by their interactions (Kinzig et al. 2006).

#### 10.2.3.2 Towards Social Learning

Participation is often linked with the concept of social learning (Webler et al. 1995). However, for social learning to occur, participants should have a good understanding of their interdependencies as well as of the system's complexity. Social simulation can provide these bases, provided that the communication is well developed (Pahl-Wostl and Hare 2004).

This learning comes from exchanges among stakeholders involved in the participatory process but also from new knowledge which emerges in the interaction. Externalisation of tacit knowledge in boundary objects (Star and Griesemer 1989) is useful for both: it facilitates communication in giving a joint framework to make one's knowledge explicit; and it enhances individual, as well as social, creativity (Fischer et al. 2005).

Simulation models are good candidates to become such boundary objects. Agent based models have long been considered as blackboards upon which various disciplines could cooperate (Hochman et al. 1995). Through simulation outputs, they provide the necessary feedback for reflexivity, be it individual or collective.

The question then remains whether such models constrain the format of knowledge which might be externalised.

### 10.2.4 *Synthesis: A Key Role of the Interaction Pattern Between Model and Stakeholders*

These three categories of expectations have led to specific requests for the development of participation in relation to social simulation models. In the following section, we provide an overview of these techniques. On the basis of the previous requests, these techniques and methods have to be analysed according to the following dimensions:

– Set of connections between the participation arena and simulation model: its structure, its content, and organisation of its mobilisation;
– Control of the process; and
– Format of information which can travel from one pole to another: openness and suitability to the diversity of stakeholders' competencies.

## 10.3 A Diversity of Settings

In this section, we describe some examples of participatory techniques and approaches associated with social simulation models. Settings described in this overview stem from various fields and disciplines. Most of these have already produced some reviews on participatory approaches. For the purpose of the discussion in relation with social simulation, a synthesis of these reviews is provided here with a focus on the requests identified in the previous section.

### 10.3.1 *From System Science and Cybernetics*

Cybernetics and system sciences have produced a first category of simulation models of social complexity (Gilbert and Troitzsch 1999). These models are based on tools originating from system dynamics, using specific software. They focus on flows of resources and information between stocks which can be controlled.

Two main types of interactions between these models and stakeholders have so far emerged: group model building (Vennix 1996); and management flight simulators or microworlds (Maier and Grössler 2000).

Group Model Building experiments focus on the interaction with stakeholders in the design stage of a modelling process. It associates techniques of system dynamics modelling with brainstorming tools and other techniques of group work, mainly based on workshops and meetings. This trend consists of integrating future users of the model in the design stage. The participants are supposed to be the clients of the modelling process. Rouwette and colleagues analysed 107 cases of such experiments and proposed a number of guidelines to facilitate consistent reporting on participatory modelling exercises. These guidelines focus on three categories: context, mechanisms and results (Rouwette et al. 2002). The second category focuses predominately on preparation activities and description of meetings, along with factual elements and the modelling process.

This category of participatory modelling deals with the expectations identified in the first section in the following manner:

– The participation arena is constituted of a rather small or medium size well identified group. The structure of the interaction is rather global: debates tackle the whole model, and participants are supposed to be concerned by the model entity as a whole. The connections may convey information on the tacit knowledge of stakeholders, as well as on their purposes. This is still very diverse among the experiments. The group of stakeholders is mobilised within specific events, workshops, which might be repeated. The aim is to feed the model but also to increase the probability of use of the models produced.
– The process is predominately controlled by the modellers; and
– The format of information is generally not well formalised, even though techniques, such as hexagons brainstorming or causal diagrams (Akkermans 1995), appear to organise the knowledge brought by stakeholders. This low formalisation allows the issues related to stakeholder diversity to be tackled and alleviated in the problem framing phase, but it leaves a large place to the modellers' interpretation.

Management flight simulators or microworlds constitute a complementary technique, which focuses more on the stages of use and simulation outcomes analysis, even though this technique may also be used in a design stage to elicit tacit knowledge. A key characteristic of this type of technique is to encourage learning by doing. Participants, who might be the clients or other concerned people without any formal relation to the modelling team, have to play through a simulation of the model. Martin and colleagues have used this technique to validate a system dynamics model on the hen industry (Martin et al. 2007). Participants were asked to play with some parameters of the model.

When used to elicit knowledge, microworlds attempt to provide events that are similar to those that participants already face or are likely to face in their activities related to the issue at stake in the model. Le Bars and colleagues have thus developed a game setting to lead farmers to understand the dynamics of their territory with regard to water use and changes in EU Common Agricultural Policy (Le Bars et al. 2004). In flight simulator experiments, interaction between stakeholders and the simulation model is structured around future users of the model or people whose stakes are represented in the model, with a slightly deeper connection than with previous group modelling building approaches. Participants are asked to deal with parameters of the model and are framed in the categories used in the model. There is no a priori differentiation among participants. The connections convey information about the object from the model to participants. It also conveys the participants' reactions to this object, and some behavioural patterns observed that can provide new information for the modellers. This connection is activated by the participants working through specific events and focus on the use of the tool. Control is still on the side of modellers, who frame the interactions. The format of information is largely formalised from model to stakeholders. It is not formalised from stakeholders to model.

### *10.3.2 Knowledge Engineering: Between Artificial Intelligence and Social Psychology*

Knowledge engineering focuses on a specific time of the interaction between stakeholders and a simulation model in the design stage: the process of translating tacit knowledge into conceptual or sometimes computational models. Many knowledge elicitation techniques are useful in transforming written or oral text into pieces of simulation models. The purpose of these techniques is to separate the contributions made directly to the model from the design of the model itself. Knowledge engineering aims to provide interfaces for this gap.

To deal with this interface, techniques have been developed, grounded in artificial intelligence, (social) psychology and cognitive science. Behavioural patterns in social simulation models are often borrowed in simplified versions from these fields (Moss et al. 2000; Pahl-Wostl ND Hare 2004). This cross-pollination of disciplines can be potentially fruitful for model design. As an example, Abel and colleagues have built upon the concept of a mental model. They assume that individuals have representations of their world which may be formalised in causal rules. Working in the Australian bush, they have designed specific individual interview protocols and analysis frameworks to elicit these mental models (Abel et al. 1998). In this case, interaction with the model occurs through the interviewer who in this case was also the modeller. There was no collective interaction. Researchers dealing with the interviews and the corresponding model design clearly guide the process. The format of information is speech (in the form of a transcribed text), which is transformed into a modelling language in this elicitation process.

Building upon Abel's work, Becu has further minimised the involvement of the modeller, still using individual interviews. This has led him to collaborate with an anthropologist and to use ethnographic data as a benchmark. Individual interviews, with the interviewee in the environment suitable to the purpose of the interview, led him to identify objects and relations among these objects. These constitute the initial basis for an exercise, labelled as playable stories: stakeholders, in his case farmers from Northern Thailand, are asked to choose the key elements to describe their world from their own viewpoints (with the possibility of adding new elements), then to draw relations among them and to tell a story with this support

(Becu 2006; Becu et al. 2006). In this case, interaction between stakeholders and the simulation model is still on an individual basis. The format of conveyed information is finally less formal, but the work of translation is less important. However, control of the process still remains largely in the hand of the modeller, but to a lesser degree than in previous examples. This technique was further associated with semi-automatic ontology building procedures by Dray and colleagues in order to generate collective representations of water management in the atoll of Tarawa (Dray et al. 2006a).

With inspiration coming similarly from the domain of ethnography, Bharwani and colleagues have developed the KNeTS method to elicit knowledge. Apart from a first stage with a focus group, this method is also based on individual interviews. As in Becu's work, interaction occurs in two phases: elicitation through questionnaires and involvement in the model design at the validation stage, which is also considered as a learning phase for stakeholders. These authors used an interactive decision tree to check with stakeholders whether the output of simulation would fit their points of view (Bharwani 2006). Control of this process is on the modeller's side. The stakeholders' interaction is marginally deeper in the model than in previous examples, since there is a direct interaction with the model as in management flight simulator. On the other hand, the ontology which is manipulated seems to be poorer, since the categories of choices open in the interaction are rather reduced. The format of information is open in the first phase and very structured in the decision tree in the second phase. The structuration process used in the modelling process occurs outside of the field of interaction with the stakeholders.

On its side, Group Decision Support System design domain is based on a collective interaction with stakeholders as early as the design stage. These systems tend to be used to address higher level stakeholders. In the method he developed, ACKA, Hamel organised a simulation exercise with the stakeholders of a poultry company. In this exercise, the participants were requested to play their own roles in the company. He constrained the exchanges taken place during the exercise through the use of an electronic communication medium so that he could analyse them and keep track of them later. All of the participants' communication was transformed into graphs and dynamic diagrams (Hamel and Pinson 2005). In this case, the format of information was quite structured.

### *10.3.3 From Software Engineering*

Close to the artificial intelligence trend, working like Hamel and Pinson on the design of Agent Based Models, there is an emerging trend in computing science based on agent based participatory simulations (Guyot and Honiden 2006) or participatory agent based design (Ramanath and Gilbert 2004). This trend focuses on the development of computer tools, multi-agent systems, which originate from software engineering. Guyot proposes the implementation of hybrid agents, with agents in the software controlled by real agents, as avatars (Guyot 2006). These avatars help the players' understanding the system (Guyot and Honiden 2006). They can be thought as learning agents: they learn from choices of their associated player and are progressively designed (Rouchier 2003). The approaches working on hybrid agents implement a deep connection between participants and the social simulation model. Information conveyed in the interaction is relative to the model assumptions, as well as to the model content.

Ramanath and Gilbert have reviewed a number of software engineering techniques which may be coupled to participatory approaches (Ramanath and Gilbert 2004). This union between software design and participatory approaches is based on joint production not only between developers but also with end-users. Not only interaction with stakeholders contributes to better software ergonomics – the Computer Supported Cooperative Work (CSCW) workshops series being an example – but their participation tends to improve their acceptation and further appropriation of the mode.

The implementation of interactive techniques may take place at all stages of a software development process. In early stages, joint application design (Wood and Silver 1995) allows issues raised to be dealt with during the software development phase, attributing a champion to each issue. It is also concerned with technical issues. This protocol might involve other developers, as well as potential users. It may also increase the computing literacy of the participants involved in the process. This process is based on the implementation of rather well framed workshops.

Joint application design is supported by using prototypes. It is here we find a link with a second technique: prototyping. This technique can be used all the way through a software development cycle. It is based around providing rough versions or parts of the targeted product. For example, it allows the pre-product to be criticised, respecified, or the interface improved. Quite close to prototyping, in the final stages of the process, user panels can be used to involve end-users in assessment of the product. These panels are based on a demonstration or a test of the targeted product.

In these cases, control of the process is dependent on the hiring of a skilful facilitator. Otherwise, control of the process may become rather implicit. The content of the interaction is rather technical, which makes it potentially unbalanced according to participants' literacy in computer science. An assessment of 37 joint application design experiments has shown that the participation of users during the process is actually rather poor, notably due to the technical nature of debates, which is hardly compatible with the time allocated to a joint application design process by users, compared to the time allocated by developers (Davidson 1999). Interaction is rather superficial and needs translation. However, identification of a champion of specific tasks gives a little bit more control to participants, as does involvement in the content of pieces of the tool being developed.

Besides these approaches originating from software engineering, people working in thematic fields such as the environmental sciences propose co-design workshops that focus on the development of simulation models. Such workshops are a type of focus group, organised around the identification of actors, resources, dynamics and interactions, suitable for a set of stakeholders to represent from a socio-ecological system on which they express their own point of view (Étienne 2006). This approach, which occurs at the design stage of the modelling process is supposed to lead participants to design the simulation model by themselves, by formalising the conceptual model through a series of diagrams and a set of logical sentences. The final interaction diagram and the attached logical sentences are then translated by the modeller in computer code. It is in this type of process that a deep interaction can occur between participants and the model. This interaction conveys information on the model content, which is attached to the representations and knowledge of each participant.

### *10.3.4 From Statistical Modelling*

Bayesian Belief Networks have been developed to include in the computation of probabilities, their dependence on the occurrence of any event. They can be useful to represent complex systems and increasingly used in participatory settings because their graphical nature facilitates discussion (Henriksen et al. 2004). A group of participants can be asked individually or collectively to generate relations between events and possibly probabilities as well. Henriksen and his colleagues propose a method in seven stages which alternates between individual and collective assessment and revision of an existent Bayesian Belief Network diagram.

This approach is reported to still present some difficulties in encouraging strong participant involvement due to the mathematical functions behind the network structure. However, other researchers and practitioners have improved their communication and facilitation of the technique with their own Bayesian Belief Network processes and are receiving positive stakeholder engagement in the modelling processes (Ticehurst et al. 2005). In the example of Henriksen and colleagues, the process is controlled by the modeller and includes only a rather superficial coupling between participants and the model. The translation of participant-provided information into probabilities is mediated by the modeller and is rather opaque, as in many participatory modelling approaches.

### *10.3.5 From the Social Sciences*

The association of participatory approaches and social simulation modelling also originates from disciplines not focussing on the production of tools but on understanding social systems. Social psychology, economics, management and policy sciences have all developed their own interactive protocols to involve stakeholders in the design and/or use of their models. Sociology is still at the beginning of this process (Nancarrow 2005). These protocols propose a variety of structures of experimental settings, from laboratory to in vivo experiments through interactive platforms (Callon and Muniesa 2006). These three categories vary according to their openness to the influence given to participants. The in vivo category is beyond the scope of this paper since it does not involve modelling: the society in which the experiment is embedded provides its own model (Callon and Muniesa 2006).

Laboratory settings are very controlled experiments, involving human subjects. This is the case for most economic experiments. Participants are encouraged to behave with a given rationality through instructions and payments at the end of the session. In canonical experiments, analysis of the experiments is performed by the scientist. The focus of the analysis is to understand the individual and collective behavioural patterns generated by these settings. The purpose of these experiments is either: the testing of theories and models; new knowledge on human behavioural patterns in given situations; or the testing of new institutional configurations (Friedman and Sunder 1994). These experiments are particularly efficient for situations with strong communication issues or with important inter-individual interactivity (Ostrom et al. 1994). The issue of simulating a real situation is not considered, but rather the testing of a theoretical model. This field is currently very active and evolve with the emergence of field experiments involving stakeholders concerned by the issues idealised in the model tested, asking them to play in their environment (Cardenas et al. 2000). With this configuration, interactions are rather deep since participants act as parts of the model. The participants convey action choices. However, the experimentalist strongly controls the process.

A platform is an intermediary setting more open to compromise and hybridisation than the laboratory. Heterogeneity of participants is also more welcome, since the setting is designed to enhance sharing interests. Through experimentation, a platform is supposed to bridge through experimentation the gap between the world of the model and that of the stakeholders (Callon and Muniesa 2006). Policy exercises and role playing games, as developed in the companion modelling approach, are kinds of these platforms (Richard and Barreteau 2006). Policy exercises embed stakeholders in potential situations they might have to face in the future (Toth 1988). They stem from war games that have been developed since the time of Ancient China and are now used in public policy assessment (Duke and Geurts 2004) or environmental foresighting (Mermet 1993). They are actually quite similar to the business games and the system dynamics trend explained previously in subsection 10.2.1. However, the underlying social simulation model is rather implicit; though it exists to create the potential situation and to help identify the participants relevant to the exercise. Association with a computer tool tends to be with a simulation model of the environment, that does not necessary involve a social component. The interaction between participants and the social model is rather deep since they are pieces of the model and connect with the model of their environment. Control of the process is rather diffuse. There might be a genuine empowerment of participants since they have the possibility of bringing their own parts of the social model to the process, and can adapt it in ways different to what the designers expected. Alike with laboratory settings, platforms provide information to the modeller about behavioural patterns of the participants. Reaction to taboos or innovative behaviours in situations new to the participants, tacit routines, and collective behavioural patterns can be elicited using these platforms, while it is difficult with classical interviewing techniques.

Between experimental laboratory settings and policy exercises, the companion modelling approach proposes an association of role playing games and agent based simulations (Bousquet et al. 2002). Even though authors in this approach claim not to limit themselves to these two categories of tools, they predominately rest in the trend of participatory agent based simulations, and are thus close to the software design and artificial intelligence trends presented above. This approach makes a full use of similarities in architecture between role playing games and agent based simulations (Barreteau 2003). Both implement autonomous agents that interact within a shared dynamic environment. Joint use of both agent based simulation and role playing games builds upon these similarities to express the same conceptual model. Authors in this approach use this to reinforce a principle of making all the assumptions underlying a model used or design interactively with stakeholders explicit and understood. At the design stage, this approach aims to incorporate stakeholders' viewpoints in the model. At the model use stage, it aims to improve the appropriation of the tool produced, as well as to increase its legitimacy for further operational use. However, this appropriation is still under discussion and might be rather heterogeneous (Barreteau et al. 2005).

## 10.4 Participation in the Modelling Process: Diversity of Phases and Intensity

While many authors claim to use participatory approaches for the simulation of social complexity, there remains a large diversity of actual involvement of stakeholders and of activities hidden behind this involvement. Associations of participatory methods with social simulation models are rather heterogeneous. It is thus important to qualify the actual involvement of stakeholders in these processes. This level of participation can range from mere information received by concerned parties related to the output of a process to the full involvement of a wide range of stakeholders at all stages of a process. There are also many intermediary situations imaginable. Participation should not be thought of as just talking, and diversity should be made explicit so that criticisms towards participation as a global category (Irvin and Stansbury 2004) can focus on specific implementations. This section explores the potential consequences of this diversity in three dimensions: stage in the modelling process, degree of involvement and heterogeneity of stakeholders involved.

### *10.4.1 Stages in the Modelling Process*

The modelling process can be subdivided into the following stages, with the possibility of iterating along them:

– Preliminary synthesis/diagnosis (through previously available data). This includes making explicit the goal of the modelling process
– Data collection (specific to the modelling purpose)

– Conceptual model design
– Implementation
– Calibration and verification
– Simulation process (might be running a computer simulation model, playing a game session, etc.)
– Validation
– Discussion of results

Involvement of stakeholders in each of the different stages of the modelling process does not generate the same level of empowerment or learning, even if we assume that this involvement is sincere. Preliminary synthesis, conceptual model design, validation and, to some extent, discussion of results are framing stages; stakeholder involvement at these levels gives power to stakeholders to orientate the process. In the preliminary synthesis/diagnosis, stakeholders have the opportunity to play a part in setting the agenda. This is the stage of problem structuring which is identified as a key one in all participatory processes (Daniell et al. 2006). Even if the agenda developed with stakeholder involvement might further evolve, its initialisation generates a strong irreversibility in the process: data collection, participants selection and (partially) modelling choices (architecture, platform) are related to this agenda and are costly, either directly or through the necessity of re-programming. The modelling process is a sequential decision process, and as shown in theory of sequential decisions: initial decisions are often at the source of more consequences than envisaged (Henry 1974; Richard and Trometter 2001). Conceptual model design constitutes a landmark in the process. It is the crystallisation of viewpoints that serves as a reference in further stages. Validation is the compulsory stage where stakeholders will have the opportunity to check the effectiveness of the computer model in representing correctly their behaviours and ways of acting. Discussion of results may also constitute a framing phase, according to the purpose of the discussion. If dimensions of discussion are to be defined and model is open to be modified, there is some place for participants to (re-)orientate the modelling process. Otherwise, if the discussion of results aims to choose from a few scenarios for example, the choice is very narrow and might be completely manipulated. In this regard, it has been shown that for any vote among composite baskets, it is possible to maintain that one item always selected according to the way the baskets are constituted (Marengo and Pasquali 2003). A scenario in this case is a kind of composite basket.

In other stages of a modelling process, the influence of stakeholder involvement on the overall process is less important. When data collection, or calibration and verification involve participants, stakeholders tend to take the role of informants. Among the various levels proposed in the classical ladder of participation explained in the following subsection, these stages deal predominately with consultation. Their involvement is framed by the format of information which is expected, and on the parts of the model which are to be calibrated or validated. If the process is open to modification in these frames, the level of participation might be higher, but still with a limited scope.

Implementation stage is another mean to empower participants. It is often implicitly framing. But empowerment through involving stakeholders in this technical activity is rather to raise their literacy in this part and raise the probability of their appropriation of the model. Simulation stage is basically providing information to stakeholders on what is being done. This is a technical stage (running the simulation) which keeps a part of strategic choices (design of scenarios and indicators to track the simulation progress). Involvement of stakeholders in the technical part, such as through role playing games, increases their knowledge of the model from inside, provided stakeholders have the literacy for that. Involvement in strategic part is connected to the initial stage which has set the agenda. The further this initialisation has gone in formalising the questions, the less empowering is this involvement.

### *10.4.2 Level of Involvement*

Level of involvement is a more classical dimension. It is inspired by the classical hierarchy of participation levels proposed first by Arnstein (1969). Several reviews and adaptations have been made since then, with the same focus on power issues (Mostert 2006; Asselt et al. 2001). These works focus on what participation means in decision making terms (the bases of many political or democratic theories), with Democracy Cube (Fung 2006) or the work of Pateman (1990) and Rocha (1997). In most of these examples, the emphasis is placed on who (citizens, managers or policy makers) has the balance of power for final decision-making (i.e. the choice phase of a decision process (Simon 1977)) but other issues of process are not specifically mentioned. Such participation classifications, although useful in a very general sense for the question of participation in modelling processes, do not explicitly treat the issue of the place of a modeller or researchers with expert knowledge (Daniell et al. 2006).

On these bases, we consider here the five following levels in which there are at least some interactions between a group of citizens and a group of decision makers:

- Information supply: citizens are provided access to information. This is not genuine participation since it is a one way interaction;
- Consultation: solicitation of citizens' views;
- Co-thinking: real discussions between both groups;
- Co-design: citizens have an active contribution in policy design; and
- Co-decision making: decisions are taken jointly by members of both groups.

Since a modelling process is a kind of decision process, this hierarchy might apply to modelling process as well. This is a little bit more complicated because two processes are behind the modelling process and the network of interactions cannot be represented with a group of citizens and a group of decision or policy makers only.

A modelling process with the purpose of simulation has two dimensions along which these scales might be assessed: model content and building on one hand; and control over model use on the other. Though these two dimensions are related, it is useful to consider them separately as they provide power and knowledge: either within the process; or in the system in which the process takes place. Each of these dimensions is more closely related to specific stages in the modelling process presented in the previous subsection. However, some stages, such as model design or implementation, contribute to both dimensions.

Therefore we consider the following categories:

– Information on a model's content and no control over model use;
– Consultation and no control over model use;
– Dialogue with modellers and no control over model use;
– Dialogue with modellers and control over model use;
– Co-building of a model and no control over model use; and
– Co-building of a model and control over model use;

Each category is described in the following sub-section by a flow of interactions within an interaction network based on four poles: **A**, **R**, **M**, and **P**. **A** stands for all people who are involved in and/or concerned by the social complexity at stake in the modelling process. This includes policy makers and citizens. **R** stands for researchers involved in the modelling process. **M** stands for the model. **P** stands for policy makers. **P** is a subset of **A**, which gathers the actors who might use the model and its output for the design of new regulations or policies concerning the system as a whole. We chose to gather citizens and policy makers in **A**, as in the modelling process they are rather equivalent in their interactions with the researchers about the model. Their distinction is useful for the second dimension: model use and dissemination. We assume that the default situation is an access of **P** members to the output of the modelling process.

#### 10.4.2.1 Information and No Control

$$\begin{tikzcd} A & R \arrow[l] \arrow[d] \\ P & M \arrow[l] \end{tikzcd}$$

Participants are informed about the model's content and the simulation by researchers, who are the only designers. No control over the model's use or dissemination is deputed to participants as such. Whatever the use of the model may be afterwards, citizens become only better aware of the basis on which this model has been built. However, the model exists and can be used by members of **P**. This is the classical situation with simulation demonstration and explanation of a model's assumptions. This explanation might be achieved by more active means, such as a role playing game. A switch to the following category occurs when this explanation leads to a debate that makes the model open to modifications. Otherwise, it remains mere information.

#### 10.4.2.2 Consultation and No Control

$$\begin{tikzcd} A \arrow[r] & R \arrow[d] \\ P & M \arrow[l] \end{tikzcd}$$

Participants are consulted about the model's content and its simulation that is by the researchers, who are the only designers. They provide information and solicit comments on the model. Mere data collection through a survey does not fall in this category because it assumes active involvement from participants in providing information to the modellers. Some knowledge elicitation techniques, such as BBN design, tend to fall mostly in this category. Translation of the inputs originating from participants into pieces of a model is performed only by researchers. This translation is not necessary transparent. No control over use or dissemination of the model is deputed to participants as such. Compared to previous category, participants have the ability to frame marginally more of what is performed by the model through their inputs to the model's content However, the extent of this ability depends on the participants' skills to identify potential uses of a model. As in any participatory process, when there is an unbalanced power relation between parties, the process is also a way for policy makers to gain information from stakeholders; information that could be used for strategic purposes. This bias can be alleviated if the involvement of **A** includes all members of **A**, including the subset **P**. The constructed model in this case may be used by the members of **P**.

#### 10.4.2.3 Dialogue with Modellers and No Control

$$\begin{tikzcd} A \arrow[r] & R \arrow[d] \\ P & M \arrow[l] \arrow[ul] \end{tikzcd}$$

In this category, iterative and genuinely interactive processes between stakeholders and modellers start to appear. There is still a translation of inputs from participants into the model through the researchers, but there is feedback about these developments to the stakeholders. This leads to discussion about the model. Convergence of the discussion remains on the researchers' side. Group Model Building experiments predominately fall into this category. In this case, stakeholders may increase their influence on the framing of the model with better prior assessment of the scope of simulations to be examined. Biases related to strategic information being revealed in the dialogue process are still present if there is unbalanced involvement of Members of **A**, and notably if members of **P** are less active, but still present. However, this category still represents indirect control and no specification of model use is left open to the stakeholders. At the end of the process, the created model can be used by members of **P** without any control or any roadmap set by other members of **A**.

#### 10.4.2.4 Dialogue with Modellers and Control

$$\begin{tikzcd} A \arrow[r] \arrow[d] & R \arrow[d] \\ P & M \arrow[ul] \end{tikzcd}$$

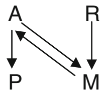
This category is the same as the previous one with translation of stakeholders' inputs and feedback from the researchers about them. However, the output of the discussion, the model, is appropriated by stakeholders. They have control over its use and dissemination of models which may have been produced through the modelling process: who might use them; with which protocol; and what is the value of their outputs. They can decide whether the model and simulations are legitimate to be used for the design of policies that may concern them. However, this appropriation raises issues of dialogue between researchers and stakeholders about the suitability of model for various uses. Comparison of several participatory agent based simulations has shown that there is a need for dialogue about not only a model's content but also on its domain of validity (Barreteau et al. 2005).

#### 10.4.2.5 Co-building of a Model and No Control

$$\begin{tikzcd} A \arrow[dr] & R \arrow[d] \\ P & M \arrow[l] \end{tikzcd}$$

A further stage of empowerment of stakeholders through participation in a modelling process is their co-building of the model. The design and/or implementation of such a model are joint activities between the researchers and stakeholders. Co-design workshops or joint application development fall into this category, provided that there is genuinely no translation of stakeholders' inputs by the researchers. Techniques originating from Artificial Intelligence and knowledge engineering, as presented above, aim to reach this level, either through the implementation of virtual agents extending stakeholders, or through constraining the interactions between actors through a computer network. This involvement increases the fidelity of the model to match stakeholders' viewpoints and behavioural patterns. However, at the end of the process, the created model can still be used by members of **P** without any control or any roadmap set by other members of **A**.

#### 10.4.2.6 Co-building of a Model and Control




This category is the same as the previous one, but actors now have control over use and dissemination of models which may be produced through the process. This leads to possible stakeholder appropriation of the models, raising the same issues as in Sect. 3.2.4.

### *10.4.3 Heterogeneity of Actors*

Eversole points out the need for participatory processes to take into account the complexity of the society involved including: power relations; institutions; and the diversity of viewpoints (Eversole 2003). This is all the more true when applied to the participatory process of social simulation modelling. Most settings presented in Sect. 10.2 have a limited capacity to involve a large numbers of people in interactions with a given version of a model. When interactions convey viewpoints or behavioural patterns, heterogeneity may not appear if no attention is paid to it. Due to limits in terms of number of participants, participatory approaches that deal with social simulation modelling involve usually representatives or spokespeople. The issue of their statistical representativeness is left aside here, as the aim is to comprehend the diversity of possible viewpoints and behavioural patterns. There is still an issue of their representativeness through their legitimacy to speak for the group they represent, as well as their competency to do so. The feedback of these spokespersons to their group should also be questioned. When issues of empowerment are brought to the fore, the potential for framing or controlling the process is dedicated to the participants. This might induce echoes in power relations within the group, notably due to training that may be induced.

Van Daalen and Bots have proposed a categorisation of participatory modelling according to this dimension with three scales: individual involvement; a group considered as homogeneous, and a heterogeneous group (Daalen and Bots 2006). Table 10.1 provides examples of each level according to the two processes involved that were explained in previous subsection.

These three categories are represented in the diagrams below, as expansions of the relation between A and (M ∪ R) in the previous subsection. The third category corresponds to the deep connection mentioned in the first section (Figs. 10.1, 10.2, and 10.3).

Some other ways are currently explored with hybrid agents to technically overcome the difficulty of dealing with representatives: by involving them all in large systems. The internet or mobile phone networks provide the technical substrate for

**Table 10.1** Categories of participation according to level of heterogeneity embraced (From Daalen and Bots 2006)

| Level | Model construction | Model use | |
|---|---|---|---|
| | | Computer model | Gaming simulation |
| 1. Individual stakeholders | Knowledge elicitation involving one or more individuals separately; depending on the modelling method this may consist of interviews about (perceptions on) a system or questionnaires related to the aspects being modelled (e.g. Molin 2005) | Model can be executed and individual stakeholders are informed of the result (e.g. Dudley 2003) | Individual can 'play' an actor in a flight simulator setting (e.g. Maier and Grössler 2000; Sterman 1992) |
| 2. Homogenous group | Same as 1, but group model building includes interaction between stakeholders (e.g. Castella et al. 2005) | Use of a model in a homogenous group means that the model can be run in a workshop setting and model results are discussed (e.g. Daalen et al. 1998) | Multi-player gaming simulation can be conducted, the game is followed by a debriefing (e.g. Mayer et al. 2005) |
| 3. Heterogeneous group | Same as 2, but group model building interaction between stakeholders with different perceptions/beliefs (e.g. Van den Belt 2004) | Same as 2, but results discussed with stakeholders with different perceptions/beliefs | Same as 2, but full stakeholder group involved (e.g. Étienne et al. 2003) |

such interaction. A large number of participants have a virtual component in a large system, interacting with other components, possibly with the purpose of building a model (Klopfer et al. 2004). However, in this case it is rather an individual interaction of these participants with the system, than genuine interactions amongst the participants.

### *10.4.4 Which Configurations Can Meet the Expectations of the First Section?*

In this subsection we revisit the expectations towards the joint use of participatory approaches and social simulation presented in the first section, through the

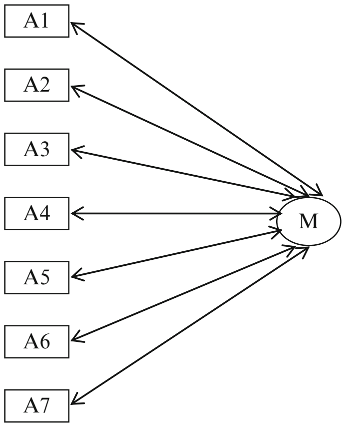



**Fig. 10.1** Individual involvement

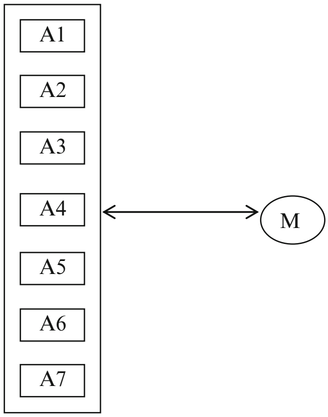



**Fig. 10.2** Homogenous group involvement

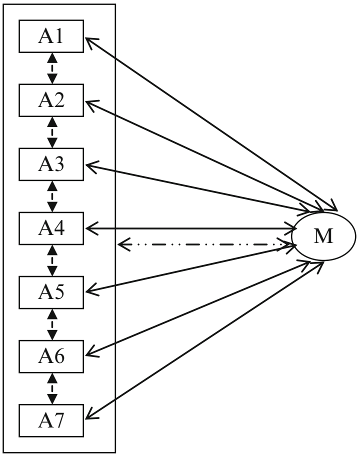



**Fig. 10.3** Heterogenous group involvement

**Table 10.2** Matching expectations on joint use of participatory approaches and social simulation modelling with categories of participation

| Expectation | Key stage(s) for participation | Minimum level of empowerment | Level of heterogeneity |
|---|---|---|---|
| Increase model's quality with social diversity and capacity to evolve | Simulation | Information and no control | Heterogeneous group |
| Increase model's quality through distribution of control | Simulation | Information and no control | Heterogeneous group |
| Improve suitability of simulation model's use for increasing knowledge | Design | Dialogue and no control | Individual |
| Improve suitability of simulation model's use for policy making | Design and discussion of results | Dialogue and control | Homogeneous group |
| Simulation as a means to support participation to deal with dynamics and uncertainties | Discussion of results | Consultation and no control (depend on participatory process to be supported) | Homogeneous group |
| Simulation as a means to support participation through social learning | Preliminary diagnosis, design and discussion of results | Co-building and control (to be preferred) | Heterogeneous group |

categorisations above. This is a tentative mapping of participatory approach categorisation with model expectations. Table 10.2 below synthesises this mapping.

The two expectations dealing with increasing a model's quality often actually use participants as (sometimes cheap) resources in the simulation modelling process. The most important stage is simulation, because participants are supposed to bring missing information to the simulation, as well as the missing complexity. The minimum level of empowerment is rather low. These processes are hardly participatory in that sense, because participants are not supposed to benefit from the process, except a potential payment. A higher level of empowerment might increase the quality of participants' involvement in the process through a deeper concern in the outcome of the simulation. Finally, the heterogeneous group level is obviously to be respected because it can instil a deep connection between stakeholders and the model still and concurrently profit from their interactions with each other.

To make simulation models match their intended use, the key stage is the design process. Stakeholders are supposed to aid the building of an appropriate model. The main difference between targets of simulation model's use is in the necessity to give control over the process to stakeholders in case of policy making. New knowledge is of individual benefit to all participants and the emergence of fruitful interactions can also become an individual benefit. There are few direct consequences of this new knowledge. Therefore, control over the process in this case is useless and involvement might be individual, as with knowledge elicitation techniques. However, higher level of stakeholder heterogeneity might raise the knowledge acquired in the process.

When simulation is used to support participation, discussion of results is a key stage. Previous stages aid in the problem framing and literacy increase of participants that allow them to reach more solid interpretations. The empowerment level is rather dependent on the participatory process that is being supported. However, consultation in the modelling process should be a minimum requirement so that uncertainties and dynamics tackled by the simulations are relevant to the stakeholders. When focusing on social learning, co-building and control should be preferred because this category increases the potential for exploration and creativity. However, some social learning might take place in lesser levels, provided that group heterogeneity is well encouraged in the process.

## 10.5 Combining Approaches and Techniques at Work

We present in this section two case studies implementing various methods for joining social simulation modelling and participatory approaches. The first deals with fire hazard prevention in southern France, and the second one with groundwater management in the Atoll of Kiribati.

### *10.5.1 The Fire Hazard Case Study*

In December 2005, the Forest Service of the Gard Department of Agriculture (DDAF), decided to start a fire prevention campaign focused on fire hazard at the interface between urban and forest areas. Interested in the participatory approaches developed by INRA researchers on fire prevention and forest management planning (Etienne 2003), they ask for an adaptation of the SylvoPast model to the periurban context in order to make local politicians aware of the increasing fire hazard problem. The District of Nîmes City (NM) who was already interested in the use of role playing games for empowering stakeholders and decision makers, asked the Ecodevelopment Unit of the INRA of Avignon to develop a companion modelling approach based on social simulations and a participatory involvement of all the mayors of the district.

The modelling process was subdivided into seven stages:

1. Collection and connection on a GIS of relevant cartographic data on forests, land-use and urbanization, and individual interviews with local extensionists on farmers, foresters and property developers practices.
2. Co-construction with DDAF and NM of a virtual but implicit map representing three villages typical from the northern area of Nîmes city and validation of the map (shape, land-use attributes and scale) by a group of experts (EX) covering the

main activities of the territory (agriculture and livestock extensionists, forest managers, hunting manager, land tenure regulator, fire brigade captain and town planner).

3. Co-construction, with NM, DDAF and experts of a conceptual model accounting for the current functioning of the territory and the probable dynamic trends to occur during the next 15 years. This participatory process followed the ARDI methodology mentioned in Sect. 10.2.3 (Étienne 2006).
4. Implementation of the NimetFeu model on Cormas multi-agent platform by INRA researchers and validation of the model by simulating with the co-construction group, the current situation and its consequences on fire hazard and landscape dynamics for the 15 following years.
5. Co-construction and test of a role-playing game (RPG) using the NimetFeu model as a way to simulate automatically natural processes and some social decisions (vineyard abandonment, horse herding, fire fighting). The other social decisions were programmed to be taken directly by the players and used as an input to the model.
6. Use of the RPG during several sessions gathering 6 players (3 mayors, 1 developer, 1 NM representative, 1 DDAF technician) until the 14 villages involved in the project did participate to a session.
7. Adaptation of NimetPasLeFeu to other ecological conditions, and decision of the Gard Department to become autonomous in running RPG sessions. A facilitator and a data manager were trained and tested during sessions organized in the framework of an INTEREG project with mayors and fire prevention experts from France, Spain, Italy and Portugal

The approach is based on a mutual comprehension of the elements of the territory that make sense with the question asked. This sharing of representations is done by means of a series of collective workshops during which Actors, Resources, Dynamics and Interactions (ARDI) which make the stakes of the territory are identified and elicited. To facilitate this sharing, the answers to the questions are formalized into easily comprehensible diagrams, with a minimum of coding making it possible to classify the provided information. The role of the facilitator only consists in calling upon each participant, writing down the proposals in a standard way, and asking for reformulating when the proposal is too generic, enounced with a polysemous word or can lend to confusion.

In both models, the environment is divided into three neighbouring villages covering the gradient of urbanization and agricultural land/woodland ratio currently observed around Nîmes city. It is visualized by means of a cellular automaton through a spatial grid representing 18 land-use types that can change according to natural transitions or human activities.

Four categories of social entities are identified: property developers, mayors, farmers and fire prevention managers. The developers propose new urban developments according to social demand and land prices. They have to respect the government regulations (flood hazard, protected areas, urban zoning). Mayors select an urbanization strategy (to densify, to develop on fallow land, olive groves or forests), update their urban zoning according to urban land availability and social demand and make agreements with the developers. When updating the urban zoning, they can create new roads. Farmers crop their fields using or not current practices that impact fire hazard (vineyards weeding, stubble ploughing) or adapt to the economic crisis of certain commodities by uprooting and setting aside lowland vineyards or olive groves near to urban zones. The fire prevention manager establishes a fuel-break in a strategic place, selected according to fire hazard ranges in the forest and the possible connections with croplands, as well as available funds and forest cleaning costs.

Four biophysical models issued from previous researches and adjusted to the local conditions are integrated to the MAS to account for fallow development, shrub encroachment, pine overspreading and fire propagation. The model is run at a 1-year time step, the state represented on the map corresponding to the land cover at the end of June (beginning of the wildfire period). Each participant was invited to propose a set of key indicators that permit them to monitor key changes on ecological or socio-economic aspects. A common agreement was made on what to measure, on which entity, with which unit, and on the way to represent the corresponding qualitative or quantitative value (visualizing probes on graphs or viewpoints on maps). They were encouraged to elaborate simple legends, in order to be able to share their point of view with the other participants while running the model.

The first MAS was exclusively used to support the collective thinking on which procedures and agents will be affected to players, and which ones would be automatically simulated by the computer. In the RPG model, the playing board was strongly simplified with only four types of land cover. Running the game gives participants the opportunity to play individually or collectively by turns, according to a precisely defined sequence. While the mayors players draw the limits of the urban zone and rank the price of constructible land according to its current land-use, the developer player sorts randomly a development demand and elaborates a strategy (village, density, livelihood). Then begin a series of negotiations between the developer and the three mayors in order to decide where to build, at which density and with which type of fire prevention equipment. All the players' decisions are input into the computer and landscape dynamics is simulated by running the model. Players get different types of output from the simulation run: budget updating, new land-use mapping, popularity scoring. Each round corresponds to a 3-year lapse and is repeated three to four times according to players' availability.

A specific effort is made in the RPG design to account for physical remoteness and territory identity among participant: the playing room is set up into three neighbouring but distinct boxes for the three mayors (each box represents one village), one isolated small table for the developer, and another game place with two tables, one small for the DDAF and a huge one for NM. Lastly, in a corner, the computer stuff is placed with an interactive board than can both be used as a screen to project different viewpoints on the map or as an interactive town plan to identify the parcels' number.

At the end of the game, all the participants are gathered in the computer room and discuss collectively, with the support of fast replays of the game played.

Table 10.3 Classification of type of participation in various stages of the NimetPasleFeu experiment

| | Involvement | Heterogeneity | nb |
|---|---|---|---|
| Preliminary diagnosis | Consultation | Individuals | 10 |
| Data collecting | Consultation | Individuals | 3 |
| Conceptual model designing | Co-design | Heterogeneous group | 14 |
| Implementing | Information | Individuals | 2 |
| Calibrating and validating | Co-thinking | Heterogeneous group | 14 |
| Role-playing game designing | Co-design | Heterogeneous group | 14 |
| RPG playing and debriefing | Co-decision making | Heterogeneous group | 30 |
| Getting self sufficient | Information | Individuals | 3 |

Different topics are tackled related to ecological processes (effect of fire, main dynamics observed), attitudes (main concerns, changes in practices), and social behaviours (negotiations, alliances, strategies).

Along these various stages, this experiment features a diversity of involvement as well as of structure of interactions. This is synthesised in the Table 10.3 above.

### 10.5.2 The AtollGame Experiment

This study is carried out in the Republic of Kiribati, on the low-lying atoll of Tarawa. The water resources are predominantly located in freshwater lenses on the largest islands of the atoll. South Tarawa is the capital and main population centre of the Republic. The water supply for the urban area of South Tarawa is pumped from horizontal infiltration galleries in groundwater protection zones. These currently supply about 60 % of the needs of South Tarawa's communities. The government's declaration of water reserves over privately owned land has lead to conflicts, illegal settlements and vandalism of public assets (Perez et al. 2004).

The AtollGame experiment aims at providing the relevant information to the local actors, including institutional and local community representatives, in order to facilitate dialogue and to help devise together sustainable and equitable water management practices. Knowledge elicitation techniques as well as Multi Agent-Based Simulations (MABS) coupled with a role-playing game have been implemented to fulfil this aim. In order to collect, understand and merge viewpoints coming from different stakeholders, the following five-stage methodology is applied: (1) collecting local and expert knowledge; (2) blending the different viewpoints into a game-based model; (3) playing the game with the different stakeholders; (4) formalising the different scenarios investigated in computer simulations; and (5) exploring the simulated outcomes with the different stakeholders (Dray et al. 2006b).

Initial knowledge elicitation (Stages 1 and 2) relies on three successive methods. First, a Global Targeted Appraisal focuses on social group leaders in order to collect different standpoints and their articulated mental models. These collective models are partly validated through Individual Activities Surveys focusing on behavioural patterns of individual islanders. Then, these individual representations are merged into one collective model using qualitative analysis techniques. This conceptual model is further simplified in order to create a computer-assisted role-playing game (AtollGame). The range of contrasted viewpoints confirms the need for an effective consultation, and engagement of the local population in the design of future water management schemes in order to warrant the long-term sustainability of the system. Clear evidence of the inherent duality between land and water use rights on one hand, and between water exploitation and distribution on the other hand, provides essential features to frame the computer-assisted Role Playing Game.

The assistance of a computer is needed as far as interactions between groundwater dynamics and surface water balance involve complex spatial and time-dependent interactions (Perez et al. 2004). The use of Agent-Based Modelling (ABM) enables us to take full advantage of the structure of the conceptual model. We developed the AtollGame simulator with the CORMAS© platform (Bousquet et al. 1998).[1]

A board game version reproduces the main features of the AtollGame simulator (Dray et al. 2006a). 16 players – 8 on each island – are able to interact according to a set of pre-defined rules. Their choices and actions are directly incorporated into the simulator at the end of each round of the game. During the game, players can ask for more information from the simulator or discuss the results provided by the simulator (salinity index, global demand). Landowners, traditional or new buyers, are the essential actors in the negotiations with the government. The connection between land tenure issues and water management is essential. It drives the land use restrictions and land leases discussions. The population increase, mainly through immigration, is perceived as a threat in terms of water consumption, pollution generation and pressure on the land. Financial issues linked with water management usually deal with land leases, equipment investment and, seldom, with water pricing. Hence, the model features:

– Agents/players becoming a local landowners;
– Land and water allocation conflicting rules, and various sources of incomes;
– An increasing number of new settlers on agents/players' land.

The individual objective of the players is to minimize the number of angry or sick people in their house. People may become ANGRY because they didn't have enough water to drink during the round. People may become SICK if they drank unhealthy (polluted or salty) water during the round. POLLUTION depends on the number of people living on the island and contaminating the freshwater lens. SALTY WATER depends on the recharge rate of the fresh water lens and the location of the people on the island.

At first, representatives from the different islands displayed different viewpoints about the Water Reserves. Hence, the group meetings organized in the villages prior

[1] More details about the AtollGame can be found online at http://cormas.cirad.fr/en/applica/atollGame.htm.

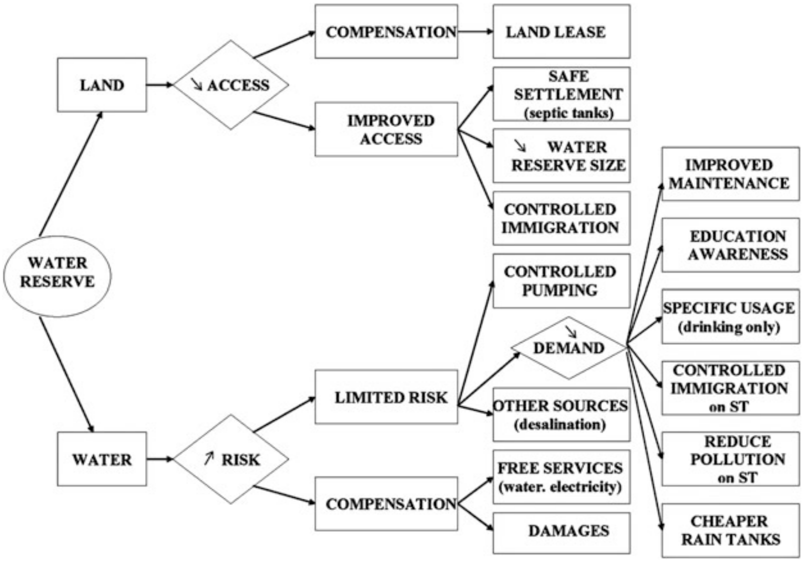



**Fig. 10.4** Flowchart of financial, technical, and social solutions agreed on by the participants of the AtollGame experiment

the workshop allowed for a really open debate. On the institutional side, the position of the different officers attending the workshop demonstrated a clear commitment to the project. All the participants showed the same level of motivation either to express their views on the issue or to genuinely try to understand other viewpoints. Participants also accepted to follow the rules proposed by the project team, especially the necessity to look at the problem from a broader perspective. During the first rounds, the players quickly handled the game and entered into interpersonal discussions and comparisons. The atmosphere was good and the game seemed playful enough to maintain the participants' interest alive. The second day, the introduction of a Water Management Agency and the selection of its (virtual) Director created a little tension among the participants. But, after a while, the players accepted the new situation as a gaming scenario and started to interact with the newly created institution. At this stage, players started to mix arguments based on the game with other ones coming directly from the reality. On Island 1, players entered direct negotiations with the (virtual) Director of the Water Management Agency. On Island 2, discussions opposed players willing or not to pay the fee.

Finally, the project team introduced the fact that the Water Management Agency was no longer able to maintain the reticulated system due to a poor recovery of the service fees. It had for immediate consequence a sharp decrease of the water quantity offered on Island 2.

Then, players from both tables were asked to list solutions to improve the situation on their island. When the two lists were completed, the project team and the participants built a flowchart of financial, technical and social solutions, taking into account issues from both islands (Fig. 10.4).

A collective analysis of the flowchart concluded that the actual situation was largely unsustainable either from a financial or social viewpoint. The flowchart above provides a set of inter-dependent solutions that should be explored in order to gradually address the present situation.

## 10.6 Discussion: Relations Between Participants and Models

The diverse categories of joint implementation of participatory approaches and social simulation modelling feature a diversity of relations between a set of people, participants, and a model.

Classical social simulation models do not feature any participant. People are represented in the model, sometimes from assumed or theoretical behavioural patterns. This entails exploring potential emergent phenomena from interactions among these behavioural patterns. Some participatory approaches involve only an implicit social model. Within this scope, there is a large diversity of relations. This diversity is based on the role undertaken by stakeholders, their actual involvement and issues tackled by the model.

In all the processes allying social simulation models and participation, stakeholders take on various roles: pieces in simulation, interfaces for coupling various sources of knowledge, beneficiaries of the process, key informants... As pointed out by Ryan, managers are overwhelmed by the complexity to be managed. Participation is a way to share this burden (Ryan 2000). Stakeholders provide the missing interactions and add missing pieces of knowledge, such as tacit knowledge (Johannessen et al. 2001). If involvement of stakeholders is useful for principal agents such as managers, we propose it as a rule that they should gain some empowerment in the process.

Stakeholders can be key pieces of the modelling process itself as well. In the simulation they are an alternative to computer code to provide the engine (Hanneman 1995). They provide an answer to issues of coupling several viewpoints (Robinson 1991).

However actual involvement of people in a participatory modelling process might largely differ from formal involvement planned. Leaving aside cases of manipulation and announce effect, people have also to find their place in the participatory process. Suitability of participatory approaches in a specific society has to be taken into account: context (including social) is a key driver for success in stakeholder involvement (Kujala 2003), and practice of interactive policy making processes depend on local culture (Driessen et al. 2001). Representation mechanisms have already be pointed out as a major factor. It has to be tuned to this local social and cultural context. At a finer grain, facilitator has a key role to lead people towards the level of involvement they are invited (Akkermans and Vennix 1997).

## 10.7 Conclusion

This chapter provides a review of the diversity of association of participatory approaches and social simulation, for their mutual benefit. This diversity of approaches allows tackling expectations about increasing model's quality, model's suitability to its intended use and improving participation. Their diversity is built upon ingredients coming from various disciplines from social sciences to computer sciences and management. It is expressed according to the implementation of interactions between the participants and the simulation model, the control of the process and the format of information. This leads to expand the classical ladder of participation towards categorization according to the stage in the modelling process when participation takes place and the structure of the interaction to cope with the heterogeneity of stakeholders.

This diversity requires a cautious description of each implementation in situation, so that any evaluation is specific to the implementation of a given association in its context. Generalisation can then be done only on the relation of this practice of participatory simulation and its suitability to its context and purpose. Efficiency to induce changes in practice or knowledge depends on the respect of a triple contingency of collective decision processes: time, people, means (Miettinen and Virkkunen 2005). This means to respect and take into account the own dynamics within the social system at stake, to allow the participation of people with their whole essence (including tacit knowledge, networks, relations to the world), and to be adaptive to means and competences present within the system (Barreteau 2007). Another dimension of evaluation should be democracy, since it is often put to the front. This raises the issue of the existence of a control of the process. Does it rely only on modellers or is it more shared? Finally there is a necessity of being more explicit on the kind of PA which is used because of the potential deconsideration of the whole family if expectations are deceived.

## Further Reading

Participatory modelling is increasingly present in special sessions of conferences or special features of scientific journals. A first source of further readings consists in case studies. Among others, Environmental Modelling and Software had a special issue on modelling with stakeholders (Bousquet and Voinov 2010), where readers will find a whole set of well described case studies using various methods. The biennial international Environmental Modelling and Software conferences have also specific tracks for participatory modelling; proceedings are available online (see http://www.iemss.org/society/ under publications). For specific tools, refer to the papers of a symposium on simulation and gaming in natural resource management, published as a special issue of Simulation and Gaming (volume 38, issues 2 and 3). The introductory paper giving an overview is (Barreteau et al. 2007).

Reflexivity is crucial for practitioners of participatory processes, as part of the need for more cautious evaluation of participatory processes as pointed out by Rowe and Frewer (2004). Another direction for reading consists in methods for evaluation and assessment of stakeholder involvement in modelling processes. Etienne edited a whole book aiming at assessing consequences of a specific approach, so-called companion modelling (Étienne 2011).

Readers who are more interested in stakeholder involvement in modelling at a more technical level should go for the review paper of Ramanath and Gilbert (2004) which provides a nice overview of this point of view.

## References


Abel N, Ross H, Walker P (1998) Mental models in rangeland research, communication and management. Rangel J 20:77–91

Akkermans HA (1995) Developing a logistics strategy through participative business modelling. Int J Oper Prod Manag 15:100–112

Akkermans HA, Vennix JAM (1997) Clients' opinions on group model building: an exploratory study. Syst Dyn Rev 13:3–31

Arnstein S (1969) A ladder of citizen participation. J Am Plann Assoc 35:216–224

Barreteau O (2003) The joint use of role-playing games and models regarding negotiation processes: characterization of associations. J Artif Soc Soc Simul 6(2), http://jasss.soc.surrey.ac.uk/6/2/3.html

Barreteau O (2007) Modèles et processus de décision collective: entre compréhension et facilitation de la gestion concertée de la ressource en eau (HDR Thesis). Paris Dauphine University, Paris

Barreteau O, Bousquet F (1999) Jeux de rôles et validation de systèmes multi-agents. In: Gleizes M-P, Marcenac P (eds) Ingénierie des systèmes multi-agents, actes des 7$^{\text{èmes}}$ JFIADSMA. Hermès, Paris, pp 67–80

Barreteau O, Bousquet F, Attonaty J-M (2001) Role-playing games for opening the black box of multi-agent systems: method and teachings of its application to Senegal River Valley irrigated systems. J Artif Soc Soc Simul 4(2), http://jasss.soc.surrey.ac.uk/4/2/5.html

Barreteau O, Hare M, Krywkow J, Boutet A (2005) Model designed through participatory processes: whose model is it? In: Ferrand N, Perez P, Batten D (eds) Joint conference on multiagent modelling for environmental management, CABM-HEMA-SMAGET 2005. Bourg St Maurice – Les Arcs, 21–25 Mar 2005

Barreteau O, Le Page C, Perez P (2007) Contribution of simulation and gaming to natural resource management issues: An introduction. Simul Gaming 38(2):185–194

Becu N (2006) Identification et modélisation des représentations des acteurs locaux pour la gestion des bassins versants. PhD thesis, Sciences de l'eau, Université Montpellier 2, Montpellier, France

Becu N, Barreteau O, Perez P, Saising J, Sungted S (2006) A methodology for identifying and formalizing farmers' representations of watershed management: a case study from Northern Thailand. In: Bousquet F, Trebuil G, Hardy B (eds) Companion modeling and multi-agent systems for integrated natural resource management in Asia. IRRI, Los Baños, pp 41–62

Bharwani S (2006) Understanding complex behavior and decision making using ethnographic knowledge elicitation tools (KnETs). Soc Sci Comput Rev 24:78–105

Bousquet F, Voinov A (eds) (2010) Thematic issue – modelling with stakeholders. Environ Modell Softw 25(11):1267–1488

Bousquet F, Bakam I, Proton H,Le Page C (1998) Cormas: common-pool resources and multi-agent systems. In: Lecture notes in artificial intelligence 1416, Springer, Berlin pp 826–838
Bousquet F, Barreteau O, Le Page C, Mullon C, Weber J (1999) An environmental modelling approach: the use of multi-agent simulations. In: Blasco F, Weill A (eds) Advances in environmental and ecological modelling. Elsevier, Amsterdam, pp 113–122
Bousquet F et al (2002) Multi-agent systems and role games: an approach for ecosystem co-management. In: Janssen M (ed) Complexity and ecosystem management: the theory and practice of multi-agent approaches. Edward Elgar, Northampton, pp 248–285
Callon M, Muniesa F (2006) Economic experiments and the construction of markets. In: MacKenzie D, Muniesa F, Siu L (eds) Do economists make markets? On the performativity of economics. Princeton University Press, Princeton, pp 163–189
Cardenas J-C, Stranlund J, Willis C (2000) Local environmental control and institutional crowding-out. World Dev 28:1719–1733
Castella JC, Tran Ngoc T, Boissau S (2005) Participatory simulation of land-use changes in the northern mountains of Vietnam: the combined use of an agent-based model, a role-playing game, and a geographic information system. Ecol Soc 10(1):27, http://www.ecologyandsociety.org/vol10/iss1/art27/
Cockes D, Ive J (1996) Mediation support for forest land allocation: the SIRO-MED system. Environ Manage 20(1):41–52
Daalen CE van, Bots PWG (2006) Participatory model construction and model use in natural resource management. In: Proceedings of the workshop on formalised and non-formalised methods in resource management – knowledge and learning in participatory processes, Osnabrück, 21–22 Sept 2006, http://www.partizipa.uni-osnabrueck.de/wissAbschluss.html
Daniell KA, Ferrand N, Tsoukias A (2006) Investigating participatory modelling processes for group decision aiding in water planning and management. In: Seifert S, Weinhardt C (eds) Proceedings of group decision and negotiation (GDN) 2006, International conference, Karlsruhe, 25–28 June 2006, Universitätsverlag Karlsruhe, Karlsruhe, pp 207–210
D'Aquino P, Le Page C, Bousquet F, Bah A (2003) Using self-designed role-playing games and a multi-agent system to empower a local decision-making process for land use management: the SelfCormas experiment in Senegal. J Artif Soc Soc Simul 6(3), http://jasss.soc.surrey.ac.uk/6/3/5.html
Davidson EJ (1999) Joint application design (JAD) in practice. J Syst Softw 45:215–223
DeSanctis G, Gallupe RB (1987) A foundation for the study of group decision support systems. Manag Sci 33:589–609
Dewulf A, Bouwen R, Tailleu T (2006) The multi-actor simulation 'Podocarpus National Park' as a tool for teaching and researching issue framing. In: Proceedings of IACM 2006 Montreal Meetings. http://ssrn.com/abstract=915943
Dray A et al (2006a) The AtollGame experience: from knowledge Engineering to a computer-assisted role playing game. J Artif Soc Soc Simul 9(1), http://jasss.soc.surrey.ac.uk/9/1/6.html
Dray A, Perez P, Le Page C, D'Aquino P, White I (2006b) AtollGame: a companion modelling experience in the pacific. In: Perez P, Batten D (eds) Complex science for a complex world: exploring human ecosystems with agents. ANU E Press, Canberra, pp 255–280
Driessen PPJ, Glasbergen P, Verdaas C (2001) Interactive policy making: a model of management for public works. Eur J Oper Res 128:322–337
Drogoul A, Vanbergue D, Meurisse T (2003) Multi-agent based simulation: where are the agents? In: Sichman JS, Bousquet F, Davidsson P (eds) Multi-agent-based simulation II: third international workshop, MABS 2002, Bologna, Revised papers (Lecture notes in computer science, 2581), Springer, Berlin/Heidelberg, 15–16 July 2002, pp 1–15
Dudley RG (2003) Modeling the effects of a log export ban in Indonesia. Syst Dyn Rev 20:99–116
Duke RD, Geurts JLA (2004) Policy games for strategic management. Dutch University Press, Amsterdam
Étienne M (2006) Companion modelling: a tool for dialogue and concertation in biosphere reserves. In: Bouamrane M (ed) Biodiversity and stakeholders: concertation itineraries, Biosphere reserves – technical notes 1, Unesco, Paris, pp 44–52

Étienne M (ed) (2011) Companion modeling: a participatory approach to support sustainable development. QUAE, Versailles
Etienne M (2003) SYLVOPAST a multiple target role-playing game to assess negotiation processes in sylvopastoral management planning. Journal of Artificial Societies and Social Simulations 6, http://jasss.soc.surrey.ac.uk/6/2/5.html
Étienne M, Le Page C, Cohen M (2003) A step by step approach to build up land management scenarios based on multiple viewpoints on multi-agent systems simulations. J Artif Soc Soc Simul 6(2), http://jasss.soc.surrey.ac.uk/6/2/2.html
Eversole R (2003) Managing the pitfalls of participatory development: some insight from Australia. World Dev 31:781–795
Fischer G, Giaccardi E, Eden H, Sugimoti M, Ye Y (2005) Beyond binary choices: integrating individual and social creativity. Int J Hum-Comput Stud 63:482–512
Friedman D, Sunder S (1994) Experimental methods, a primer for economists. Cambridge University Press, Cambridge
Fung A (2006) Varieties of participation in complex governance. Public Adm Rev 66:66–75
Funtowicz SO, Martinez-Alier J, Munda G, Ravetz JR (1999) Information tools for environmental policy under conditions of complexity, vol 9, Environmental issues series. European Environment Agency, Copenhagen
Gilbert N, Troitzsch KG (1999) Simulation for the social scientist. Open University Press, Buckingham
Green KC (2002) Forecasting decisions in conflict situations: a comparison of game theory, role-playing and unaided judgment. Int J Forecast 18:321–344
Guyot P (2006) Simulations multi-agents participatives. PhD thesis, Informatique, Université Paris VI, Paris
Guyot P, Honiden S (2006) Agent-based participatory simulations: merging multi-agent systems and role-playing games. J Artif Soc Soc Simul 9(4), http://jasss.soc.surrey.ac.uk/9/4/8.html
Hamel A, Pinson S (2005) Conception participative de simulations multi-agents basée sur une approche d'analyse multi-acteurs. In: Drogoul A, Ramat E (eds) Systèmes multi-agents: Vers la conception de systèmes artificiels socio-mimétiques (JFSMA 2005). Hermès, Paris, pp 1–15
Hanneman RA (1995) Simulation modeling and theoretical analysis in sociology. Sociol Perspect 38:457–462
Henriksen HJ, Rasmussen P, Brandt G, von Bülow D, Jensen FV (2004) Engaging stakeholders in construction and validation of Bayesian belief networks for groundwater protection. In: Proceedings of IFAC workshop on modelling and control for participatory planning and managing water systems, Venice, 29 Sept–1 Oct 2004
Henry C (1974) Investment decisions under uncertainty: the irreversibility effect. Am Econ Rev 64:1006–1012
Hochman Z, Hearnshaw H, Barlow R, Ayres JF, Pearson CJ (1995) X-breed: a multiple domain knowledge based system integrated through a blackboard architecture. Agr Syst 48:243–270
Irvin RA, Stansbury J (2004) Citizen participation in decision making: is it worth the effort. Public Adm Rev 64:55–65
Johannessen J-A, Olaisen J, Olsen B (2001) Mismanagement of tacit knowledge: the importance of tacit knowledge, the danger of information technology, and what to do about it. Int J Inf Manag 21:3–20
Kinzig A, Ryan P, Etienne M, Allyson H, Elmqvist T, Walker B (2006) Resilience and regime shifts: assessing cascading effects. Ecol Soc 11(1):20, http://www.ecologyandsociety.org/vol11/iss1/art20
Klopfer E, Yoon S, Rivas L (2004) Comparative analysis of palm and wearable computers for participatory simulations. J Comput Assist Lear 20:347–359
Kujala S (2003) User involvement: a review of the benefits and challenges. Behav Inf Technol 22:1–16
Landry M, Banville C, Oral M (1996) Model legitimation in operational research. Eur J Oper Res 92:443–457

Le Bars M, Le Grusse P, Allaya M, Attonaty J-M, Mahjoubi R (2004) NECC: Un jeu de simulation pour l'aide à la décision collective; Application à une région méditerranéenne virtuelle. In: Projet INCO-WADEMED, Séminaire Modernisation de l'Agriculture Irriguée, Rabat

Loucks DP, Kindler J, Fedra K (1985) Interactive water resources modeling and model use: an overview. Water Resour Res 21:95–102

Maier FH, Grössler A (2000) What are we talking about? A taxonomy of computer simulations to support learning. Syst Dynam Rev 16:135–148

Manson SM (2002) Validation and verification of multi-agent systems. In: Janssen M (ed) Complexity and ecosystem management: the theory and practice of multi-agent approaches. Edward Elgar, Northampton, pp 63–74

Marengo L, Pasquali C (2003) How to construct and share a meaning for social interactions? In Conventions et Institutions: Approfondissements théoriques et Contributions au Débat Politique, Paris

Martin L et al (2007) Microworld gaming of a local agricultural production chain in Poland. Simul Gaming 38(2):211–232

Mayer IS, van Bueren EM, Bots PWG, van der Voort HG, Seijdel RR (2005) Collaborative decision-making for sustainable urban renewal projects: a simulation-gaming approach. Environ Plan B Plan Des 32:403–423

McKinnon J (2005) Mobile interactive GIS: bringing indigenous knowledge and scientific information together; a narrative account. In: Neef A (ed) Participatory approaches for sustainable land use in Southeast Asia. White Lotus, Bangkok, pp 217–231

Mermet L (1993) Une méthode de prospective: Les exercices de simulation de politiques. Nature Sciences Sociétés 1:34–46

Miettinen R, Virkkunen J (2005) Epistemic objects, artefacts and organizational change. Organization 12:437–456

Molin E (2005) A causal analysis of hydrogen acceptance. Transport Res Rec 1941:115–121

Moss S, Downing T, Rouchier J (2000) Demonstrating the role of stakeholder participation: an agent based social simulation model of water demand policy and response (CPM report, 00–76). Centre for Policy Modelling, Manchester Metropolitan University, Manchester, http://cfpm.org/cpmrep76.html

Mostert E (2006) Participation for sustainable water management. In: Giupponi C, Jakeman AJ, Karssenberg D, Hare MP (eds) Sustainable management of water resources. Edward Elgar, Northampton, pp 153–176

Nancarrow B (2005) When the modeller meets the social scientist or vice-versa. In: Zerger A, Argent RM (eds) MODSIM 2005 international congress on modelling and simulation. Modelling and Simulation Society of Australia and New Zealand, Melbourne, pp 38–44

Ostrom E, Gardner R, Walker J (1994) Rules, games and common-pool resources. The University of Michigan Press, Ann Arbor

Pahl-Wostl C, Hare M (2004) Processes of social learning in integrated resources management. J Community Appl Soc Psychol 14:193–206

Pateman C (1990) Participation and democratic theory. Cambridge University Press, Cambridge

Perez P, Dray A, Le Page C, D'Aquino P, White I (2004) Lagoon, agents and kava: a companion modelling experience in the Pacific. In: van Dijkum C, Blasius J, Durand C (eds) Recent developments and applications in social research methodology: Proceedings of RC33 sixth international conference on social science methodology, Amsterdam Barbara Budrich Publishers, Opladen, p 282

Ramanath AM, Gilbert N (2004) The design of participatory agent based simulations. J Artif Soc Soc Simul 7(4), http://jasss.soc.surrey.ac.uk/7/4/1.html

Reitsma R, Zigurs I, Lewis C, Wilson V, Sloane A (1996) Experiment with simulation models in water-resources negotiations. J Water Resour Plan Manag 122:64–70

Richard A, Barreteau O (2006) Concert'eau: un outil de sociologie expérimentale pour l'étude de dispositifs de gestion locale et concertée de l'eau. In: Proceedings of 2e Congrès de l'Association Française de Sociologie, Bordeaux, 5–8 Sept 2006

Richard A, Trometter M (2001) Les caractéristiques d'une décision séquentielle: Effet irréversibilité et endogénéisation de l'environnement. Revue Economique 52:739–752

Rittel HWJ, Webber MM (1973) Dilemmas in a general theory of planning. Policy Sci 4:155–169

Robinson JB (1991) Modelling the interactions between human and natural systems. Int Soc Sci J 130:629–647

Rocha EM (1997) A ladder of empowerment. J Plan Edu Res 17:31–44

Rouchier J (2003) Re-implementation of a multi-agent model aimed at sustaining experimental economic research: the case of simulations with emerging speculation. J Artif Soc Soc Simul 6(4), http://jasss.soc.surrey.ac.uk/6/4/7.html

Rouwette EAJA, Vennix JAM, van Mullekorn T (2002) Group model building effectiveness: a review of assessment studies. Syst Dyn Rev 18:5–45

Rowe G, Frewer LJ (2004) Evaluating public-participation exercises: a research agenda. Sci Technol Hum Val 29(4):512–556

Ryan T (2000) The role of simulation gaming in policy making. Syst Res Behav Sci 17:359–364

Schelling TC (1961) Experimental games and bargaining theory. World Polit 14:47–68

Schuler D, Namioka A (eds) (1993) Participatory design: principles and practices. Lawrence Erlbaum, Hillsdale

Shakun ME (1996) Modeling and supporting task-oriented group processes: purposeful complex adaptive systems and evolutionary systems design. Group Decis Negot 5:305–317

Simon H (1977) The new science of management decision. Prentice-Hall, Englewood Cliffs

Star SL, Griesemer JR (1989) Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berleley's museum of vertebrate zoology, 1907–39. Soc Stud Sci 19:387–420

Sterman JD (1992) Teaching takes off – flight simulators for management education. OR/MS Today, pp 40–44

Ticehurst J, Rissik D, Letcher RA, Newham LTH, Jakeman AJ (2005) Development of decision support tools to assess the sustainability of coastal lakes. In: Zerger A, Argent RM (eds) MODSIM 2005 international congress on modelling and simulation, Modelling and Simulation Society of Australia and New Zealand, Melbourne, pp 2414–2420, http://www.mssanz.org.au/modsim05/papers/ticehurst.pdf

Toth FL (1988) Policy exercises: objectives and design elements. Simul Games 19:235–255

van Asselt M et al (2001) Building blocks for participation in integrated assessment: a review of participatory methods. ICIS, Maastricht

van Daalen CE, Thissen WAH, Berk MM (1998) The Delft process: experiences with a dialogue between policy makers and global modellers. In: Alcamo J, Leemans R, Kreileman E (eds) Global change scenarios of the 21st century: results from the IMAGE 2.1 model. Elsevier, London, pp 267–285

Van den Belt M (2004) Mediated modeling: a system dynamics approach to environmental consensus building. Island Press, Washington, DC

Vennix JAM (1996) Group model building, facilitating team learning using system dynamics. Wiley, Chichester

Webler T, Kastenholz H, Renn O (1995) Public participation in impact assessment: a social learning perspective. Environ Impact Assess Rev 15:443–463

Whitworth B, Gallupe B, McQueen R (2000) A cognitive three-process model of computer mediated group interaction. Gr Decis Negot 9:431–456

Wood J, Silver D (1995) Joint application development. Wiley, New York

# Chapter 11
# Combining Mathematical and Simulation Approaches to Understand the Dynamics of Computer Models

**Luis R. Izquierdo, Segismundo S. Izquierdo, José M. Galán, and José I. Santos**

**Why Read This Chapter?** To learn how to better understand the dynamics of computer models using both simulation and mathematical analysis. Our starting point is a computer model which is already implemented and ready to be run; the objective is to gain a thorough understanding of its dynamics. Combining computer simulation with mathematical analysis can help to provide a picture of the model dynamics that could not be drawn by using only one of the two techniques.

**Abstract** This chapter shows how computer simulation and mathematical analysis can be used together to understand the dynamics of computer models. For this purpose, we show that it is useful to see the computer model as a particular implementation of a formal model in a certain programming language. This formal model is the abstract entity which is defined by the input-output relation that the computer model executes, and can be seen as a function that transforms probability distributions over the set of possible inputs into probability distributions over the set of possible outputs.

It is shown here that both computer simulation and mathematical analysis are extremely useful tools to analyse this formal model, and they are certainly complementary in the sense that they can provide fundamentally different insights on the same model. Even more importantly, this chapter shows that there are plenty of synergies to be exploited by using the two techniques together.

The mathematical approach to analyse formal models consists in examining the rules that define the model directly. Its aim is to deduce the logical implications of these rules for any particular instance to which they can be applied. Our analysis of mathematical

L.R. Izquierdo (✉) • J.M. Galán • J.I. Santos
Departamento de Ingeniería Civil, Universidad de Burgos, E-09001, Burgos, Spain
e-mail: luis@izquierdo.name; jmgalan@ubu.es; jisantos@ubu.es

S.S. Izquierdo
Departamento de Organización de Empresas y C.I.M., Universidad de Valladolid, E-47011, Valladolid, Spain
e-mail: segis@eis.uva.es

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_11,

techniques to study formal models is focused on the theory of Markov Chains, which is particularly useful to characterise the dynamics of computer models.

In contrast with mathematical analysis, the computer simulation approach does not look at the rules that define the formal model directly, but instead tries to infer general properties of these rules by examining the outputs they produce when applied to particular instances of the input space. Thus, conclusions obtained with this approach may not be general. On a more positive note, computer simulation enables us to explore formal models beyond mathematical tractability, and we can achieve any arbitrary level of accuracy in our computational approximations by running the model sufficiently many times.

Bearing in mind the relative strengths and limitations of both approaches, this chapter explains three different ways in which mathematical analysis and computer simulation can be usefully combined to produce a better understanding of the dynamics of computer models. In doing so, it becomes clear that mathematical analysis and computer simulation should not be regarded as alternative – or even opposed – approaches to the formal study of social systems, but as complementary. Not only can they provide fundamentally different insights on the same model, but they can also produce hints for solutions for each other. In short, there are plenty of synergies to be exploited by using the two techniques together, so the full potential of each technique cannot be reached unless they are used in conjunction.

## 11.1 Introduction

This chapter is about how to better understand the dynamics of computer models using both simulation and mathematical analysis. Our starting point is a computer model which is already implemented and ready to be run; our objective is to gain a thorough understanding of its dynamics. Thus, this chapter is *not* about how to design, implement, verify, or validate a model; this chapter is about how to better understand its behaviour.

Naturally, we start by clearly defining our object of study: a computer model. The term ‘computer model’ can be understood in many different ways – i.e. seen from many different perspectives –, and not all of them are equally useful for every possible purpose. Thus, we start by interpreting the term ‘computer model’ in a way that will prove useful for our objective: to characterise and understand its behaviour. Once our object of study has been clearly defined, we then describe two techniques that are particularly useful to understand the dynamics of computer models: mathematical analysis and computer simulation.

In particular, this chapter will show that mathematical analysis and computer simulation should not be regarded as alternative – or even opposed – approaches to the formal study of social systems, but as complementary (Gotts et al. 2003a, b). They are both extremely useful tools to analyse formal models, and they are certainly complementary in the sense that they can provide fundamentally different insights on the same model. Even more importantly, this chapter will show that there are plenty of synergies to be exploited by using the two techniques together, i.e. the full potential of each technique will not be reached until they are used in conjunction. The remaining of this introduction outlines the structure of the chapter.

Sections 11.2, 11.3 and 11.4 are devoted to explaining in detail what we understand by 'computer model', and they therefore provide the basic framework for the rest of the chapter. In particular, Sect. 11.2 shows that a computer model can be seen as an implementation – i.e. an explicit representation – of a certain deterministic input-output function in a particular programming language. This interpretation is very useful since, in particular, it will allow us to abstract from the details of the modelling platform where the computer model has been programmed, and focus on analysing the formal model that the computer model implements. This is clarified in Sect. 11.3, which explains that any computer model can be re-implemented in many different formalisms (in particular, in any sophisticated enough programming language), leading to alternative representations of the same input-output relation.

Most computer models in the Social Simulation literature make use of pseudo-random number generators. Section 11.4 explains that – for these cases and given our purposes – it is useful to abstract from the details of how pseudo-random numbers are generated, and look at the computer model as an implementation of a stochastic process. In a stochastic process, a certain input does not necessarily lead to one certain output only; instead, there are many different paths that the process may take with potentially different probabilities. Thus, in a stochastic process a certain input will generally lead to a particular probability distribution over the range of possible outputs, rather than to a single output only. Stochastic processes are used to formally describe how a system subjected to random events evolves through time.

Having explained our interpretation of the term 'computer model', Sect. 11.5 introduces and compares the two techniques to analyse formal models that are assessed in this chapter: computer simulation and mathematical analysis. The following two sections sketch possible ways in which each of these two techniques can be used to obtain useful insights about the dynamics of a model. Section 11.8 is then focused on the *joint use* of computer simulation and mathematical analysis. It is shown here that the two techniques can be used together to provide a picture of the dynamics of the model that could not be drawn by using one of the two techniques only. Finally, our conclusions are summarised in Sect. 11.9.

## 11.2 Computer Models as Input-Output Functions

At the most elementary level, a computer model can be seen as an implementation – i.e. an explicit representation – of a certain deterministic input-output function in a particular programming language. The word 'function' is useful because it correctly conveys the point that any particular input given to the computer model will lead to one and only one output.[1] (Obviously, different inputs may lead to the same output.) Admittedly, however, the word ‘function’ may also mislead the reader into thinking that a computer model is necessarily simple. The computer model may be as complex and sophisticated as the programmer wants it to be but, ultimately, it is just an entity that associates a specific output to any given input, i.e. a function. In any case, to avoid confusion, we will use the term ‘formal model’ to denote the function that a certain computer model implements.[2] To be sure, the ‘formal model’ that a particular computer model implements is the abstract entity which is defined by the input-output relation that the computer model executes.[3]

Thus, running a computer model is just finding out the logical implications of applying a set of unambiguously defined formal rules (which are coded in the program and define the input-output function or formal model) to a set of inputs (Balzer et al. 2001). As an example, one could write the computer program “$y = 4x$” and apply it to the input “$x = 2$” to obtain the output “$y = 8$”. The output ($y = 8$), which is fully and unequivocally determined by the input ($x = 2$) and the set of rules coded in the program ($y = 4x$), can be seen as a theorem obtained by pure deduction ($\{x = 2; y = 4x\} \Rightarrow y = 8$). Naturally, there is no reason why the inputs or the outputs should be numbers[4]; they could equally well be e.g. strings of characters. In the general case, a computer run is a logical theorem that reads: *the output obtained from running the computer simulation follows* (*with logical necessity*) *from applying to the input the algorithmic rules that define the model*. Thus, regardless of its inherent complexity, a computer run constitutes a perfectly valid sufficiency theorem (see e.g. Axtell 2000).

It is useful to realise that we could always apply the same inference rules ourselves to obtain – by logical deduction – the same output from the given input. While useful as a thought, when it comes to actually doing the job, it is much more convenient, efficient and less prone to errors to let computers derive the output for us. Computers are inference engines that are able to conduct many algorithmic processes at a speed that the human brain cannot achieve.

[1] Note that simulations of stochastic models are actually using pseudo-random number generators, which are deterministic algorithms that require a seed as an input.

[2] A formal model is a model expressed in a formal system (Cutland 1980). A formal system consists of a formal language and a deductive apparatus (a set of axioms and inference rules). Formal systems are used to derive new expressions by applying the inference rules to the axioms and/or previously derived expressions in the same system.

[3] The mere fact that the model has been implemented and can be run in a computer is a proof that the model is formal (Suber 2002).

[4] As a matter of fact, strictly speaking, inputs and outputs in a computer model are *never* numbers. We may interpret strings of bits as numbers, but we could equally well interpret the same strings of bits as e.g. letters. More importantly, a bit itself is already an abstraction, an interpretation we make of an electrical pulse that can be above or below a critical voltage threshold.

## 11.3 Different Ways of Representing the Same Formal Model

A somewhat controversial issue in the Social Simulation literature refers to the allegedly unique features of some modelling platforms. It is important to realise that any formal model implemented in a computer model can be re-implemented in many different programming languages, leading to exactly the same input-output relation. Different implementations are just different ways of representing one same formal model, much in the same way that one can say 'Spain' or 'España' to express the same concept in different languages: same thing, different representation, that's all.

Thus, when analysing the dynamics of a computer model, it is useful to abstract from the details of the modelling platform that has been used to implement the computer model, and focus strictly on the formal model it represents, which could be re-implemented in any *sophisticated enough*[5] modelling platform. To be clear, let us emphasise that *any* computer model implemented in Objective-C (e.g. using Swarm) can be re-implemented in Java (e.g. using RePast or Mason), NetLogo, SDML, Mathematica© or Matlab©. Similarly, *any* computer model can be expressed as a well-defined mathematical function (Epstein 2006; Leombruni and Richiardi 2005; Richiardi et al. 2006).

Naturally, the implementation of a particular formal model may be more straightforward in some programming languages than in others. Programming languages differ in where they position themselves in the well-known trade-offs between ease of programming, functionality and performance; thus, different programming languages lead to more or less natural and more or less efficient implementations of any given formal model. Nonetheless, the important point is this: whilst we may have different implementations of the same formal model, and whilst each of these implementations may have different characteristics (in terms of e.g. code readability), ultimately they are all just different representations of the same formal model, and they will therefore return the same output when given the same input.

In the same way that using one or another formalism to represent a particular formal model will lead to more or less natural implementations, different formalisms also make more or less apparent certain properties of the formal model they implement. For example, we will see in this chapter that representing a computer model as a Markov chain, i.e. looking at the formal model implemented in a

[5] A sufficient condition for a programming language to be "sophisticated enough" is to allow for the implementation of the following three control structures:

- Sequence (i.e. executing one subprogram, and then another subprogram),
- Selection (i.e. executing one of two subprograms according to the value of a boolean variable, e.g. IF[boolean == true]-THEN[subprogram1]-ELSE[subprogram2]), and
- Iteration (i.e. executing a subprogram until a boolean variable becomes false, e.g. WHILE [boolean == true]-DO[subprogram]).

Any programming language that can combine subprograms in these three ways can implement any computable function; this statement is known as the "structured program theorem" (Böhm and Jacopini 1966; Harel 1980; Wikipedia 2007).

computer model through Markov's glasses, can make apparent various features of the computer model that may not be so evident without such glasses. In particular, as we will show later, Markov theory can be used to find out whether the initial conditions of a model determine its asymptotic dynamics or whether they are actually irrelevant in the long term. Also, the theory can reveal whether the model will sooner or later be trapped in an absorbing state.

## 11.4 'Stochastic' Computer Models as Stochastic Processes

Most computer models in the Social Simulation literature contain stochastic components. This section argues that, for these cases and given our purposes, it is convenient to revise our interpretation of computer models as *deterministic* input-output relations, abstract from the (deterministic) details of how pseudo-random numbers are generated, and reinterpret the term 'computer model' as an implementation of a stochastic process. This interpretation will prove useful in most cases and, importantly, does not imply any loss of generality: even if the computer model to be analysed does not contain any stochastic components, our interpretation will still be valid.

In the general case, the computer model to be analysed will make use of (what are meant to be) random numbers, i.e. the model will be stochastic. The word 'stochastic' requires some clarification. Strictly speaking, there does not exist a *truly stochastic* computer model, but one can approximate randomness to a very satisfactory extent by using pseudo-random number generators. The pseudo-random number generator is a deterministic algorithm that takes as input a value called the random seed, and generates a sequence of numbers that approximates the properties of random numbers. The sequence is not truly random in that it is completely determined by the value used to initialise the algorithm, i.e. the random seed. Therefore, if given the same random seed, the pseudo-random number generator will produce exactly the same sequence of (pseudo-random) numbers. (This fact is what made us define a computer model as an implementation of a certain *deterministic* input-output function in Sect. 11.2.)

Fortunately, the sequences of numbers provided by current off-the-shelf pseudo-random number generators approximate randomness remarkably well. This basically means that, for most intents and purposes in this discipline, it seems safe to assume that the pseudo-random numbers generated in one simulation run will follow the intended probability distributions to a satisfactory degree. The only problem we might encounter appears when running several simulations which we would like to be statistically independent. As mentioned above, if we used the same random seed for every run, we would obtain the same sequence of pseudo-random numbers, i.e. we would obtain exactly the same results. How can we *truly randomly* select a random seed? Fortunately, for most applications in this discipline, the state of the computer system at the time of starting a new run can be considered a truly random variable; and, conveniently, if no seed is explicitly provided to the pseudo-random number

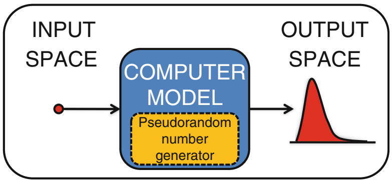



**Fig. 11.1** A computer model can be usefully seen as the implementation of a function that transforms any given input into a certain probability distribution over the set of possible outputs

generator, most platforms generate a seed from the state of the computer system (e.g. using the time). When this is done, the sequences of numbers obtained with readily available pseudo-random number generators approximate statistical randomness and independence remarkably well.

Given that – for most intents and purposes in this discipline – we can safely assume that pseudo-random numbers are random and independent enough, we dispense with the qualifier 'pseudo' from now on for convenience. Since every random variable in the model follows a specific probability distribution, the computer model will indeed generate a particular probability distribution over the range of possible outputs. Thus, to summarise, a computer model can be usefully seen as the implementation of a stochastic process, i.e. a function that transforms any given input into a certain probability distribution over the set of possible outputs (Fig. 11.1).

Seeing that we can satisfactorily simulate random variables, note that studying the behaviour of a model that has been parameterised stochastically does not introduce any conceptual difficulties. In other words, we can study the behaviour of a model that has been parameterised with probability distributions rather than certain values. An example would be a model where agents start at a random initial location.

To conclude this section, let us emphasise an important corollary of the previous paragraphs: *any statistic that we extract from a parameterised computer model follows a specific probability distribution* (even if the values of the input parameters have been expressed as probability distributions).[6] Thus, a computer model can be seen as the implementation of a function that transforms probability distributions over the set of possible inputs into probability distributions over the set of possible outputs (Fig. 11.2). The rest of the chapter is devoted to characterising this function.

[6] Note that statistics extracted from the model can be of any nature, as long as they are unambiguously defined. For example, they can refer to various time-steps, and only to certain agents (e.g. "average wealth of female agents in odd time-steps from 1 to 99").

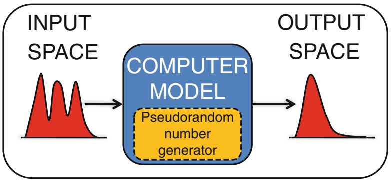



**Fig. 11.2** A computer model can be seen as the implementation of a function that transforms probability distributions over the set of possible inputs into probability distributions over the set of possible outputs

## 11.5 Tools to Understand the Behaviour of Formal Models

Once it is settled that a computer model can be seen as a particular implementation of a (potentially stochastic) function in a certain programming language, let us refer to such a function as the 'formal model' that the computer model implements. As mentioned before, this formal model can be expressed in many different formalisms – in particular, it can always be expressed as a set of well defined mathematical equations (Leombruni and Richiardi 2005) – and our objective consists in understanding its behaviour. To do that, we count with two very useful tools: mathematical analysis[7] and computer simulation.

The advantages and limitations of these two tools to formally study social systems have been discussed at length in the literature (see e.g. Axtell 2000; Axtell and Epstein 1994; Edmonds 2005; Gilbert 1999; Gilbert and Troitzsch 1999; Gotts et al. 2003a; Holland and Miller 1991; Ostrom 1988). Here we only highlight the most prominent differences between these two techniques (see Fig. 11.3).

In broad terms, when using mathematical analysis, one examines the rules that define the formal model directly, and tries to draw general conclusions about these rules. These conclusions are obtained by using logical deduction; hence they follow with logical necessity from the premises of the formal model (and the axioms of the mathematics employed). The aim when using mathematical analysis is usually to "solve" the formal system (or, most often, certain aspects of it) by producing general closed-form solutions that can be applied to *any* instance of the whole input set (or, at least, to large portions of the input set). Since the inferences obtained with mathematical analysis pertain to the rules themselves, such inferences can be safely particularised to any specific parameterisation of the model, even if such a parameterisation was never explicitly contemplated when analysing the model mathematically. This greatly facilitates conducting sensitivity analyses and assessing the robustness of the model.

Computer simulation is a rather different approach to the characterisation of the formal model (Epstein 2006; Axelrod 1997a). When using computer simulation,

[7] We use the term "mathematical analysis" in its broadest sense, i.e. we do not refer to any particular branch of mathematics, but to the general use of (any type of) mathematical technique to analyse a system.

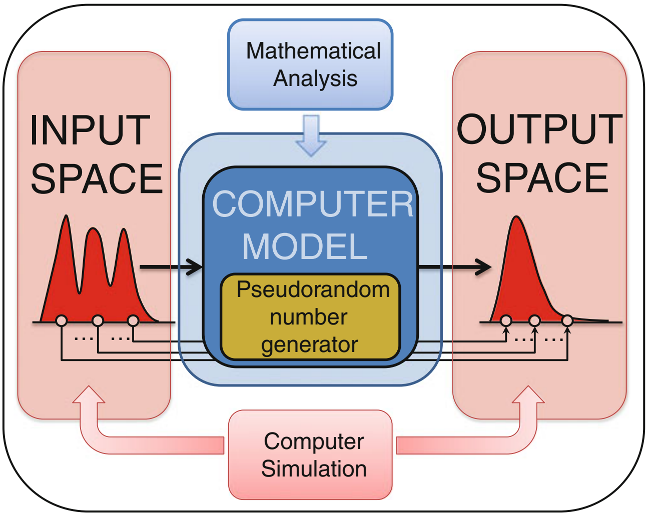



**Fig. 11.3** In general terms, mathematical analysis tends to examine the rules that define the formal model directly. In contrast, computer simulation tries to infer general properties of such rules by looking at the outputs they produce when applied to particular instances of the input space

one often treats the formal model as a black box, i.e. a somewhat obscure abstract entity that returns certain outputs when provided with inputs. Thus, the path to understand the behaviour of the model consists in obtaining many input-output pairs and – using generalisation by induction – inferring general patterns about how the rules transform the inputs into the outputs (i.e. how the formal model works).

Importantly, the execution of a simulation run, i.e. the logical process that transforms any (potentially stochastic) given input into its corresponding (potentially stochastic) output is pure deduction (i.e. strict application of the formal rules that define the model). Thus, running the model in a computer provides a formal proof that a particular input (together with the set of rules that define the model) is sufficient to generate the output that is observed during the simulation. This first part of the computer simulation approach is therefore, in a way, very "mathematical": outputs obtained follow with logical necessity from applying to the inputs the algorithmic rules that define the model.

In contrast, the second part of the computer simulation approach, i.e. inferring general patterns from particular instances of input-output pairs, can only lead to probable – rather than necessarily true – conclusions.[8] The following section explains how to rigorously assess the confidence we can place on the conclusions obtained using computer simulation, but the simple truth is irrefutable: inferences obtained using generalisation by induction can potentially fail when applied to instances that were not used to infer the general pattern. This is the domain of statistical extrapolation.

[8] Unless, of course, all possible particular instances are explored.

So why bother with computer simulation at all? The answer is clear: computer simulation enables us to study formal systems in ways that go beyond mathematical tractability. This role should not be underestimated: most models in the Social Simulation literature are mathematically intractable, and in such cases computer simulation is our only chance to move things forward. As a matter of fact, the formal models that many computer programs implement are often so complicated and cumbersome that the computer code itself is not that far from being one of the best descriptions of the formal model that can be provided.

Computer simulation can be very useful even when dealing with formal models that are mathematically tractable. Valuable uses of computer simulation in these cases include conducting insightful initial explorations of the model and presenting dynamic illustrations of its results.

And there is yet another important use of computer simulation. Note that understanding a formal model in depth requires identifying the parts of the model (i.e. the subset of rules) that are responsible for generating particular (sub)sets of results or properties of results. Investigating this in detail often involves changing certain subsets of rules in the model, so one can pinpoint which subsets of rules are necessary or sufficient to produce certain results. Importantly, changing subsets of rules can make the original model mathematically intractable and in such (common) cases, computer simulation is, again, our only hope. In this context, computer simulation can be very useful to produce counter-examples. This approach is very common in the literature of e.g. evolutionary game theory, where several authors (see e.g. Hauert and Doebeli 2004; Imhof et al. 2005; Izquierdo and Izquierdo 2006; Lieberman et al. 2009; Nowak and May 1992; Nowak and Sigmund 1992, 1993; Santos et al. 2006; Traulsen et al. 2006) resort to computer simulations to assess the implications of assumptions made in mathematically tractable models (e.g. the assumptions of "infinite populations" and "random encounters").

It is important to note that the fundamental distinction between mathematical analysis and computer simulation as presented here is *not* about whether one uses pen and paper or computers to analyse formal models. We can follow either approach with or without computers, and it is increasingly popular to do mathematical analysis with computers. Recent advancements in symbolic computation have opened up a new world of possibilities to conduct mathematical analyses (using e.g. Mathematica©). In other words, nowadays it is perfectly possible to use computers to *directly* examine the rules that define a formal model (see Fig. 11.3).

Finally, as so often in life, things are not black or white, but involve some shade of grey. Similarly, most models are not tractable *or* intractable in mathematical terms; most often they are *partially* tractable. It is in these cases where an adequate combination of mathematical analysis and computer simulation is particularly useful. We illustrate this fact in Sect. 11.8, but first let us look at each technique separately. The following two sections provide some guidelines on how computer simulation (Sect. 11.6) and mathematical analysis (Sect. 11.7) can be usefully employed to analyse formal models.

## 11.6 Computer Simulation: Approximating the Exact Probability Distribution by Running the Model

The previous sections have argued that *any* statistic obtained from a (stochastically or deterministically) parameterised model follows a specific probability distribution. The statistic could be anything as long as it is unambiguously defined; in particular, it could refer to one or several time-steps, and to one or various subcomponents of the model. Ideally, one would like to calculate the exact probability distribution for the statistic using mathematical analysis, but this will not always be possible. In contrast, using computer simulation we will always be able to approximate this probability distribution to any arbitrary level of accuracy; this section provides basic guidelines on how to do that.

The output probability distribution – which is fully and unequivocally determined by the input distribution – can be approximated to any degree of accuracy by running enough simulation runs. Note that any specific simulation run will be conducted with a particular certain value for every parameter (e.g. a particular initial location for every agent), and will produce one and only one particular certain output (see Fig. 11.3). Thus, in order to infer the probability distribution over the set of outputs that a particular probability distribution over the set of inputs leads to, there will be a need to run the model many times (with different random seeds); this is the so-called Monte Carlo method.

The method is straightforward: obtain as many random samples as possible (i.e. run as many independent simulations as possible), since this will get us closer and closer to the exact distribution (by the law of large numbers). Having conducted a large number of simulation runs, the question that naturally comes to mind is: How close to the exact distribution is the one obtained by simulation?

To illustrate how to assess the quality of the approximation obtained by simulation, we use CoolWorld, a purpose-built agent-based model (Gilbert 2007) implemented in NetLogo 4.0 (Wilensky 1999). A full description of the model, an applet and the source code can be found at the dedicated model webpage http://luis.izquierdo.name/models/coolworld. For our purposes, it suffices to say that in CoolWorld there is a population of agents called walkers, who wander around a 2-dimensional grid made of square patches; some of the patches are empty whilst others contain a house (see Fig. 11.4). Patches are at a certain predefined temperature, and walkers tend to walk towards warmer patches, staying for a while at the houses they encounter in their journey.

Let us assume that we are interested in studying the number of CoolWorld walkers staying in a house in time-step 50. Initial conditions (which involve 100 walkers placed at a random location) are unambiguously defined at the model webpage and can be set in the implementation of CoolWorld provided by clicking on the button "Special conditions". Figure 11.4 shows a snapshot of CoolWorld after having clicked on that button.

As argued before, given that the (stochastic) initial conditions are unambiguously defined, the number of CoolWorld walkers in a house after 50 time-steps will follow

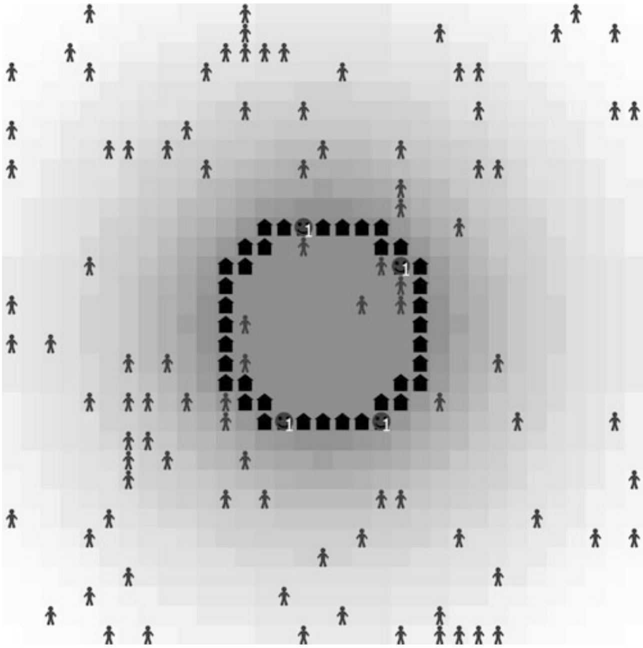



**Fig. 11.4** Snapshot of CoolWorld. Patches are coloured according to their temperature: the higher the temperature, the darker the shade of *grey*. Houses are coloured in *black*, and form a circle around the central patch. Walkers are coloured in *grey*, and represented as a person if standing on a patch without a house, and as a *smiling face* if standing on a patch with a house. In the latter case, the *white label* indicates the number of walkers in the same house

a specific probability distribution that we are aiming to approximate. For that, let us assume that we run 200 runs, and plot the relative frequency of the number of walkers in a patch with a house after 50 time-steps (see Fig. 11.5).

Figure 11.5 does not provide all the information that can be extracted from the data gathered. In particular, we can plot error bars showing the standard error for each calculated frequency without hardly any effort.[9] Standard errors give us information about the error we may be incurring when estimating the exact probabilities with the empirical frequencies. Another simple task that can be conducted consists in partitioning the set of runs into two batteries of approximately

---

[9] The frequency of the event "there are $i$ walkers in a patch with a house" calculated over $n$ simulation runs can be seen as the mean of a sample of $n$ i.i.d. Bernouilli random variables where success denotes that the event occurred and failure denotes that it did not. Thus, the frequency $f$ is the maximum likelihood (unbiased) estimator of the exact probability with which the event occurs. The standard error of the calculated frequency $f$ is the standard deviation of the sample divided by the square root of the sample size. In this particular case, the formula reads:

$$\text{Std. error } (f, n) = (f\,(1-f)/(n-1))^{1/2}$$

Where $f$ is the frequency of the event, $n$ is the number of samples, and the standard deviation of the sample has been calculated dividing by $(n - 1)$.

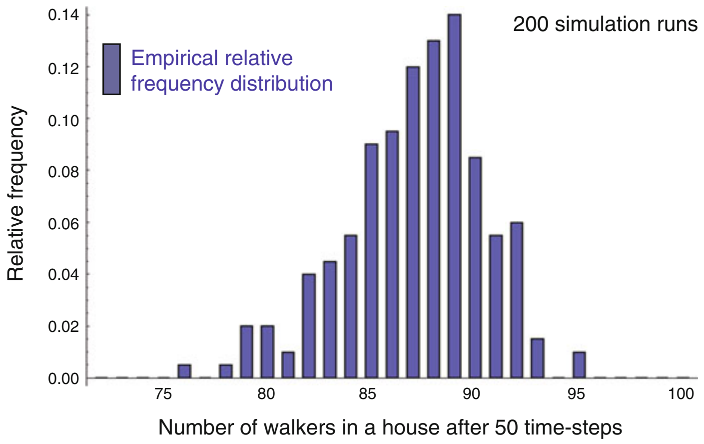



**Fig. 11.5** Relative frequency distribution of the number of walkers in a house after 50 time-steps, obtained by running CoolWorld 200 times, with initial conditions set by clicking on "Special conditions"

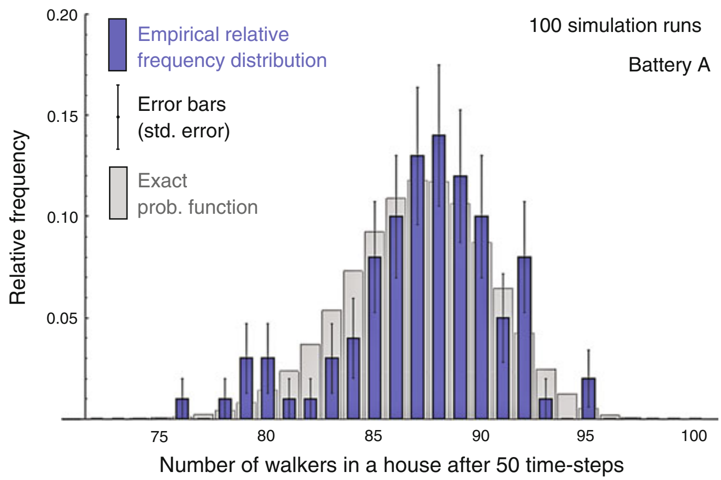



**Fig. 11.6** In *darker grey*: Relative frequency distribution of the number of walkers in a house after 50 time-steps, obtained by running CoolWorld 100 times (Battery A), with initial conditions set by clicking on "Special conditions". In *lighter grey*: Exact probability distribution (Calculated using Markov chain analysis)

equal size and comparing the two distributions. If the two distributions are not similar, then there is no point in proceeding: we are not close to the exact distribution, so there is a need to run more simulations.

Figures 11.6 and 11.7 show the data displayed in Fig. 11.5 partitioned in two batteries of 100 simulation runs, including the standard errors. Figures 11.6 and 11.7 also show the exact probability distribution we are trying to approximate, which has been calculated using mathematical methods that are explained later in this chapter.

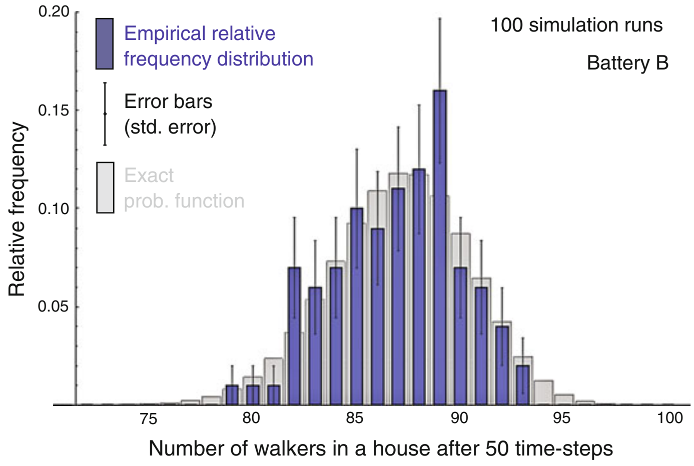



**Fig. 11.7** In *darker grey*: Relative frequency distribution of the number of walkers in a house after 50 time-steps, obtained by running CoolWorld 100 times (Battery B), with initial conditions set by clicking on "Special conditions". In *lighter grey*: Exact probability distribution (Calculated using Markov chain analysis)

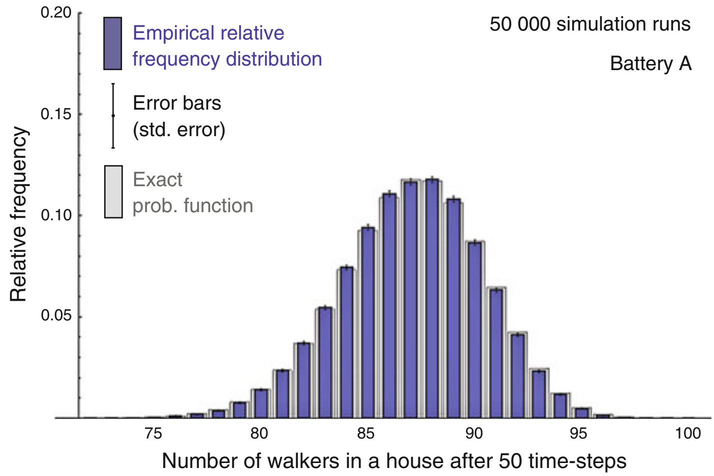



**Fig. 11.8** In *darker grey*: Relative frequency distribution of the number of walkers in a house after 50 time-steps, obtained by running CoolWorld 50,000 times (Battery A), with initial conditions set by clicking on "Special conditions". In *lighter grey*: Exact probability distribution (Calculated using Markov chain analysis)

Figures 11.6 and 11.7 indicate that 100 simulation runs may not be enough to obtain a satisfactory approximation to the exact probability distribution. On the other hand, Figs. 11.8 and 11.9 show that running the model 50,000 times does seem to get us close to the exact probability distribution. The standard error, which

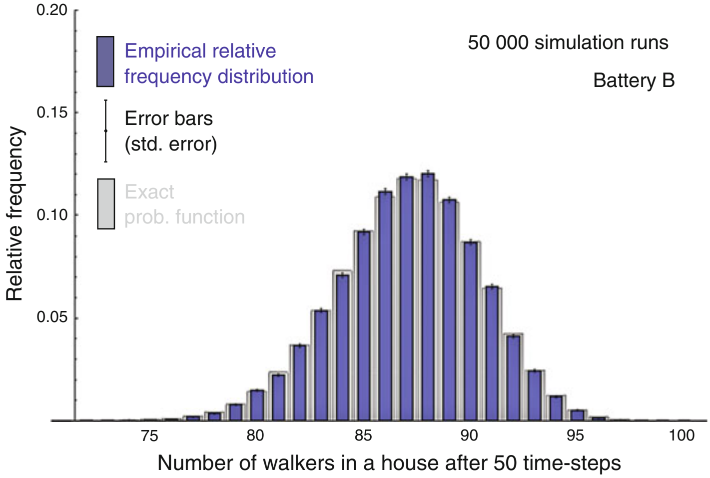



**Fig. 11.9** In *darker grey*: Relative frequency distribution of the number of walkers in a house after 50 time-steps, obtained by running CoolWorld 50,000 times (Battery B), with initial conditions set by clicking on "Special conditions". In *lighter grey*: Exact probability distribution (Calculated using Markov chain analysis)

is inversely proportional to the square root of the sample size (i.e. the number of runs), is naturally much lower in these latter cases.

When, like in this example, the space of all possible outcomes in the distribution under analysis is finite (the number of walkers in a house must be an integer between 0 and 100), one can go further and calculate confidence intervals for the obtained frequencies. This is easily conducted when one realises that the exact probability distribution is a multinomial. Genz and Kwong (2000) show how to calculate these confidence intervals.

To conclude this section, let us emphasise that all that has been written here applies to *any* statistic obtained from *any* computer model. In particular, the statistic may refer to predefined regimes (e.g. "number of time-steps between 0 and 100 where there are more than 20 walkers in a house") or to various time-steps (e.g. "total number of walkers in a house in odd time-steps in between time-steps 50 and 200"). These statistics, like any other one, follow a specific probability distribution that can be approximated to any degree of accuracy by running the computer model.

## 11.7 Mathematical Analysis: Time-Homogenous Markov Chains

The whole range of mathematical techniques that can be used to analyse formal systems is too broad to be reviewed here. Instead, we focus on one specific technique that seems to us particularly useful to analyse Social Simulation models:

Markov chain analysis. Besides, there are multiple synergies to be exploited by using Markov chain analysis and computer simulation together, as we will see in the next section.

Our first objective is to learn how to represent a particular computer model as a time-homogeneous Markov chain. This alternative representation of the model will allow us to use several simple mathematical results that will prove useful to understand the dynamics of the model. We therefore start by describing time-homogeneous Markov chains.

### 11.7.1 What Is a Time-Homogeneous Markov Chain?

Consider a system that in time-step $n = \{1,2,3,\ldots\}$ may be in one of a finite number of possible states $S = \{s_1, s_2,\ldots, s_M\}$. The set $S$ is called the state space; in this chapter we only consider *finite* state spaces.[10] Let the sequence of random variables $X_n \in S$ represent the state of the system in time-step $n$. As an example, $X_3 = s_9$ means that at time $n = 3$ the system is in state $s_9$. The system starts at a certain initial state $X_0$ and moves from one state to another. The system is stochastic in that, given the present state, the system may move to one or another state with a certain probability (see Fig. 11.10). The probability that the system moves from state $i$ to state $j$ in one time-step, $\mathrm{P}(X_{n+1} = j \mid X_n = i)$, is denoted by $p_{i,j}$. As an example, in the Markov chain represented in Fig. 11.10, $p_{4,6}$ equals 0 since the system cannot go from state 4 to state 6 in one single time-step. The system may also stay in the same state $i$, and this occurs with probability $p_{i,i}$. The probabilities $p_{i,j}$ are called transition probabilities and they are often arranged in a matrix, namely the transition matrix $P$.

Implicitly, our definition of transition probabilities assumes two important properties about the system:

(a) The system has the **Markov property**. This means that the present state contains all the information about the future evolution of the system that can be obtained from its past, i.e. given the present state of the system, knowing the past history about how the system reached the present state does not provide any additional information about the future evolution of the system. Formally,

$$\mathrm{P}(X_{n+1} = x_{n+1} \mid X_n = x_n,\ X_{n-1} = x_{n-1}, \ldots,\ X_0 = x_0) = \mathrm{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

(b) In this chapter we focus on **time-homogeneous** Markov chains, i.e. Markov chains with time-homogeneous transition probabilities. This basically means that transition probabilities $p_{i,j}$ are independent of time, i.e. the one-step transition probability $p_{i,j}$ depends on $i$ and $j$ but is the same at all times $n$. Formally,

[10] The term 'Markov chain' allows for countably infinite state spaces too (Karr 1990).

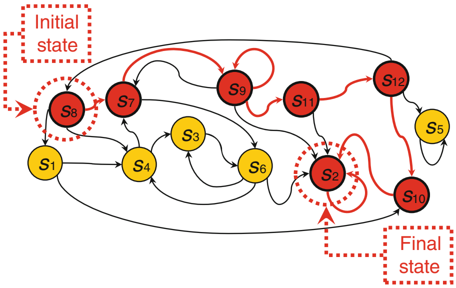



**Fig. 11.10** Schematic transition diagram of a Markov chain. *Circles* denote states and *directed arrows* indicate possible transitions between states. In this figure, thicker *circles* and *arrows* represent one possible path where the initial state $X_0$ is $s_8$ and the final state is $s_2$

$$\mathrm{P}(X_{n+1} = j \,|\, X_n = i) = \mathrm{P}(X_n = j \,|\, X_{n-1} = i) = p_{i,j}$$

The crucial step in the process of representing a computer model as a time-homogeneous Markov chain (THMC) consists in identifying an appropriate set of state variables. A particular combination of specific values for these state variables will define one particular state of the system. Thus, the challenge consists in choosing the set of state variables in such a way that the computer model can be represented as a THMC. In other words, the set of state variables must be such that one can see the computer model as a transition matrix that unambiguously determines the probability of going from any state to any other state.

**Example: A Simple Random Walk**

Let us consider a model of a simple 1-dimensional random walk and try to see it as a THMC. In this model – which can be run and downloaded at the dedicated model webpage http://luis.izquierdo.name/models/randomwalk – there are 17 patches in line, labelled with the integers between 1 and 17. A random walker is initially placed on one of the patches. From then onwards, the random walker will move randomly to one of the spatially contiguous patches in every time-step (staying still is not an option). Space does not wrap around, i.e. patch 1's only neighbour is patch 2 (Fig. 11.11).

This model can be easily represented as a THMC by choosing the agent's position (e.g. the number of the patch she is standing on) as the only state variable. To be sure, note that defining the state of the system in this way, it is true that there is a fixed probability of going from any state to any other state, independent of time. The transition matrix $P = [p_{i,j}]$ corresponding to the model is:

$$P = [p_{i,j}] = \begin{bmatrix} 0 & 1 & 0 & & & & \cdots & 0 \\ 0.5 & 0 & 0.5 & 0 & & & & \vdots \\ 0 & 0.5 & 0 & 0.5 & 0 & & & \\ & 0 & 0.5 & 0 & 0.5 & 0 & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & 0 & 0.5 & 0 & 0.5 & 0 \\ \vdots & & & & 0 & 0.5 & 0 & 0.5 \\ 0 & \cdots & & & & 0 & 1 & 0 \end{bmatrix} \qquad (11.1)$$

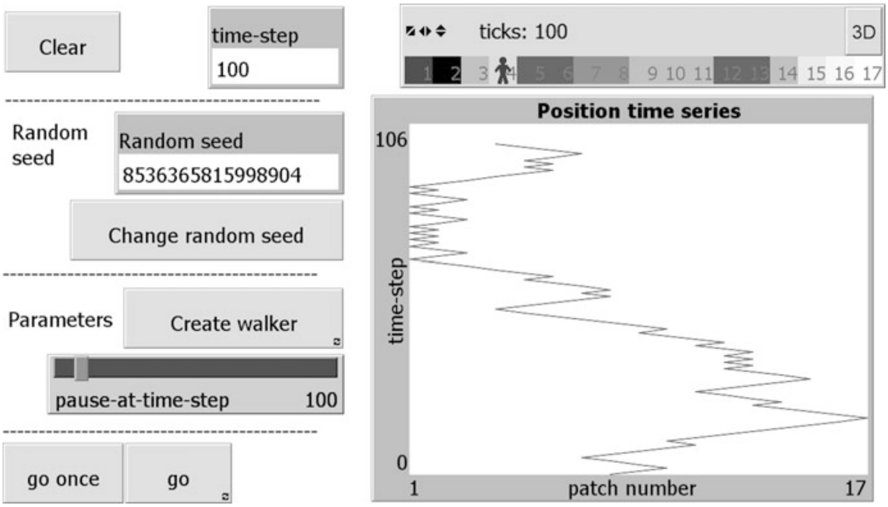



**Fig. 11.11** Snapshot of the 1-dimensional random walk applet. Patches are arranged in a horizontal line on the *top right corner* of the figure; they are labelled with integers, and coloured in shades of grey according to the number of times that the random walker has visited them: the higher the number of visits, the darker the shade of blue. The plot beneath the patches shows the time series of the random walker's position

Where, as explained above, $p_{i,j}$ is the probability $P(X_{n+1} = j \mid X_n = i)$ that the system will be in state $j$ in the following time-step, knowing that it is currently in state $i$.

## *11.7.2 Transient Distributions of Finite THMCs*

The analysis of the dynamics of THMCs is usually divided into two parts: transient dynamics (finite time) and asymptotic dynamics (infinite time). The transient behaviour is characterised by the distribution of the state of the system $X_n$ for a fixed time-step $n \geq 0$. The asymptotic behaviour (see Sects. 11.7.3 and 11.7.4) is characterised by the limit of the distribution of $X_n$ as $n$ goes to infinity, when this limit exists.

This section explains how to calculate the transient distribution of a certain THMC, i.e. the distribution of $X_n$ for a fixed $n \geq 0$. In simple words, we are after a vector $a^{(n)}$ containing the probability of finding the process in each possible state in time-step $n$. Formally, $a^{(n)} = [a_1^{(n)}, \ldots, a_M^{(n)}]$, where $a_i^{(n)} = P(X_n = i)$, denotes the distribution of $X_n$ for a THMC with $M$ possible states. In particular, $a^{(0)}$ denotes the initial distribution over the state space, i.e. $a_i^{(0)} = P(X_0 = i)$. Note that there is no problem in having uncertain initial conditions, i.e. probability functions over the space of possible inputs to the model.

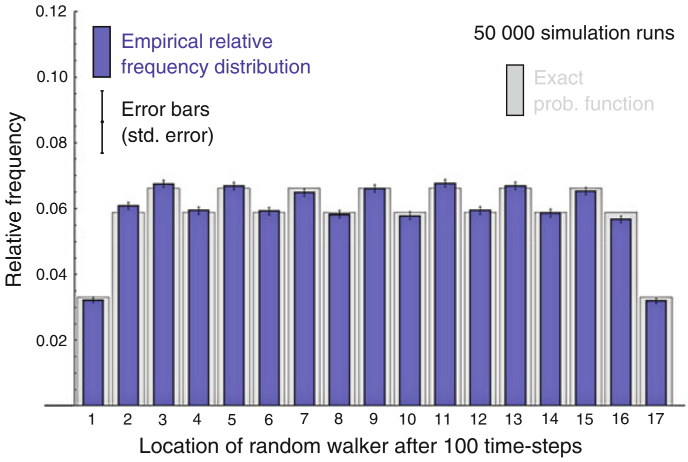



**Fig. 11.12** Probability function of the position of the 1-dimensional random walker in time-step 100, starting at an initial random location

It can be shown that one can easily calculate the transient distribution in time-step $n$, simply by multiplying the initial conditions by the $n$-th power of the transition matrix $P$.

**Proposition 1.** $a^{(n)} = a^{(0)} \cdot P^n$

Thus, the elements $p^{(n)}{}_{i,j}$ of $P^n$ represent the probability that the system is in state $j$ after $n$ time-steps having started in state $i$, i.e. $p^{(n)}{}_{i,j} = \mathrm{P}(X_n = j \mid X_0 = i)$. A straightforward corollary of Proposition 1 is that $a^{(n+m)} = a^{(n)} \cdot P^m$.

As an example, let us consider the 1-dimensional random walk again. Imagine that the random walker starts at an initial random location, i.e. $a^{(0)} = [1/17, \ldots, 1/17]$. The exact distribution of the walker's position in time-step 100 would then be $a^{(100)} = a^{(0)} \cdot P^{100}$. This distribution is represented in Fig. 11.10, together with an empirical distribution obtained by running the model 50,000 times (Fig. 11.12).

Having obtained the probability function over the states of the system for any fixed $n$, namely the probability mass function of $X_n$, it is then straightforward to calculate the distribution of *any* statistic that can be extracted from the model. As argued in the previous sections, the state of the system fully characterises it, so *any* statistic that we obtain about the computer model in time-step $n$ must be, ultimately, a function of $\{X_0, X_1, \ldots, X_n\}$.

Admittedly, the transition matrix of most computer models cannot be easily derived, or it is unfeasible to operate with it. Nonetheless, this apparent drawback is not as important as one might expect. As we shall see below, it is often possible to infer many properties of a THMC even without knowing the exact values of its transition matrix, and these properties can yield useful insights about the dynamics of the associated process. Knowing the exact values of the transition matrix allows us to calculate the exact transient distributions using Proposition 1; this is desirable but not critical, since we can always approximate these distributions by conducting many simulation runs, as explained in Sect. 11.6.

### 11.7.3 Important Concepts

This section presents some basic concepts that will prove useful to analyse the dynamics of computer models. The notation used here follows the excellent book on stochastic processes written by Kulkarni (1995).

**Definition 1: Accessibility**
A state $j$ is said to be accessible from state $i$ if starting at state $i$ there is a chance that the system may visit state $j$ at some point in the future. By convention, every state is accessible from itself. Formally, a state $j$ is said to be accessible from state $i$ if for some $n \geq 0$, $p^{(n)}{}_{i,j} > 0$.

Note that $j$ is accessible from $i \neq j$ if and only if there is a directed path from $i$ to $j$ in the transition diagram. In that case, we write $i \rightarrow j$. If $i \rightarrow j$ we also say that $i$ leads to $j$. As an example, in the THMC represented in Fig. 11.10, $s_2$ is accessible from $s_{12}$ but not from $s_5$. Note that the definition of accessibility does not depend on the actual magnitude of $p^{(n)}{}_{i,j}$, only on whether it is exactly zero or strictly positive.

**Definition 2: Communication**
A state $i$ is said to communicate with state $j$ if $i \rightarrow j$ and $j \rightarrow i$.

If $i$ communicates with $j$ we also say that $i$ and $j$ communicate and write $i \leftrightarrow j$. As an example, note that in the simple random walk presented in Sect. 11.7.1, every state communicates with every other state. It is worth noting that the relation "communication" is transitive, i.e.

$$i \leftrightarrow j,\ j \leftrightarrow k \Rightarrow\ i \leftrightarrow k.$$

**Definition 3: Communicating Class**
A set of states $C \subset S$ is said to be a communicating class if:

- Any two states in the communicating class communicate with each other. Formally,

$$i \in C,\ j \in C \Rightarrow\ i \leftrightarrow j$$

- The set $C$ is maximal, i.e. no strict superset of a communicating class can be a communicating class. Formally,

$$i \in C,\ \ i \leftrightarrow j\ \ \Rightarrow j \in C$$

As an example, note that in the simple random walk presented in Sect. 11.7.1 there is one single communicating class that contains all the states. In the THMC represented in Fig. 11.10 there are four communicating classes: $\{s_2\}$, $\{s_5\}$, $\{s_{10}\}$, $\{s_1, s_3, s_4, s_6, s_7, s_8, s_9, s_{11}, s_{12}\}$.

**Definition 4: Closed Communicating Class (i.e. Absorbing Class). Absorbing State**
A communicating class $C$ is said to be closed if no state within $C$ leads to any state outside $C$. Formally, a communicating class $C$ is said to be closed if $i \in C$ and $j \notin C$ implies that $j$ is not accessible from $i$.

Note that once a Markov chain visits a closed communicating class, it cannot leave it. Hence we will sometimes refer to closed communicating classes as "absorbing classes". This latter term is not standard in the literature, but we find it useful here for explanatory purposes. Note that if a Markov chain has one single communicating class, it must be closed.

As an example, note that the communicating classes $\{s_{10}\}$ and $\{s_1, s_3, s_4, s_6, s_7, s_8, s_9, s_{11}, s_{12}\}$ in the THMC represented in Fig. 11.10 are not closed, as they can be abandoned. On the other hand, the communicating classes $\{s_2\}$ and $\{s_5\}$ are indeed closed, since they cannot be abandoned. When a closed communicating class consists of one single state, this state is called absorbing. Thus, $s_2$ and $s_5$ are absorbing states. Formally, state $i$ is absorbing if and only if $p_{i,i} = 1$ and $p_{i,j} = 0$ for $i \neq j$.

**Proposition 2. Decomposition Theorem (Chung 1960)**
The state space $S$ of any Markov chain can be uniquely partitioned as follows:

$$S = C_1 \cup C_2 \cup \ldots \cup C_k \cup T$$

where $C_1, C_2, \ldots, C_k$ are closed communicating classes, and $T$ is the union of all other communicating classes.

Note that we do not distinguish between non-closed communicating classes: we lump them all together into $T$. Thus, the unique partition of the THMC represented in Fig. 11.10 is $S = \{s_2\} \cup \{s_5\} \cup \{s_1, s_3, s_4, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}\}$. The simple random walk model presented in Sect. 11.7.1 has one single (closed) communicating class $C_1$ containing all the possible states, i.e. $S \equiv C_1$.

**Definition 5: Irreducibility**
A Markov chain is said to be irreducible if all its states belong to a single closed communicating class; otherwise it is called reducible. Thus, the simple random walk presented in Sect. 11.7.1 is irreducible, but the THMC represented in Fig. 11.10 is reducible.

**Definition 6: Transient and Recurrent States**
A state $i$ is said to be transient if, given that we start in state $i$, there is a non-zero probability that we will never return back to $i$. Otherwise, the state is called recurrent. A Markov chain starting from a recurrent state will revisit it with probability 1, and hence revisit it infinitely often. On the other hand, a Markov chain starting from a transient state has a strictly positive probability of never coming back to it. Thus, a Markov chain will visit any transient state only finitely many times; eventually, transient states will not be revisited anymore.

**Definition 7: Periodic and Aperiodic States. Periodic and Aperiodic Communicating Classes**
A state $i$ has period $d$ if any return to state $i$ must occur in multiples of $d$ time-steps. If $d = 1$, then the state is said to be aperiodic; otherwise ($d > 1$), the state is said to be periodic with period $d$. Formally, state $i$'s period $d$ is the greatest common divisor of the set of integers $n > 0$ such that $p^{(n)}{}_{i,i} > 0$. For our purposes, the

concept of periodicity is only relevant for recurrent states. As an example, note that every state in the simple random walk presented in Sect. 11.7.1 is periodic with period 2.

An interesting and useful fact is that if $i \leftrightarrow j$, then $i$ and $j$ must have the same period (see Theorem 5.2. in Kulkarni (1995)). In particular, note that if $p_{i,i} > 0$ for any $i$, then the communicating class to which $i$ belongs must be aperiodic. Thus, it makes sense to qualify communicating classes as periodic with period $d$, or aperiodic. A closed communicating class with period $d$ can return to its starting state only at times $d$, $2d$, $3d$, …

The concepts presented in this section will allow us to analyse the dynamics of any finite Markov chain. In particular, we will show that, given enough time, any finite Markov chain will necessarily end up in one of its closed communicating classes (i.e. absorbing classes).

## *11.7.4 Limiting Behaviour of Finite THMCs*

This section is devoted to characterising the limiting behaviour of a THMC, i.e. studying the convergence (in distribution) of $X_n$ as $n$ tends to infinity. Specifically, we aim to study the behaviour of $a_i^{(n)} = \mathrm{P}(X_n = i)$ as $n$ tends to infinity. From Proposition 1 it is clear that analysing the limiting behaviour of $P^n$ would enable us to characterise $a_i^{(n)}$. There are many introductory books in stochastic processes that offer clear and simple methods to analyse the limiting behaviour of THMCs when the transition matrix $P$ is tractable (see e.g. Chap. 5 in (Kulkarni 1999), Chaps. 2–4 in (Kulkarni 1995), Chap. 3 in (Janssen and Manca 2006) or the book chapter written by Karr (1990)). Nonetheless, we focus here on the general case, where operating with the transition matrix $P$ may be computationally unfeasible.

### 11.7.4.1 General Dynamics

The first step in the analysis of any THMC consists in identifying all the closed communicating classes, so we can partition the state space $S$ as indicated by the decomposition theorem (see proposition 2). The following proposition (Theorems 3.7 and 3.8 in Kulkarni (1995)) reveals the significance of this partition:

**Proposition 3. General Dynamics of Finite THMCs**
Consider a finite THMC that has been partitioned as indicated in proposition 2. Then:

1. All states in $T$ (i.e. not belonging to a closed communicating class) are transient.
2. All states in $C_v$ (i.e. in any closed communicating class) are recurrent; $v \in \{1, 2, \ldots, k\}$.

Proposition 3 states that sooner or later the THMC will enter one of the absorbing classes and stay in it forever. Formally, for all $i \in S$ and all $j \in T$:

$\lim_{n\to\infty} p_{i,j}^{(n)} = 0$, i.e. the probability of finding the process in a state belonging to a non-closed communicating class goes to zero as $n$ goes to infinity. Naturally, if the initial state already belongs to an absorbing class $C_v$, then the chain will never abandon such a class. Formally, for all $i \in C_v$ and all $j \notin C_v$: $p^{(n)}{}_{i,j} = 0$ for all $n \geq 0$.

As an example of the usefulness of Proposition 3, consider the THMC represented in Fig. 11.10. This THMC has only two absorbing classes: $\{s_2\}$ and $\{s_5\}$. Thus, the partition of the state space is: $S = \{s_2\} \cup \{s_5\} \cup \{s_1, s_3, s_4, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}\}$. Hence, applying Proposition 3 we can state that the process will eventually end up in one of the two absorbing states, $s_2$ or $s_5$. The probability of ending up in one or the other absorbing state depends on the initial conditions $a^{(0)}$ (and on the actual numbers $p_{i,j}$ in the transition matrix, of course). Slightly more formally, the limiting distribution of $X_n$ exists, but it is not unique, i.e. it depends on the initial conditions.

#### 11.7.4.2 Dynamics Within Absorbing Classes

The previous section has explained that any simulation run will necessarily end up in a certain absorbing class; this section characterises the dynamics of a THMC that is already “trapped” in an absorbing class. This is precisely the analysis of irreducible Markov chains, since irreducible Markov chains are, by definition, Markov chains with one single closed communicating class (see definition 5). In other words, one can see any THMC as a set of transient states $T$ plus a finite number of irreducible Markov sub-chains.

Irreducible THMCs behave significantly different depending on whether they are periodic or not. The following sections characterise these two cases.

##### Irreducible and Aperiodic THMCs

Irreducible and aperiodic THMCs are often called ergodic. In these processes the probability function of $X_n$ approaches a limit as $n$ tends to infinity. This limit is called the **limiting distribution**, and is denoted here by π. Formally, the following limit exists and is unique (i.e. independent of the initial conditions $a_i^{(0)}$):

$$\lim_{n\to\infty} a_i^{(n)} = \pi_i > 0 \qquad i \in S$$

Thus, in ergodic THMCs the probability of finding the system in each of its states in the long run is strictly positive and independent of the initial conditions (Theorems 3.7 and 3.15 in Kulkarni (1995)). As previously mentioned, calculating such probabilities may be unfeasible, but we can estimate them sampling many simulation runs at a sufficiently large time-step.

Importantly, in ergodic THMCs the limiting distribution π coincides with the **occupancy distribution** π*, which is the long-run fraction of the time that the THMC spends in each state.[11] Naturally, the occupancy distribution π* is also independent of the initial conditions. Thus, in ergodic THMCs, running just one simulation for long enough (which enables us to estimate π*) will serve to estimate π just as well.

The question that comes to mind then is: How long is long enough? I.e. when will I know that the empirical distribution obtained by simulation resembles the limiting distribution π? Unfortunately there is no answer for that. The silver lining is that knowing that the limiting and the occupancy distribution coincide, that they must be stable in time, and that they are independent of the initial conditions, enables us to conduct a wide range of tests that may tell us when it is certainly *not* long enough. For example, we can run a battery of simulations and study the empirical distribution over the states of the system across samples as time goes by. If the distribution is not stable, then we have not run the model for long enough. Similarly, since the occupancy distribution is independent of the initial conditions, one can run several simulations with widely different initial conditions, and compare the obtained occupancy distributions. If the empirical occupancy distributions are not similar, then we have not run the model for long enough. Many more checks can be conducted.

Admittedly, when analysing a computer model one is often interested not so much in the distribution over the possible states of the system, but rather in the distribution of a certain statistic. The crucial point is to realise that if the statistic is a function of the state of the system (and all statistics that can be extracted from the model are), then the limiting and the occupancy distributions of the statistic exist, coincide and are independent of the initial conditions.

## Irreducible and Periodic THMCs

In contrast with aperiodic THMCs, the probability distribution of $X_n$ in periodic THMCs does not approach a limit as $n$ tends to infinity. Instead, in an irreducible THMC with period $d$, as $n$ tends to infinity, $X_n$ will in general cycle through $d$ probability functions depending on the initial distribution.

As an example, consider the simple random walk again (which is irreducible and periodic, with period 2), and assume that the random walker starts at patch number 1 (i.e. $X_0 = 1$). Given these settings, it can be shown that

[11] Formally, the occupancy of state $i$ is defined as:

$$\pi_i^* = \lim_{n\to\infty} \frac{E(N_i(n))}{n+1}$$

where $N_i(n)$ denotes the number of times that the THMC visits state $i$ over the time span $\{0, 1,\ldots, n\}$.

$$\lim_{n\to\infty} a_i^{(2n)} = \left[\frac{1}{16}, 0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{8} 0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{16}\right]$$

$$\lim_{n\to\infty} a_i^{(2n+1)} = \left[0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{8}, 0, \frac{1}{8}, 0\right]$$

In particular, the limits above show that the random walker cannot be at a patch with an even number in any even time-step, and he cannot be at a patch with an odd number in any odd time-step. In contrast, if the random walker started at patch number 2 (i.e. $X_0 = 2$), then the limits above would be interchanged.

Fortunately, every irreducible (periodic or aperiodic) THMC does have a unique occupancy distribution π*, independent of the initial conditions (see Theorem 5.19 in Kulkarni (1999)). In our particular example, this is:

$$\pi* = \left[\frac{1}{32}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{32}\right]$$

Thus, the long-run fraction of time that the system spends in each state in any irreducible THMC is unique (i.e. independent of the initial conditions). This is a very useful result, since any statistic which is a function of the state of the system will also have a unique occupancy distribution independent of the initial conditions. As explained before, this occupancy distribution can be approximated with one single simulation run, assuming it runs for long enough.

## 11.8 Synergies Between Mathematical Analysis and Computer Simulation

In this section we present various ways in which mathematical analysis and computer simulation can be combined to produce a better understanding of the dynamics of a model. Note that a full understanding of the dynamics of a model involves not only characterising (i.e. describing) them, but also finding out *why* such dynamics are being observed, i.e. identifying the subsets of rules that are necessary or sufficient to generate certain aspects of the observed dynamics. To do this, one often has to make changes in the model, i.e. build supporting models that differ only slightly from the original one and may yield useful insights about its dynamics. These supporting models will sometimes be more tractable (e.g. if heterogeneity or stochasticity are averaged out) and sometimes more complex (e.g. if interactions that were assumed to be global in the original model may only take place locally in the supporting model). Thus, for clarity, we distinguish three different cases and deal with them in turn (see Fig. 11.13):

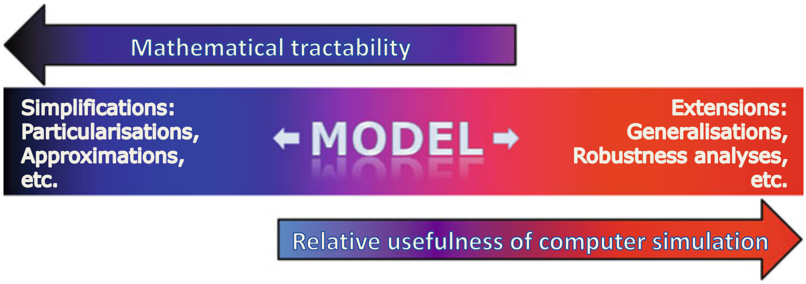



**Fig. 11.13** To fully understand the dynamics of a model, one often has to study supporting models that differ only slightly from the original one. Some of these supporting models may be more tractable whilst others may be more complex

1. Characterisation of the dynamics of a model.
2. Moves towards greater mathematical tractability. This involves creating and studying supporting models that are simpler than the original one.
3. Moves towards greater mathematical complexity. This involves creating and studying supporting models that are less tractable than the original one.

### *11.8.1 Characterising the Dynamics of the Model*

There are many types of mathematical techniques that can be usefully combined with computer simulation to characterise the dynamics of a model (e.g. Stochastic Approximation Theory (Benveniste et al. 1990; Kushner and Yin 1997)), but for limitations of space we focus here on Markov chain analysis only.

When using Markov chain analysis to characterise the dynamics of a model it may happen that the transition matrix can be easily computed and we can operate with it, or it may not. In the former case – which is quite rare in Social Simulation models –, one can provide a full characterisation of the dynamics of the model just by operating with the transition matrix (see Proposition 1 and the beginning of Sect. 11.7.4 for references). In general, however, deriving and operating with the transition matrix may be unfeasible, and it is in this common case where there is a lot to gain in using Markov chain analysis and computer simulation together. The overall method goes as follows:

- Use Markov chain analysis to assess the relevance of initial conditions and to identify the different regimes in which the dynamics of the model may end up trapped.
- Use the knowledge acquired in the previous point to design suitable computational experiments aimed at estimating the exact probability distributions for the relevant statistics (which potentially depend on the initial conditions).

The following describes this overall process in greater detail. Naturally, the first step consists in finding an appropriate definition of the state of the system, as explained in Sect. 11.7.1. The next step is to identify all the closed communicating (i.e. absorbing) classes in the model $C_v$ ($v \in \{1, 2, \ldots, k\}$). This allows us to partition the state space of the Markov chain as the union of all the closed communicating classes $C_1, C_2, \ldots, C_k$ in the model plus another class $T$ containing all the states that belong to non-closed communicating classes. Izquierdo et al. (2009) illustrate how to do this in ten well-known models in the Social Simulation literature.

In most cases, conducting the partition of the state space is not as difficult as it may seem at first. In particular, the following proposition provides some simple sufficient conditions that guarantee that the computer model contains one single aperiodic absorbing class, i.e. the finite THMC that the computer model implements is irreducible and aperiodic (i.e. ergodic).

**Proposition 4. Sufficient Conditions for Irreducibility and Aperiodicity**

1. If it is possible to go from any state to any other state in one single time-step ($p_{i,j} > 0$ for all $i \neq j$) and there are more than two states, then the THMC is irreducible and aperiodic.
2. If it is possible to go from any state to any other state in a finite number of time-steps ($i \leftrightarrow j$ for all $i \neq j$), and there is at least one state in which the system may stay for two consecutive time-steps ($p_{i,i} > 0$ for some $i$), then the THMC is irreducible and aperiodic.
3. If there exists a positive integer $n$ such that ${p^{(n)}}_{i,j} > 0$ for all $i$ and $j$, then the THMC is irreducible and aperiodic (Janssen and Manca 2006, p. 107).

If one sees the transition diagram of the Markov chain as a (directed) network, the conditions above can be rewritten as:

1. The network contains more than two nodes and there is a directed link from every node to every other node.
2. The network is strongly connected and there is at least one loop.
3. There exists a positive integer $n$ such that there is at least one walk of length $n$ from any node to every node (including itself).

Izquierdo et al. (2009) show that many models in the Social Simulation literature satisfy one of these sufficient conditions (e.g. Epstein and Axtell's (1996) Sugarscape, Axelrod's (1986) metanorms models, Takahashi's (2000) model of generalized exchange, and Miller and Page's (2004) standing ovation model with noise). This is important since, as explained in Sect. 11.7.4.2, in ergodic THMCs the limiting and the occupancy distributions of any statistic exist, coincide and are independent of the initial conditions (so running just one simulation for long enough, which enables us to estimate the occupancy distribution, will serve to estimate the limiting distribution just as well).

Let us return to the general case. Having partitioned the state space, the analysis of the dynamics of the model is straightforward: all states in $T$ (i.e. in any finite communicating class that is not closed) are transient, whereas all states in $C_v$ (i.e. in any finite closed communicating class) are recurrent. In other words, sooner or later any simulation run will enter one of the absorbing classes $C_v$ and stay in it forever.

Here computer simulation can play a crucial role again, since it allows us to estimate the probability of ending up in each of the absorbing classes for any (stochastic or deterministic) initial condition we may be interested in. A case-in-point would be a model that has only a few absorbing states, or where various absorbing states are put together into only a few groups. Izquierdo et al. (2009) analyse models that follow that pattern: Axelrod's (1997b) model of dissemination of culture, Arthur's (1989) model of competing technologies, and Axelrod and Bennett's (1993) model of competing bimodal coalitions. CharityWorld (Polhill et al. 2006; Izquierdo and Polhill 2006) is an example of a model with a unique absorbing state.

The following step consists in characterising the dynamics of the system within each of the absorbing classes. Once the system has entered a certain absorbing class $C_v$, it will remain in it forever exhibiting a unique conditional[12] occupancy distribution $\pi_v$* over the set of states that compose $C_v$. Naturally, the same applies to any statistic we may want to study, since all statistics that can be extracted from the model are a function of the state of the system.

The conditional occupancy distribution $\pi_v$* denotes the (strictly positive) long-run fraction of the time that the system spends in each state of $C_v$ given that the system has entered $C_v$. Importantly, the conditional occupancy distribution $\pi_v$* is the same regardless of the specific state through which the system entered $C_v$. The role of simulation here is to estimate these conditional occupancy distributions for the relevant statistics by running the model for long enough.

Finally, recall that some absorbing classes are periodic and some are aperiodic. Aperiodic absorbing classes have a unique conditional limiting distribution $\pi_v$ denoting the long-run (strictly positive) probability of finding the system in each of the states that compose $C_v$ given that the system has entered $C_v$. This conditional limiting distribution $\pi_v$ coincides with the conditional occupancy distribution $\pi_v$* and, naturally, is also independent of the specific state through which the system entered $C_v$. (Again, note that this also applies to the distribution of any statistic, as they are all functions of the state of the system, necessarily.)

In contrast with aperiodic absorbing classes, periodic absorbing classes do not generally have a unique limiting distribution; instead, they cycle through $d$ probability functions depending on the specific state through which the system entered $C_v$ (where $d$ denotes the period of the periodic absorbing class). This is knowledge that one must take into account at the time of estimating the relevant probability distributions using computer simulation.

Thus, it is clear that Markov chain analysis and computer simulation greatly complement each other. Markov chain analysis provides the overall picture of the dynamics of the model by categorising its different dynamic regimes and

[12] Given that the system has entered the absorbing class $C_v$.

identifying when and how initial conditions are relevant. Computer simulation uses this information to design appropriate computational experiments that allow us to quantify the probability distributions of the statistics we are interested in. As explained above, these probability distributions can always be approximated with any degree of accuracy by running the computer model several times.

There are several examples of this type of synergetic combination of Markov chain analysis and computer simulation in the literature. Galán and Izquierdo (2005) analysed Axelrod's (1986) agent-based model as a Markov chain, concluded that the long-run behaviour of that model was independent of the initial conditions, in contrast to the initial conclusions of the original analysis. Galán and and Izquierdo (2005) also used computer simulation to estimate various probability distributions. Ehrentreich (2002, 2006) used Markov chain analysis on the Artificial Stock Market (Arthur et al. 1997; LeBaron et al. 1999) to demonstrate that the mutation operator implemented in the model is not neutral to the learning rate, but introduces an upward bias.[13] A more positive example is provided by Izquierdo et al. (2007, 2008b), who used Markov chain analysis and computer simulation to confirm and advance various insights on reinforcement learning put forward by Macy and Flache (2002) and Flache and Macy (2002).

### *11.8.2 Moves Towards Greater Mathematical Tractability: Simplifications*

There are at least two types of simplifications that can help us to better understand the dynamics of a model. One consists in studying specific parameterisations of the original model that are thought to lead to particularly simple dynamics, or to more tractable situations (Gilbert and Terna 2000; Gilbert 2007). Examples of this type of activity would be to run simulations without agents or with very few agents, explore the behaviour of the model using extreme parameter values, model very simple environments, etc. This activity is common practice in the field (see e.g. Gotts et al. 2003c, d).

A second type of simplification consists in creating an abstraction of the original model (i.e. a model of the model) which is mathematically tractable. An example of one possible abstraction would be to study the *expected* motion of the dynamic system (see the studies conducted by Galán and Izquierdo (2005), Edwards et al. (2003), Castellano et al. (2000), Huet et al. (2007), Mabrouk et al. (2007), Vilà (2008) and Izquierdo et al. (2007, 2008b) for illustrations of mean-field approximations). Since these mathematical abstractions do not correspond in a one-to-one way with the specifications of the formal model, any results obtained with them will not be conclusive in general, but they may give us insights

[13] This finding does not refute some of the most important conclusions obtained by the authors of the original model.

suggesting areas of stability and basins of attraction, clarifying assumptions, assessing sensitivity to parameters, or simply giving the option to illustrate graphically the expected dynamics of the original model. This approach can also be used as a verification technique to detect potential errors and artefacts (Galán et al. 2009).

### 11.8.3 Moves Towards Greater Mathematical Complexity: Extensions

As argued before, understanding the dynamics of a model implies identifying the set of assumptions that are responsible for particular aspects of the obtained results. Naturally, to assess the relevance of any assumption in a model, it is useful to replace it with other alternatives, and this often leads to greater mathematical complexity.[14]

Ideally, the evaluation of the significance of an assumption is conducted by generalisation, i.e. by building a more general model that allows for a wide range of alternative competing assumptions, and contains the original assumption as a particular case. An example would be the introduction of arbitrary social networks of interaction in a model where every agent necessarily interacts with every other agent. In this case, the general model with arbitrary networks of interaction would correspond with the original model if the network is assumed to be complete, but any other network could also be studied within the same common framework. Another example is the introduction of noise in deterministic models.

Building models by generalisation is useful because it allows for a transparent, structured and systematic way of exploring the impact of various alternative assumptions that perform the same role in the model, but it often implies a loss in mathematical tractability (see e.g. Izquierdo and Izquierdo 2006). Thus, it is often the case that a rigorous study of the impact of alternative assumptions in a model requires being prepared to slide up and down the tractability continuum depicted in Fig. 11.13 (Gotts et al. 2003b). In fact, all the cases that are mentioned in the rest of this section involved greater complexity than the original models they considered, and computer simulation had to be employed to understand their dynamics.

In the literature there are many examples of the type of activity explained in this section. For example, Klemm et al. studied the relevance of various assumptions in Axelrod's model of dissemination of culture (1997b) by changing the network topology (Klemm et al. 2003a), investigating the role of dimensionality (Klemm

[14] This is so because many assumptions we make in our models are, to some extent, for the sake of simplicity. As a matter of fact, in most cases the whole purpose of modelling is to build an abstraction of the world which is simpler than the world itself, so we can make inferences about the model that we cannot make directly from the real world (Edmonds 2001; Galán et al. 2009; Izquierdo et al. 2008a).

et al. 2003b, 2005), and introducing noise (Klemm et al. 2003c). Another example is given by Izquierdo and Izquierdo (2007), who analysed the impact of using different structures of social networks in the efficiency of a market with quality variability.

In the context of decision-making and learning, Flache and Hegselmann (1999) and Hegselmann and Flache (2000) compared two different decision-making algorithms that a set of players can use when confronting various types of social dilemmas. Similarly, Takadama et al. (2003) analysed the effect of three different learning algorithms within the same model.

Several authors, particularly in the literature of Game Theory, have investigated the effect of introducing noise in the decision-making of agents. This is useful not only to investigate the general effect of potential mistakes or experimentation, but also to identify the stochastic stability of different outcomes (see Sect. 10 in Izquierdo et al. (2009)). An illustrative example is given by Izquierdo et al. (2008b), who investigate the reinforcement learning algorithm proposed by Bush and Mosteller (1955) using both mathematical analysis and simulation, and find that the inclusion of small quantities of randomness in players' decisions can change the dynamics of the model dramatically.

Another assumption investigated in the literature is the effect of different spatial topologies (see e.g. Flache and Hegselmann (2001), who generalised two of their cellular automata models by changing their – originally regular – grid structure). Finally, as mentioned in Sect. 11.5, it is increasingly common in the field of evolutionary game theory to assess the impact of various assumptions using computer simulation (see e.g. Galán and Izquierdo 2005; Santos et al. 2006; Traulsen et al. 2006; Izquierdo and Izquierdo 2006).

## 11.9 Summary

In this chapter we have provided a set of guidelines to understand the dynamics of computer models using both simulation and mathematical analysis. In doing so, it has become clear that mathematical analysis and computer simulation should not be regarded as alternative – or even opposed – approaches to the formal study of social systems, but as complementary (Gotts et al. 2003a, b). Not only can they provide fundamentally different insights on the same model, but they can also produce hints for solutions for each other. In short, there are plenty of synergies to be exploited by using the two techniques together, so the full potential of each technique cannot be reached unless they are used in conjunction.

To understand the dynamics of any particular computer model, we have seen that it is useful to see the computer model as the implementation of a function that transforms probability distributions over the set of possible inputs into probability distributions over the set of possible outputs. We refer to this function as the formal model that the computer model implements.

The mathematical approach to analyse formal models consists in examining the rules that define the model directly; the aim is to deduce the logical implications of these rules for any particular instance to which they can be applied. Our analysis of mathematical techniques to study formal models has been focused on the theory of Markov Chains. This theory is particularly useful for our purposes since many computer models can be meaningfully represented as time-homogenous Markov chains.

In contrast with mathematical analysis, the computer simulation approach does not look at the rules that define the formal model directly, but instead tries to infer general properties of these rules by examining the outputs they produce when applied to particular instances of the input space. Thus, in the simulation approach, the data is produced by the computer using strict logical deduction, but the general patterns about how the rules transform the inputs into the outputs are inferred using generalisation by induction. Thus, in the general case – and in contrast with mathematical analysis –, the inferences obtained using computer simulation will not be necessarily correct in a strict logical sense; but, on the other hand, computer simulation enables us to explore formal models beyond mathematical tractability, and the confidence we can place on the conclusions obtained with this approach can be rigorously assessed in statistical terms. Furthermore, as shown in this chapter, we can achieve any arbitrary level of accuracy in our computational approximations by running the model sufficiently many times.

Bearing in mind the relative strengths and limitations of both approaches, we have identified at least three different ways in which mathematical analysis and computer simulation can be usefully combined to produce a better understanding of the dynamics of computer models.

The first synergy appears at the time of characterising the dynamics of the formal model under study. To do that, we have shown how Markov chain analysis can be used to provide an overall picture of the dynamics of the model by categorising its different dynamic regimes and identifying when and how initial conditions are relevant. Having conducted such an analysis, one can then use computer simulation to design appropriate computational experiments with the aim of quantifying the probability distributions of the variables we are interested in. These probability distributions can always be approximated with any degree of accuracy by running the computer model several times.

The two other ways in which mathematical analysis and computer simulation can be combined derive from the fact that understanding the dynamics of a model involves not only characterising (i.e. describing) them, but also finding out *why* such dynamics are being observed (i.e. discovering causality). This often implies building supporting models that can be simpler or more complex than the original one. The rationale to move towards simplicity is to achieve greater mathematical tractability, and this often involves studying particularly simple parameterisations of the original model, and creating abstractions which are amenable to mathematical analysis. The rationale to move towards complexity is to assess the relevance of specific assumptions, and it often involves building generalisations of the original model to explore the impact of competing assumptions that can perform the same role in the model but may lead to different results.

Let us conclude by encouraging the reader to put both mathematical analysis and computer simulation in their backpack, and be happy to glide up and down the tractability spectrum where both simple and complex models lie. The benefits are out there.

**Acknowledgements** The authors have benefited from the financial support of the Spanish Ministry of Education and Science (projects DPI2004-06590, DPI2005-05676 and TIN2008-06464-C03-02), the Spanish Ministry for Science and Innovation (CSD2010-00034), and the JCyL (projects VA006B09, BU034A08 and GREX251-2009). We are also very grateful to Nick Gotts, Bruce Edmonds, Gary Polhill, and Cesáreo Hernández for many extremely useful discussions.

## Further Reading

Firstly we suggest three things to read to learn more about Markov Chain models. Grinstead and Snell (1997) provides an excellent introduction to the theory of finite Markov Chains, with many examples and exercises. Häggström (2002) gives a clear and concise introduction to Probability theory and Markov Chain theory, and then illustrates the usefulness of these theories by studying a range of stochastic algorithms with important applications in optimisation and other problems in computing. One of the algorithms covered is the Markov chain Monte Carlo method. Finally, Kulkarni (1995) provides a rigorous analysis of many types of useful stochastic processes, e.g. discrete and continuous time Markov Chains, renewal processes, regenerative processes, and Markov regenerative processes.

The reader may find three other papers helpful. Izquierdo et al. (2009) analyses the dynamics of ten well-known models in the social simulation literature using the theory of Markov Chains, and is thus a good illustration of the approach in practice within the context of social simulation.[15] Epstein (2006) is a more general discussion, treating a variety of foundational and epistemological issues surrounding generative explanation in the social sciences, and discussing the role of agent-based computational models in generative social science. Finally, Leombruni and Richiardi (2005) usefully discusses several issues surrounding the interpretation of simulation dynamics and the generalisation of the simulation results. For a different approach to analysing the dynamics of a simulation model we refer the interested reader to Chap. 9 in this volume (Evans et al. 2013).

[15] This comment was added by the editors as the authors are too modest to so describe their own work.

## References


Arthur WB (1989) Competing technologies, increasing returns, and lock-in by historical events. Econ J 99(394):116–131

Arthur WB, Holland JH, LeBaron B, Palmer R, Tayler P (1997) Asset pricing under endogenous expectations in an artificial stock market. In: Arthur WB, Durlauf S, Lane D (eds) The economy as an evolving complex system II. Addison-Wesley Longman, Reading, pp 15–44

Axelrod RM (1986) An evolutionary approach to norms. Am Polit Sci Rev 80(4):1095–1111

Axelrod RM (1997a) Advancing the art of simulation in the social sciences. In: Conte R, Hegselmann R, Terna P (eds) Simulating social phenomena (Lecture notes in economics and mathematical systems, 456). Springer, Berlin, pp 21–40

Axelrod RM (1997b) The dissemination of culture: a model with local convergence and global polarization. J Confl Resolut 41(2):203–226

Axelrod RM, Bennett DS (1993) A landscape theory of aggregation. Br J Polit Sci 23(2):211–233

Axtell RL (2000) Why agents? On the varied motivations for agent computing in the social sciences. In: Macal VM, Sallach D (eds) Proceedings of the workshop on agent simulation: applications, models, and tools, Argonne National Laboratory, Argonne, pp 3–24

Axtell RL, Epstein JM (1994) Agent based modeling: understanding our creations. The Bulletin of the Santa Fe Institute, Winter 1994, pp 28–32

Balzer W, Brendel KR, Hofmann S (2001) Bad arguments in the comparison of game theory and simulation in social studies. J Artif Soc Soc Simul 4(2), http://jasss.soc.surrey.ac.uk/4/2/1.html

Benveniste A, Métivier M, Priouret P (1990) Adaptive algorithms and stochastic approximations. Springer, Berlin

Böhm C, Jacopini G (1966) Flow diagrams, turing machines and languages with only two formation rules. Commun ACM 9(5):366–371

Bush RR, Mosteller F (1955) Stochastic models for learning. Wiley, New York

Castellano C, Marsili M, Vespignani A (2000) Nonequilibrium phase transition in a model for social influence. Phys Rev Lett 85(16):3536–3539

Cutland N (1980) Computability: an introduction to recursive function theory. Cambridge University Press, Cambridge

Edmonds B (2001) The use of models: making MABS actually work. In: Moss S, Davidsson P (eds) Multi-agent-based simulation (Lecture notes in artificial intelligence, 1979). Springer, Berlin, pp 15–32

Edmonds B (2005) Simulation and complexity: how they can relate. In: Feldmann V, Mühlfeld K (eds) Virtual worlds of precision: computer-based simulations in the sciences and social sciences. Lit-Verlag, Münster, pp 5–32

Edwards M, Huet S, Goreaud F, Deffuant G (2003) Comparing an individual-based model of behaviour diffusion with its mean field aggregate approximation. J Artif Soc Soc Simul 6(4), http://jasss.soc.surrey.ac.uk/6/4/9.html

Ehrentreich N (2002) The Santa Fe Artificial Stock Market re-examined: suggested corrections (Betriebswirtschaftliche Diskussionsbeiträge Nr. 45/02). Wirtschaftswissenschaftliche Fakultät, Martin-Luther-Universität, Halle/Saale, http://econwpa.wustl.edu:80/eps/comp/papers/0209/0209001.pdf

Ehrentreich N (2006) Technical trading in the Santa Fe Institute Artificial Stock Market revisited. J Econ Behav Organ 61(4):599–616

Epstein JM (2006) Remarks on the foundations of agent-based generative social science. In: Judd KL, Tesfatsion L (eds) Handbook of computational economics, vol. 2: agent-based computational economics. North-Holland, Amsterdam, pp 1585–1604

Epstein JM, Axtell RL (1996) Growing artificial societies: social science from the bottom up. Brookings Institution Press/MIT Press, Cambridge

Evans A, Heppenstall A, Birkin M (2013) Understanding simulation results. Chapter 9 in this volume

Flache A, Hegselmann R (1999) Rationality vs. learning in the evolution of solidarity networks: a theoretical comparison. Comput Math Organ Theory 5(2):97–127

Flache A, Hegselmann R (2001) Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics. J Artif Soc Soc Simul 4(4), http://jasss.soc.surrey.ac.uk/4/4/6.html

Flache A, Macy MW (2002) Stochastic collusion and the power law of learning. J Confl Resolut 46(5):629–653

Galán JM, Izquierdo LR (2005) Appearances can be deceiving: lessons learned re-implementing Axelrod's 'Evolutionary approach to norms'. J Artif Soc Soc Simul 8(3), http://jasss.soc.surrey.ac.uk/8/3/2.html

Galán JM et al (2009) Errors and artefacts in agent-based modelling. J Artif Soc Soc Simul 12(1), http://jasss.soc.surrey.ac.uk/12/1/1.html

Genz A, Kwong KS (2000) Numerical evaluation of singular multivariate normal distributions. J Stat Comput Sim 68(1):1–21

Gilbert N (1999) Simulation: a new way of doing social science. Am Behav Sci 42(10):1485–1487

Gilbert N (2007) Agent-based models, vol 153, Quantitative applications in the social sciences. Sage, London

Gilbert N, Terna P (2000) How to build and use agent-based models in social science. Mind Soc 1(1):57–72

Gilbert N, Troitzsch KG (1999) Simulation for the social scientist. Open University Press, Buckingham

Gotts NM, Polhill JG, Law ANR (2003a) Agent-based simulation in the study of social dilemmas. Artif Intell Rev 19(1):3–92

Gotts NM, Polhill JG, Adam WJ (2003b) Simulation and analysis in agent-based modelling of land use change. In: Online proceedings of the first conference of the European social simulation association, Groningen, The Netherlands, 18–21 Sept 2003, http://www.uni-koblenz.de/~essa/ESSA2003/proceedings.htm

Gotts NM, Polhill JG, Law ANR (2003c) Aspiration levels in a land-use simulation. Cybern Syst 34(8):663–683

Gotts NM, Polhill JG, Law ANR, Izquierdo LR (2003d) Dynamics of imitation in a land use simulation. In: Dautenhahn K, Nehaniv C (eds) Proceedings of the second international symposium on imitation in animals and artefacts, University of Wales, Aberystwyth, 7–11 April 2003, pp 39–46

Grinstead CM, Snell JL (1997) Chapter 11: Markov chains. In: Grinstead CM, Snell JL (eds) Introduction to probability (Second revised edition). American Mathematical Society, Providence, pp 405–470, http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html

Häggström O (2002) Finite Markov chains and algorithmic applications. Cambridge University Press, Cambridge

Harel D (1980) On folk theorems. Commun ACM 23(7):379–389

Hauert C, Doebeli M (2004) Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature 428(6983):643–646

Hegselmann R, Flache A (2000) Rational and adaptive playing. Anal Krit 22(1):75–97

Holland JH, Miller JH (1991) Artificial adaptive agents in economic theory. Am Econ Rev 81(2): 365–370

Huet S, Edwards M, Deffuant G (2007) Taking into account the variations of neighbourhood sizes in the mean-field approximation of the threshold model on a random network. J Artif Soc Soc Simul 10(1), http://jasss.soc.surrey.ac.uk/10/1/10.html

Imhof LA, Fudenberg D, Nowak MA (2005) Evolutionary cycles of cooperation and defection. Proc Natl Acad Sci USA 102(31):10797–10800

Izquierdo SS, Izquierdo LR (2006) On the structural robustness of evolutionary models of cooperation. In: Corchado E, Yin H, Botti VJ, Fyfe C (eds) Intelligent data engineering and automated learning – IDEAL 2006 (Lecture notes in computer science, 4224). Springer, Berlin/Heidelberg, pp 172–182

Izquierdo SS, Izquierdo LR (2007) The impact on market efficiency of quality uncertainty without asymmetric information. J Bus Res 60(8):858–867
Izquierdo LR, Polhill JG (2006) Is your model susceptible to floating point errors? J Artif Soc Soc Simul 9(4), http://jasss.soc.surrey.ac.uk/9/4/4.html
Izquierdo LR, Izquierdo SS, Gotts NM, Polhill JG (2007) Transient and asymptotic dynamics of reinforcement learning in games. Game Econ Behav 61(2):259–276
Izquierdo LR, Galán JM, Santos JI, Olmo R (2008a) Modelado de sistemas complejos mediante simulación basada en agentes y mediante dinámica de sistemas. Empiria 16:85–112
Izquierdo SS, Izquierdo LR, Gotts NM (2008b) Reinforcement learning dynamics in social dilemmas. J Artif Soc Soc Simul 11(2), http://jasss.soc.surrey.ac.uk/11/2/1.html
Izquierdo LR, Izquierdo SS, Galán JM, Santos JI (2009) Techniques to understand computer simulations: Markov chain analysis. J Artif Soc Soc Simul 12(1), http://jasss.soc.surrey.ac.uk/12/1/6.html
Janssen J, Manca R (2006) Applied semi-Markov processes. Springer, New York
Karr AF (1990) Markov processes. In: Heyman DP, Sobel MJ (eds) Stochastic models, vol 2, Handbooks in operations research and management science. Elsevier, Amsterdam, pp 95–123
Klemm K, Eguíluz VM, Toral R, San Miguel M (2003a) Nonequilibrium transitions in complex networks: a model of social interaction. Phys Rev E 67(2):026120
Klemm K, Eguíluz VM, Toral R, San Miguel M (2003b) Role of dimensionality in Axelrod's model for the dissemination of culture. Phys A 327(1–2):1–5
Klemm K, Eguíluz VM, Toral R, San Miguel M (2003c) Global culture: a noise-induced transition in finite systems. Phys Rev E 67(4):045101
Klemm K, Eguíluz VM, Toral R, San Miguel M (2005) Globalization, polarization and cultural drift. J Econ Dyn Control 29(1–2):321–334
Kulkarni VG (1995) Modeling and analysis of stochastic systems. Chapman & Hall/CRC, Boca Raton
Kulkarni VG (1999) Modeling, analysis, design, and control of stochastic systems. Springer, New York
Kushner HJ, Yin GG (1997) Stochastic approximation algorithms and applications. Springer, New York
Lebaron B, Arthur WB, Palmer R (1999) Time series properties of an artificial stock market. J Econ Dyn Control 23(9–10):1487–1516
Leombruni R, Richiardi M (2005) Why are economists sceptical about agent-based simulations? Phys A 355(1):103–109
Lieberman E, Havlin S, Nowak MA (2009) Evolutionary dynamics on graphs. Nature 433(7023): 312–316
Mabrouk N, Deffuant G, Lobry C (2007) Confronting macro, meso and micro scale modelling of bacteria dynamics. In: M2M 2007: Third international model-to-model workshop, Marseille, France, 15–16 Mar 2007, http://m2m2007.macaulay.ac.uk/M2M2007-Mabrouk.pdf
Macy MW, Flache A (2002) Learning dynamics in social dilemmas. Proc Natl Acad Sci USA 99(3):7229–7236
Miller JH, Page SE (2004) The standing ovation problem. Complexity 9(5):8–16
Nowak MA, May RM (1992) Evolutionary games and spatial chaos. Nature 359(6398):826–829
Nowak MA, Sigmund K (1992) Tit for tat in heterogeneous populations. Nature 355(6357): 250–253
Nowak MA, Sigmund K (1993) A strategy of win-stay, lose-shift that outperforms tit for tat in the Prisoner's Dilemma game. Nature 364(6432):56–58
Ostrom T (1988) Computer simulation: the third symbol system. J Exp Soc Psychol 24(5):381–392
Polhill JG, Izquierdo LR, Gotts NM (2006) What every agent based modeller should know about floating point arithmetic. Environ Model Software 21(3):283–309
Richiardi M, Leombruni R, Saam NJ, Sonnessa M (2006) A common protocol for agent-based social simulation. J Artif Soc Soc Simul 9(1), http://jasss.soc.surrey.ac.uk/9/1/15.html

Santos FC, Pacheco JM, Lenaerts T (2006) Evolutionary dynamics of social dilemmas in structured heterogeneous populations. Proc Natl Acad Sci USA 103(9):3490–3494

Suber P (2002) Formal systems and machines: an isomorphism (Electronic hand-out for the course “Logical Systems”, Earlham College, Richmond, IN) http://www.earlham.edu/~peters/courses/logsys/machines.htm

Takadama K, Suematsu YL, Sugimoto N, Nawa NE, Shimohara K (2003) Cross-element validation in multiagent-based simulation: switching learning mechanisms in agents. J Artif Soc Soc Simul 6(4), http://jasss.soc.surrey.ac.uk/6/4/6.html

Takahashi N (2000) The emergence of generalized exchange. Am J Sociol 10(4):1105–1134

Traulsen A, Nowak MA, Pacheco JM (2006) Stochastic dynamics of invasion and fixation. Phys Rev E 74(1):011909

Vilà X (2008) A model-to-model analysis of Bertrand competition. J Artif Soc Soc Simul 11(2), http://jasss.soc.surrey.ac.uk/11/2/11.html

Wikipedia (2007) Structured program theorem, http://en.wikipedia.org/w/index.php?title=Structured_program_theorem&oldid=112885072

Wilensky U (1999) NetLogo, http://ccl.northwestern.edu/netlogo

# Chapter 12
# Interpreting and Understanding Simulations: The Philosophy of Social Simulation

**R. Keith Sawyer**

**Why Read This Chapter?** To gain an overview of some key philosophical issues that underlie social simulation. Providing an awareness of them may help avoid the risk of presenting a very limited perspective on the social world in any simulations you develop.

**Abstract** Simulations are usually directed at some version of the question: What is the relationship between the individual actor and the collective community? Among social scientists, these questions generally fall under the topic of *emergence*. Sociological theorists and philosophers of science have developed sophisticated approaches to emergence, including the critical question: to what extent can emergent phenomena be reduced to explanations in terms of their components? Modelers often proceed without considering these issues; the risk is that one might develop a simulation that does not accurately reflect the observed empirical facts, or one that implicitly sides with one side of a theoretical debate that remains unresolved. In this chapter, I provide some tips for those developing simulations, by drawing on a strong recent tradition of analyzing scientific explanation that is found primarily in the philosophy of science but also to some extent in sociology.

## 12.1 Introduction

Researchers who develop multi-agent simulations often proceed without thinking about what the results will mean or how the results of the simulation might be used. After the simulation is completed, it is too late to begin to think about interpretation and understanding, because poorly designed simulations often turn out to be uninterpretable. An analysis of interpretation and understanding has to precede and inform the design. In this chapter, I provide some tips for those developing simulations, by

R.K. Sawyer (✉)
Washington University, St. Louis, MO, USA
e-mail: keith@keithsawyer.com

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_12,

drawing on a strong recent tradition of analyzing scientific explanation that is found primarily in the philosophy of science but also to some extent in sociology. My hope is that this exploration can help modelers by identifying hidden assumptions that often underlie simulation efforts, assumptions that sometimes can limit the explanatory power of the resulting simulation.

Modelers often proceed without considering these issues, in part, because of the historical roots of the approach. During the 1990s, two strands of work in computer science began to merge: *intelligent agents*, a tradition of studying autonomous agents that emerged from the artificial intelligence community, and *artificial life*, typically a graphically-represented two-dimensional grid in which each cell's behavior is determined by its nearby neighbors. In the 1990s, a few members of both of these communities began to use the new technologies to simulate social phenomena. Several intelligent agent researchers began to explore systems with multiple agents in communication, which became known as *distributed artificial intelligence*, and artificial life researchers began to draw parallels between their two-dimensional grids and various real-world phenomena, such as ant colonies and urban traffic flows. The new field that emerged within computer science is generally known as *multi agent systems*; the term refers to engineering computer systems with many independent and autonomous agents, which communicate with each other using a well-specified set of message types. When multi agent systems are used specifically to simulate social phenomena, the simulations are known as multi-agent based simulations (MABS), artificial societies, or social simulations.

Computer scientists had been working with multiple-processor systems since the 1980s, when computer scientists began to experiment with breaking up a computational task into sub-tasks and then assigning the subtasks to separate, standalone computers. During this time, a specialized massively parallel computer called the *thinking machine* was famously built by the Thinking Machines Corporation, and computer scientists began developing formalisms to represent distributed computational algorithms. These early efforts at parallel computation were centrally managed and controlled, with the distribution of computation being used to speed up a task or to make it more efficient.

The multi agent systems of the 1990s represented a significant shift from these earlier efforts, in that each computational agent was *autonomous*: capable of making its own decisions and choosing its own course of action. The shift to autonomous agents raised several interesting issues among computer scientists: if an agent were asked to execute a task, when and why would the agent agree to do it? Perhaps an agent would have a different understanding of how to execute a task than the requesting agent; how could two agents negotiate these understandings? Such questions would be unlikely to arise if one central agent had control over all of the distributed agents. But the rapid growth of the Internet – where there are, in fact, many autonomous computers and systems that communicate with each other every day – resulted in a real-world situation in which independent computational agents might choose not to respond to a request, or might respond differently than the requester expected.

As a result of these historical developments, computer scientists found themselves grappling with questions that have long been central to sociology and economics. Why should a person take an action on behalf of the collective good? How to prevent free-riding and social loafing – situations where agents benefit from collective action but without contributing very much? What is the relationship between the individual actor and the collective community? What configurations of social network are best suited to different sorts of collective tasks?

Among social scientists, these questions generally fall under the topic of *emergence* (Sawyer 2005). Emergence refers to a certain kind of relation between system-level phenomena or properties and properties of the system components. The traditional method of understanding complex system phenomena has been to break the system up into its components, to analyze each of the components and the interactions among them. In this reductionist approach, the assumption is that at the end of this process, the complex system will be fully explained. In contrast, scholars of emergence are unified in arguing that this reductionist approach does not work for a certain class of complex system phenomena. There are different varieties of this argument; some argue that the reduction is not possible for epistemological reasons (it is simply too hard to explain using reduction although the whole is really nothing more than the sum of the parts), while others argue that it is not possible for ontological reasons (the whole is something more and different than the sum of the parts). Emergent phenomena have also been described as novel – not observed in any of the components – and unpredictable, even if one has already developed a full and complete explanation of the components and of their interactions.

Among sociologists, this debate is generally known as the individualism-collectivism debate. *Methodological individualists* are the reductionists, those who argue that all social phenomena can be fully explained by completely explaining the participating individuals and their interactions. *Collectivists* argue, in contrast, that some social phenomena cannot be explained by reduction to the analysis of individuals and their interactions. Several scholars have recently noted that agent-based simulations are an appropriate tool to explore these issues (Neumann 2006; Sawyer 2005; Schmid 2006).

In the mid 1990s, a few sociologists who were interested in the potential of computer simulation to address these questions began to join with computer scientists who were fascinated with the more theoretical dimensions of these very practical questions, and the field of multi-agent based simulation (MABS) was born. Since that time, technology has rapidly advanced, and now there are several computer tools, relatively easy to use, that allow social scientists to develop multi-agent simulations of social phenomena.

After almost 10 years, this handbook provides an opportunity to look reflexively at this work, and to ask: what do these simulations mean? How should scientists interpret them? Such questions have traditionally been associated with the philosophy of science. For over a century, philosophers of science have been exploring topics that are fundamental to science: understanding, explanation, perception, validation, interpretation. In the following, I draw on concepts and arguments from within the philosophy of science to ask a question that is critical to scientists: What do these multi-agent based simulations mean? How should we interpret simulations?

In spite of the occasional participation of sociologists in social simulation projects, most modelers continue to proceed without an awareness of these important foundational questions. These are of more than simple theoretical importance; without an awareness of these somewhat complex issues of interpretation and understanding, the risk is that one might develop a simulation that does not accurately reflect the observed empirical facts, or one that implicitly sides with one side of a theoretical debate that remains unresolved (thus clouding the interpretation of the simulation's results).

I address these questions by delving into the nature of *explanation*. Many developers of MABS believe that the simulations provide explanations of real phenomena – that the concrete specifics of what is going on in the world are revealed by examining the simulation. Interpreting and understanding the simulation thus results in an explanation of the target phenomenon.

Although simulation has many potential benefits to offer scientists, I argue that the meaning of a simulation is rarely obvious and unequivocal. For example, there are many different ways to understand and interpret a given simulation of a social phenomenon. These run roughly along a spectrum from: a very narrowly focused and ungeneralizable simulation of a very specific instance of a social phenomenon, to a grand-theoretical type of simulation that explains a very broad range of social emergence phenomena. The specific end of the spectrum results in better understanding of a single phenomenon, but not in any lawful regularities nor in any general knowledge about social life. The general end of the spectrum is something like the tradition of grand theory in sociology, with generalizable laws that explain a wide range of social phenomena. The center of the spectrum is associated with what Merton famously called "middle range theories"; this is what most mainstream sociologists today believe is the appropriate task of sociology, as the field has turned away from grand theorizing in recent decades.

All of these are valid forms of sociological explanation, and each has the potential to increase our understanding of social life. My concern is with the specific end of the spectrum: in some cases, it could be that a simulation explains only a very narrow single case, with no generalizability. This would be of limited usefulness to our understanding.

## 12.2 Interpreting Multi-agent Simulations

How should scientists outside of the simulation community interpret a multi-agent simulation of a real phenomenon? Although there has been almost no philosophical attention to these simulations, simulation developers themselves have often engaged in discussions of the scientific status of their simulations. Within this community, there is disagreement about the scientific status of the simulations. Opinions fall into two camps: the "simulation as theory" camp, and the "simulation as experiment" camp.

Representing the first group, many of those developing computer simulations believe that in building them, they are engaged in a form of theory construction (Conte et al. 2001; Markovsky 1997; also see Ostrom 1988). They argue that social simulation, for example, is a more sophisticated and advanced form of social theory, because concepts and axioms must be rigorously specified to be implemented in a computer program, unlike the "discursive theorizing" of many sociologists, which is relatively vague and hard to empirically test (Conte et al. 2001; Turner 1993). As Markovsky (1997) noted, turning a (discursive) sociological theory into a simulation is not a transparent translation. A variable in the theory may turn out to be central to the simulation, or it may turn out not to matter very much; one cannot know which without going through the exercise of programming the simulation. Developing a simulation almost always reveals logical gaps in a theory, and these must be filled in before the simulation will work. As a result, simulations often introduce logical relationships that the original theory did not specify, and they contain gap-filling assumptions that the theory never made.

Representing the second group, other modelers have argued that a simulation is a *virtual experiment* (Carley and Gasser 1999). From this perspective, simulations cannot explain in and of themselves, but can only serve as tests of a theory – and the theory is what ultimately does the explaining. In a virtual experiment, a model is developed that simulates some real-world social phenomenon, but with one or more features modified to create experimental conditions that can be contrasted. For example, the same business organization could be modeled multiple times, but with a strong authority figure in one simulation, and a diffuse authority structure in another (Carley and Gasser 1999). Whereas it would probably be impossible to implement such an experiment with real-world societies, a computer model readily allows such manipulation. When the model is started, the simulations that result behave in ways that are argued to be analogous to how the real-world organization would have behaved, in each of the different conditions. In this view, because the simulation plays the role of a data-generating experiment, it doesn't provide an explanation; rather, it provides raw data to aid in theorizing, and the theory ultimately provides the explanation.

## 12.3 Scientific Explanation

Explanations are attempts to account for *why* things happen – singular events or regular, repeatable patterns. In the philosophy of science, there is a long history of discussion surrounding scientific explanation, including the deductive-nomological (D-N) or covering law approach (Hempel 1965), the statistical relevance approach (Salmon 1971), and mechanistic approaches (Bechtel and Richardson 1993; Salmon 1984). Here I limit the term "explanation" to *causal* explanation (cf. Little 1998; Woodward 2003). The relation between causation and explanation is complex; some philosophers of science hold that all explanation must be causal, whereas others deny this. For example, in the deductive-nomological tradition of logical empiricism, laws are said to provide explanations, even though the status of causation is questionable – causation is thought to be nothing more than an observed regularity as captured by a covering law. In the more recent mechanistic approach, in contrast, causation is central to explanation. I take the mechanistic position that causal mechanism is central to explanation, but I first briefly summarize the covering law notion of explanation.

In the covering law approach, a phenomenon is said to be explained when salient properties of the event are shown to be consequents of general laws, where the antecedents can also be identified. The phenomenon is said to be explained by the combination of the antecedent conditions and the laws that then result in the phenomenon. A strength of the covering-law approach is that laws both explain *and* predict; once a law is discovered, it can be used both to explain past phenomena and also to predict when similar phenomena will occur in the future.

Covering law models have always been problematic in the social sciences, primarily because of difficulty translating the notion of "law" to social reality. After all, advocates of the covering law model have had trouble adequately defining "law" even in the physical world (Hempel 1965). Candidates for social laws always have exceptions, and laws with exceptions are problematic in the DN approach. There is a history of debate concerning whether social laws exist at all, with prominent social theorists such as Anthony Giddens arguing that there are no social laws (Giddens 1984), and other prominent social theorists arguing that there are (e.g. Peter Blau (1977, 1983)). Philosophers of social science have taken various positions on the status of social laws (Beed and Beed 2000; Kincaid 1990; Little 1993; McIntyre 1996). Much of this discussion centers on what constitutes a law: must it be invariant and universal (Davidson's (1980) "strict law"), or can it admit of some exceptions? Even the strongest advocates of lawful explanation admit that there are no strict laws in the social sciences; these laws will typical have exceptions, and the law cannot explain those exceptions.

In the last decade or so, philosophers of biology (Bechtel 2001; Bechtel and Richardson 1993; Craver 2001, 2002; Glennan 1996; Machamer et al. 2000) and philosophers of social science (Elster 1989; Hedström 2005; Hedström and Swedberg 1998; Little 1991, 1998; Stinchcombe 1991) have begun to develop a different approach to explanation, one based on causal mechanisms rather than laws. In the mechanism approach, a phenomenon is said to be explained when the realizing mechanism that gave rise to the phenomenon is sufficiently described. Mechanistic accounts of explanation are centrally concerned with causation. For example, Salmon's (1984, 1994, 1997) causal mechanical model focuses on causal processes and their causal interactions; an explanation of an event traces the causal processes and interactions leading up to that event, and also describes the processes and interactions that make up the event.

Hedström (2005) presented an account of how social simulations correspond to mechanistic explanations. Mechanistic explanations differ from covering-law explanations by "specifying mechanisms that show how phenomena are brought about" (p. 24). Of course, how one defines "mechanism" is the crux of the approach. Some theorists believe that mechanisms provide causal explanations, whereas others do not. But what the approaches share is that "a mechanism explicates the details of how the regularities were brought about" (p. 24). Rather than explanation in terms of laws and regularities, a mechanism approach provides explanations by postulating the processes constituted by the operation of mechanisms that generate the observed phenomenon. For Hedström, "A social mechanism. . .describes a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome" (p. 25). The explanation is provided by the specification of often unobservable causal mechanisms, and the identification of the processes in which they are embedded. In other words, mechanists are willing to grant that macro-level regularities are observed; the covering-law approach to sociology has, after all, resulted in the identification of law-like regularities that are empirically supported. But for a mechanist, a covering law does not *explain*: "correlations and constant conjunctions do not explain but require explanation by reference to the entities and activities that brought them into existence" (p. 26).

## 12.4 MABS: Explaining by Simulating

Using MABS, researchers have begun to model the mechanisms whereby macro social properties emerge from interacting networked agents. A MABS contains many autonomous computational agents that negotiate and collaborate with each other, in a distributed, self-organizing fashion. The parallels with causal mechanism approaches in the philosophy of science are striking (Sawyer 2004).

Hedström (2005) refers to his social simulation method as *empirically calibrated agent-based models (ECA)* to emphasize that the models should be grounded in quantitative empirical data. His recommended method is to (1) develop a stylized agent-based model "that explicates the logic of the mechanism assumed to be operative" (p. 143); (2) use relevant data to verify the mechanism actually works this way; (3) run the model and modify it until it best matches relevant data. "Only when our explanatory account has passed all of these three stages can we claim to have an empirically verified mechanism-based explanation of a social outcome" (p. 144). Hedström provides an extended demonstration of the ECA method by modeling how social interactions might have given rise to the increase in youth unemployment in Stockholm in the 1990s. His model includes exactly as many computational agents as there were unemployed 20–24-year-olds in Stockholm during 1993–1999: 87,924. The demographic characteristics of these computational agents were an accurate reflection of their real-world counterparts. He then created a variety of different simulations as "virtual experiments" to see which simulation resulted in the best match to the observed empirical data.

The Stockholm simulation falls at the more specific end of the explanatory spectrum; even if it successfully simulates unemployment in Stockholm, it may not be helpful at understanding unemployment in any other city. Other MABS are designed to provide more general explanations. For example, many MABS have explored one of the most fundamental economic and sociological questions: What is the origin of social norms? For example, how do norms of cooperation and trust emerge? If autonomous agents seek to maximize personal utility, then under what conditions will agents cooperate with other agents? In game theory terms, this is a prisoner's dilemma problem. Many studies of cooperation in artificial societies have been implementations of the *iterated prisoner's dilemma (IPD)*, where agents interact in repeated trials of the game, and agents can remember what other agents have done in the past (Axelrod 1997).

The sociologists Macy and Skvoretz (1998) developed an artificial society to explore the evolution of trust and cooperation between strangers. In prior simulations of the prisoner's dilemma, trust emerged in the iterated game with familiar neighbors, but trust did not emerge with strangers. Macy and Skvoretz hypothesized that if the agents were grouped into neighborhoods, norms of trust would emerge among neighbors within each neighborhood, and that these norms would then extend to strangers. Their simulation contained 1,000 agents that played the prisoner's dilemma game with both familiar neighbors and with strangers. To explore the effects of community on the evolution of PD strategy, the simulation defined neighborhoods that contained varying numbers of agents – from nine agents per neighborhood to 50. Different runs of the simulation varied the *embeddedness* of interaction: the probability that in a given iteration, a player would be interacting with a neighbor or a stranger. These simulations showed that conventions for trusting strangers evolved in neighborhoods of all sizes, as long as agents interacted more with neighbors than strangers (embeddedness greater than 0.5). The rate of cooperation among strangers increased linearly as embeddedness was raised from 0.5 to 0.9. Simulations with smaller neighborhoods resulted in a higher rate of cooperation between strangers: at 0.9 embeddedness, the rate of cooperation between strangers was 0.62 in the 10-member neighborhood simulation, and 0.45 in the 50-member neighborhood simulation (p. 655).

Macy and Skvoretz concluded that these neighborhoods – characterized by relatively dense interactions – allow conventions for trusting strangers to emerge and become stable and then diffuse to other neighborhoods via weak ties. If an epidemic of distrusting behavior evolves in one segment of the society, the large number of small neighborhoods facilitates the restoration of order (p. 657). This simulation demonstrates how social structure can influence micro-to-macro emergence processes; cooperation with strangers emerges when agents are grouped into neighborhoods, but not when they are ungrouped.

An advocate of the causal mechanist approach to explanation would argue that the Macy and Skvoretz simulation provides candidate explanations of several social phenomena. First, the simulation explains how norms of cooperation could emerge among friends in small communities – because exchanges are iterated, and agents can remember their past exchanges with each other, they learn that cooperation works to everyone's advantage. Second, the simulation explains how norms of cooperation with strangers could emerge – as local conventions diffuse through weak ties. And in addition, the simulation explains how several variables contribute to these effects – variables like the size of the neighborhood and the embeddedness of each agent.

Advocates of a covering-law approach to explanation might prefer to think in terms of lawful generalizations. The above simulation suggests at least two: first, cooperation among strangers is greater when the neighborhoods are smaller, and second, cooperation among strangers increases linearly with embeddedness. In a D-N empiricist approach, such laws could be hypothesized and then tested through empirical study of existing human societies, and no understanding of the causal mechanism would be necessary. A mechanist like Hedström (2005) would counter that the identification of empirically supported lawful relations does not constitute an explanation. One has not identified a causal explanation until one has identified the underlying social mechanisms that realize the regularities captured by the law. The Macy and Skvoretz simulation helps to provide this form of causal explanation.

## 12.5 Potential Limitations of Simulations as Explanations

I am sympathetic to the mechanism approach. As Hedström points out, "It tends to produce more precise and intelligible explanations" (Hedström 2005, p. 28); this is desirable for sociology, which I believe must work toward being an empirically grounded and theoretically rigorous science. It reduces theoretical fragmentation, because a single mechanistic account might potentially explain many different observed phenomena, from crime to social movements; think of Gladwell's best seller *The Tipping Point* (Gladwell 2000). And finally, knowing the underlying mechanism allows you to make a certain kind of causal claim, whereas covering law approaches give you essentially only correlations.

However, causal mechanist accounts of scientific explanation can be epistemically demanding. For example, many behaviors of a volume of gas can be explained by knowing a single number, its pressure; yet a mechanist account requires the identification of the locations and movements of all of the contained molecules. A strict focus on mechanistic explanation would hold that the ideal gas law does not explain the behavior of a volume of gas; only an account in terms of the individual trajectories of individual molecules would be explanatory. And even that would be an incomplete explanation, because the gas would manifest the same macroscopic behavior even if the individual molecules had each taken a different trajectory; certainly, an explanation should be able to account for these multiple realizations.

Many advocates of mechanism are unwilling to accept any place for the covering law approach: they argue that mechanisms are the *only* proper form of sociological explanation. In several publications, Sawyer (2003a, 2004, 2005) has described a class of social phenomena in which mechanism-based explanations would be of limited usefulness: when macro-level regularities are multiply realized in a wide range of radically different underlying mechanisms. In such a situation, an account of the mechanism underlying one realization of the regularity would not provide an explanation of the other instances of that regularity. Hedström's Stockholm simulation does not necessarily explain the rise in youth employment in any other city, for example; other cities might have different realizing mechanisms. But even though it might not be possible to develop a single simulation that accurately represents unemployment processes in all large cities, it might nonetheless be possible to develop a law (or set of laws) that was capable of explaining (in the deductive-nomological sense of the term) unemployment in a large number of cities.

A social mechanist account often requires information that is unavailable, or that science is unable to provide. The Stockholm simulation was only possible because of the availability of detailed data gathered by the Swedish government. But covering law explanations can be developed with much less data, or with data at a much larger grain size. For example, many behaviors of a society can be explained by knowing whether it is individualist or collectivist (Markus and Kitayama 1991; Triandis 1995). Such properties figure in lawful generalizations like “individualist societies are more likely to be concerned with ownership of creative products” (Sawyer 2006) and “collectivist societies are more likely to practice co-sleeping” (Morelli et al. 1992). In contrast to such simple and easy-to-understand regularities, a mechanist explanation of the same patterns requires quite a bit of knowledge about each participant in that society, and their interactions with each other.

Even if a very good social simulation were developed, it might be very difficult to use that simulation to communicate to a broad, non-technical audience what meaning or understanding to attribute to the phenomenon (or the simulation). And in extremely complex systems like human societies, it may be impossible to develop an explanation of macro phenomena in terms of individual actions and interactions, even though we may all agree that such processes nonetheless must exist at the individual level. The issue here is identifying the right level of description, and the mechanistic or realizing level is often too detailed to provide us with understanding. There are many cases in science where it seems that reduction is not the best strategy for scientific explanation. For example, higher-level events like mental events supervene on physical processes but do not seem to be reducible to a unique set of causal relationships in terms of them.

The most accurate simulation would come very close to replicating the natural phenomenon in all its particulars. After such a simulation has been successfully developed, the task remains to explain the simulation; and for a sufficiently detailed simulation, that could be just as difficult as the original task of explaining the data (Cilliers 1998). Computer programmers often have difficulty explaining exactly why their creations behave as they do, and artificial society developers are no different. Mechanistic accounts of explanation need to more directly address issues surrounding levels of explanation and epistemic and computational limits to human explanation and understanding (see Sawyer 2003a, 2004).

Social simulation unavoidably touches on the unresolved sociological issue of how explanation should proceed. Social simulations represent only individual agents and their interactions, and in this they are methodologically individualist (Conte et al. 2001; also see Drennan 2005). Methodological individualism is a sociological position that has its roots in the nineteenth century origins of sociology; it argues that sociology should proceed by analyzing the individual participants in the social system, then their relations and the behaviors of bigger system components, and all the way up until we have an explanation of the social system. But if there are real emergent social properties, with downward causal powers over component individuals, then methodologically individualist simulation will fail to provide explanations of those social phenomena – for essentially the same reasons that philosophers of mind now believe that physicalism is inadequate to explain mental phenomena (see Sawyer 2002). Some social properties – such as the property of "being a collectivist society" or "being a church" – are multiply realized in widely different social systems. A simulation of a realizing mechanism of one instance of "being a church" would explain only one token instance, but would fail to broadly explain the full range of mechanisms that could realize the social property. To return to the Macy and Skvoretz simulation, the norm of cooperation could emerge in many other realizing social systems, yet the norm might have the same downward causal effects regardless of its realizing mechanism. If so, then a simulation of one realization is only a partial explanation of a more general social phenomenon; it does not explain the other ways that human cooperative behavior could be realized.

Social simulations which contain only individual agents deny a sociological realism that accepts social properties as real. If macro social properties are real, then they have an ontological status distinct from their realizing mechanisms, and may participate in causal relations (this point continues to be actively debated and the arguments are complex; see Sawyer 2003b). An accurate simulation of a social system that contains multiply realized macro social properties would have to represent not only individuals in interaction, but also these higher-level system properties and entities (Sawyer 2003a).

The problem is that although a social simulation may provide a plausible account of how individual actions and interactions give rise to an emergent macro pattern, it is hard to know (1) whether or not that social simulation in fact captures the empirical reality of the emergence; and more critically, (2) even if all agree that the social simulation accurately captures an instance of the emergence of a macro pattern from a network of individual agents, there may be other networks and other emergence processes that could also give rise to the same macro pattern.

Issue (1) is the issue of validation and it is addressed in another chapter in this same volume (David 2013). My concern here is with issue (2), which Sawyer (2005) called *multiple realizability*: a social simulation may accurately represent the social mechanisms by which individual actions together give rise to an emergent macro phenomenon. But for a given macro social phenomenon, there could potentially be many different networks of people, acting in different combinations, that result in different emergence processes that lead to the same macro social phenomenon. If so, the social simulation would not provide a complete explanation of the macro social phenomenon, but instead would only provide a limited and partial explanation.

The usual response to the multiple realizability issue is to argue that in many cases, the alternate realizations of the macro phenomena are not significantly different from each other. After all, every basketball team has five different players, but the fact that the five positions are occupied by different human beings does not substantially change the possible ways that the five can interact, and the possible ways that plays emerge from the interactions of the five individuals. Some pairs of realization are quite similar to each other, so similar that understanding one realization is tantamount to understanding the other one of the pair – without necessarily developing an entirely distinct social simulation.

The problem with this response is that it fails in the face of *wild disjunction* (Fodor 1974): when a given macro phenomenon has multiple realizations, and those realizations have no lawful relations with one another. If the multiple realizations are wildly disjunctive, then a social simulation of one of the realizations does not provide us with any understanding beyond that one realization. We are still left needing explanations of all of the other realizations. In many cases, wild disjunction is related to functionalist arguments (and the argument originated from functionalist perspectives in the philosophy of mind: Fodor 1974). “Being a church” is more likely to be multiply realized in wildly disjunctive fashion, if “church” is defined in terms of the functional needs it satisfies for its society rather than in terms of structural features internal to the institution.

The mechanist could respond to wild disjunction concerns by empirically identifying all of the different ways that the macro phenomenon in question might emerge, and then developing a suite of social simulations, each one of which would represent one of the realizing mechanisms. Then, could we say that the suite of social simulations, together, constituted a complete explanation of the macro phenomenon? I think so, although I would prefer to speak of a suite of “explanations” rather than to call the set a single explanation. The Stockholm unemployment simulation might only work for societies with generous social welfare systems; but then, another simulation could be developed for stingier governments, and two simulations together is not oppressively large given the outcome that unemployment everywhere is now fully explained.

The suite-of-simulations approach works fine as long as the number of realizing social simulations is manageable. But at some point, a suite of simulations would become so large that most sociologists would agree that it provided limited understanding of a phenomenon. Is 200 too many to be a meaningful explanation? Could as few as 20 still be too many? Even if it is computationally plausible and the number of person-hours required to develop all of the simulations is not excessive, it might nonetheless be of questionable value to undertake that work: because another path toward social explanation is available: the covering law model.

Many philosophical advocates of mechanism believe that mechanistic explanation is compatible with the existence of higher-level laws. Mechanisms are said to explain laws (Beed and Beed 2000; Bunge 2004; Elster 1998). Bunge (2004) and Little (1998) argued that causal mechanistic accounts are fully compatible with covering law explanations; the mechanisms do the explanatory work, and the covering laws provide a convenient shorthand that is often useful in scientific practice. However, it is possible that social laws may exist that are difficult to explain by identifying realizing mechanisms – in those cases where the laws relate wildly disjunctive, multiply realized social properties. If so, the scope of mechanistic explanation would be limited.

Many sociological theorists use the philosophical notion of emergence to argue that collective phenomena are collaboratively created by individuals, yet are not reducible to individual action (Sawyer 2005). In the social sciences, emergence refers to processes and mechanisms of the micro-to-macro transition. Many of these accounts argue that although only individuals exist, collectives possess emergent properties that are irreducibly complex and thus cannot be reduced to individual properties. Thus they reject sociological realism and are methodologically collectivist. Other accounts argue that emergent properties are real.

The resolution to the apparent contradiction between mechanistic explanation and social emergence is to develop a sufficiently robust account of emergence so that mechanistic explanation and lawful explanation can be reconciled. Sawyer (2002, 2003b) proposed a version of emergence that he called *nonreductive individualism (NRI)*. Some emergent social properties may be real, and may have autonomous causal powers, just like real properties at any other level of analysis. Nonreductive individualism argues that this is the case for social properties that are multiply realized in wildly disjunctive mechanisms. To the extent that social properties are multiply realized, artificial society simulations may be limited to the explanation of individual cases that do not generalize widely, resulting in a case study approach rather than a science of generalizable laws and theories. The emergentist nature of NRI is compatible with a more limited form of mechanism, but one that is elaborated in a sociologically realist direction – with the mechanisms containing explicit models of social properties at levels of analysis above the individual.

If a social property is multiply realized in many different (methodologically individualist) mechanisms, a mechanistic explanation of any one realizing instance will have limited explanatory power – particularly if the social property participates in causal relations across its multiple realizations. A covering law approach might be necessary to capture generalizations of higher-level phenomena across different realizing mechanisms. Alternately, a mechanism could be proposed which explicitly models emergent social properties, in addition to individuals and their interactions. Although almost all artificial societies are currently individualist – with no representation of higher-level social properties – there is no reason why computer simulations could not be extended to model both individuals and macrosocial phenomena, apart from an implicit commitment to methodological individualism.

## 12.6 Conclusion

Social simulations are almost all methodologically individualist, in that they represent agents and their interactions, but not higher level entities or properties. In other words, social simulations are representations of the realizing mechanisms of higher-level social properties. More generally, almost all agent-based simulations are representations of a realizing mechanism of some system-level phenomenon. Whether or not a complex system can be explained at the level of its realizing mechanisms, or requires explanation at the level of emergent macro properties, is an empirical question (Sawyer 2005). For example, it cannot be known a priori whether or not a given social property can be given a useful mechanistic explanation in terms of individuals – nor whether a given social property can be adequately simulated by representing only individuals and their interactions in the model.

If this "realizing mechanism" approach begins to seem limiting, then modelers could respond by incorporating system-level entities into their simulations. For example, sociologists could respond by developing simulations that contain the terms and properties of macro sociology, in addition to individual properties and relations. If macrosocial properties are indeed real, and have autonomous causal powers, then to be empirically accurate any model would have to incorporate those properties. Although the social mechanism approach is commonly associated with methodological individualism – because its advocates assume that a social mechanism must be described in terms of individuals' intentional states and relations (e.g. Elster 1998; Hedström 2005) – there is no reason why social simulations cannot include systems and mechanisms at higher levels of analysis. The system dynamics models of an earlier era focused on macrosocial properties; but with the availability of multi agent technology, new hybrid simulations could be developed that contain both societies of autonomous agents, and explicit simulations of emergent macro social properties.

To explore how we should interpret and understand simulations, I have drawn on contemporary philosophical accounts of explanation and of causal mechanism. I conclude by cautioning against being overly confident that agent-based simulations, at least those that are based on mechanistic assumptions, provide complete explanations of a given system-level phenomenon. The explanation may not be complete even in those cases where the simulation is well-conceived, is grounded in empirical observation and theory, and generates emergent processes that lead to empirically observed system outcomes. These successful simulations should be considered to be explanations of a given realizing instance of an emergence process, but not necessarily considered to be complete explanations of the target system phenomenon.

The question for sociologists is ultimately: What path should sociology take? All sociologists define their goal to be the explanation of macro social phenomena – of groups and organizations, rather than of single individuals. Many sociologists believe that they can explain macro social phenomena without attending to the specific realizing mechanisms at the level of individuals and their interactions. These sociologists define sociology as the science of an autonomous social level of analysis. The social mechanists believe that this is the wrong approach; instead, the goal of sociology should be to identify and characterize these individual realizing mechanisms. Mechanists believe that there can be no autonomous science at the macro social level of analysis. These are the issues to be faced as we attempt to interpret and understand social simulations.

The question for complexity researchers more generally is the same: Can a mechanistic approach provide a complete explanation, or will scientific explanation need to incorporate some higher-level properties and entities? The answer to this question has direct implications for modelers. If the mechanistic approach is capable of providing a complete explanation of a given phenomenon, then a strict agent-based approach is appropriate. But if it is necessary to incorporate higher-level properties or entities, then agent-based simulations will need to include model entities that represent higher levels of organization.

## Further Reading

(Bechtel and Richardson 1993) provides a discussion of a range of philosophical issues related to the likely success of reductionist strategies in understanding and explaining complex systems, inspired by connectionist accounts of cognition, but relevant to complex systems at any level of analysis. (Hedström 2005) makes a strong case for reductionist explanation of social systems, using mechanistic explanation and specifically, multi-agent based simulation in connection with empirical study.

For an examination of the philosophical accounts of mechanistic explanation and theories of emergence in sociology and philosophy see (Sawyer 2004). For an extensive review of historical and contemporary theories of emergence in the social sciences, primarily psychology and sociology see (Sawyer 2005). This advocates that sociology should be the science of social emergence. (Conte et al. 2001) is a discussion between four different viewpoints specifically as they concern social simulation.

## References


Axelrod R (1997) The complexity of cooperation: agent-based models of competition and collaboration. Princeton University Press, Princeton

Bechtel W (2001) The compatibility of complex systems and reduction: a case analysis of memory research. Minds Machines 11:483–502

Bechtel W, Richardson RC (1993) Discovering complexity: decomposition and localization as strategies in scientific research. Princeton University Press, Princeton

Beed C, Beed C (2000) Is the case for social science laws strengthening? J Theory Soc Behav 30(2):131–153

Blau PM (1977) A macrosociological theory of social structure. Am J Sociol 83(1):26–54

Blau PM (1983) Comments on the prospects for a nomothetic theory of social structure. J Theory Soc Behav 13(3):265–271

Bunge M (2004) How does it work?: the search for explanatory mechanisms. Philos Soc Sci 34(2):182–210

Carley KM, Gasser L (1999) Computational organization theory. In: Weiss G (ed) Multiagent systems: a modern approach to distributed artificial intelligence. MIT Press, Cambridge, pp 299–330

Cilliers P (1998) Complexity and postmodernism: understanding complex systems. Routledge, New York

Conte R, Edmonds B, Moss S, Sawyer RK (2001) Sociology and social theory in agent based social simulation: a symposium. Comput Math Organ Theory 7(3):183–205

Craver C (2001) Role functions, mechanisms and hierarchy. Philos Sci 68:31–55
Craver CF (2002) Interlevel experiments and multilevel mechanisms in the neuroscience of memory. Philos Sci 69(Supplement):S83–S97
David N (2013) Validating simulations. Chapter 8 in this volume
Davidson D (1980) Essays on actions and events. Oxford University Press, New York
Drennan M (2005) The human science of simulation: a robust hermeneutics for artificial societies. J Artif Soc Soc Simulat 8(1). http://jasss.soc.surrey.ac.uk/8/1/3.html
Elster J (1989) Nuts and bolts for the social sciences. Cambridge University Press, New York
Elster J (1998) A plea for mechanism. In: Hedström P, Swedberg R (eds) Social mechanisms: an analytical approach to social theory. Cambridge University Press, Cambridge, pp 45–73
Fodor JA (1974) Special sciences (Or: the disunity of science as a working hypothesis). Synthese 28:97–115
Giddens A (1984) The constitution of society: outline of the theory of structuration. University of California Press, Berkeley
Gladwell M (2000) The tipping point. Little, Brown, New York
Glennan SS (1996) Mechanisms and the nature of causation. Erkenntnis 44:49–71
Hedström P (2005) Dissecting the social: on the principles of analytic sociology. Cambridge University Press, Cambridge
Hedström P, Swedberg R (eds) (1998) Social mechanisms: an analytical approach to social theory. Cambridge University Press, Cambridge
Hempel CG (1965) Aspects of scientific explanation and other essays in the philosophy of science. Free Press, New York
Kincaid H (1990) Defending laws in the social sciences. Philos Soc Sci 20(1):56–83
Little D (1991) Varieties of social explanation: an introduction to the philosophy of science. Westview Press, Boulder
Little D (1993) On the scope and limits of generalization in the social sciences. Synthese 97(2):183–207
Little D (1998) Microfoundations, method and causation: on the philosophy of the social sciences. Transaction Publishers, New Brunswick
Machamer P, Darden L, Craver CF (2000) Thinking about mechanisms. Philos Sci 67:1–25
Macy MW, Skvoretz J (1998) The evolution of trust and cooperation between strangers: a computational model. Am Sociol Rev 63:638–660
Markovsky B (1997) Building and testing multilevel theories. In: Szmatka J, Skvoretz J, Berger J (eds) Status, network, and structure: theory development in group processes. Stanford, Palo Alto, pp 13–28
Markus HR, Kitayama S (1991) Culture and the self: implications for cognition, emotion, and motivation. Psychol Rev 98(2):224–253
McIntyre LC (1996) Laws and explanation in the social sciences: defending a science of human behavior. Westview, Boulder
Morelli GA, Rogoff B, Oppenheim D, Goldsmith D (1992) Cultural variation in infants' sleeping arrangements: questions of independence. Dev Psychol 28(4):604–613
Neumann M (2006) Emergence as an explanatory principle in artificial societies: reflections on the bottom-up approach to social theory. In: Squazzoni F (ed) Epistemological aspects of computer simulation in the social sciences: second international workshop, EPOS 2006, Brescia, 5–6 Oct 2006, revised selected and invited papers (Lecture notes in computer science), vol 5466. Springer, Berlin, pp 69–88
Ostrom T (1988) Computer simulation: the third symbol system. J Exp Soc Psychol 24:381–392
Salmon W (1971) Statistical explanation. In: Salmon W (ed) Statistical explanation and statistical relevance. University of Pittsburgh Press, Pittsburgh, pp 29–87
Salmon W (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton
Salmon W (1994) Causality without counterfactuals. Philos Sci 61:297–312
Salmon W (1997) Causality and explanation: a reply to two critiques. Philos Sci 64:461–477

Sawyer RK (2002) Nonreductive individualism, part 1: supervenience and wild disjunction. Philos Soc Sci 32(4):537–559

Sawyer RK (2003a) Artificial societies: multi agent systems and the micro–macro link in sociological theory. Sociol Methods Res 31(3):37–75

Sawyer RK (2003b) Nonreductive individualism, part 2: social causation. Philos Soc Sci 33(2):203–224

Sawyer RK (2004) The mechanisms of emergence. Philos Soc Sci 34(2):260–282

Sawyer RK (2005) Social emergence: societies as complex systems. Cambridge University Press, New York

Sawyer RK (2006) Explaining creativity: the science of human innovation. Oxford University Press, New York

Schmid A (2006) What does emerge in computer simulations? Simulation between epistemological and ontological emergence. In: Squazzoni F (ed) Epistemological aspects of computer simulation in the social sciences: second international workshop, EPOS 2006, Brescia, 5–6 Oct 2006, revised selected and invited papers (Lecture notes in computer science), vol 5466. Springer, Berlin, pp 60–68

Stinchcombe AL (1991) The conditions of fruitfulness of theorizing about mechanisms in social science. Philos Soc Sci 21(3):367–388

Triandis HC (1995) Individualism and collectivism. Westview Press, Boulder

Turner JH (1993) Classical sociological theory: a positivist perspective. Nelson-Hall, Chicago

Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University Press, New York

# Part III
# Mechanisms

# Chapter 13
# Utility, Games, and Narratives

**Guido Fioretti**

**Why Read This Chapter?** To appreciate how decision making can be modelled in terms of utility maximisation and game theory. To understand some of the paradoxes, limitations and major criticism of this approach and some of the alternatives.

**Abstract** This chapter provides a general overview of theories and tools to model decision-making. In particular, utility maximization and its application to collective decision-making, i.e. Game Theory, are discussed in detail. The most important exemplary games are presented, including the Prisoner's Dilemma, the Game of Chicken and the Minority Game, also known as the El Farol Bar Problem. After discussing the paradoxes and pitfalls of utility maximisation an alternative approach is introduced, which is based on appropriateness of decisions rather than consequences. An assessment of the pros and cons of competing approaches to modelling decision making concludes the chapter.

## 13.1 Introduction

This chapter provides a general overview of theories and tools to model individual and collective decision-making. In particular, stress is laid on the interaction of several decision-makers.

A substantial part of this chapter is devoted to utility maximization and its application to collective decision-making, known as Game Theory. However, the pitfalls of utility maximization are thoroughly discussed, and the radically alternative approach of viewing decision-making as constructing narratives is presented with its emerging computational tools. In detail, the chapter is structured as follows.

G. Fioretti (✉)
University of Bologna, Department of Management Science, Via Capo di Lucca 34, 40126 Bologna, Italy
e-mail: guido.fioretti@unibo.it

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_13,

Section 13.2 presents utility maximization and Game Theory with its Nash equilibria. The most important prototypical games are expounded in this section. Section 13.3 presents games that are not concerned with Nash equilibria. Section 13.4 illustrates the main paradoxes of utility maximization, as well as the patches that have been proposed to overcome them. Section 13.5 expounds the vision of decision-making as constructing a narrative, supported by an empirical case-study. Section 13.6 aims at providing computational tools for this otherwise literary vision of decision-making. Finally, Sect. 13.7 concludes by assessing the pros and cons of competing approaches.

This chapter touches so many issues that a complete list of references to the relevant literature would possibly be longer than the chapter itself. Instead of references, a guide to the most relevant bibliography is provided at the end of the chapter.

## 13.2 Utility and Games

Let $\{a_1, a_2, \ldots a_m\}$ be a set of *alternatives*. Let $a_i$ denote a generic alternative, henceforth called the $i$-th alternative where $i = 1,2,\ldots m$.

By selecting an alternative, a decision-maker obtains one out of several possible *consequences*. Let $\{c_{i1}, c_{i2}, \ldots c_{in}\}$ be the set of possible consequences of alternative $a_i$. Let $c_{ij}$ denote a consequence of $a_i$, where $i = 1,2,\ldots m$ and $j = 1, 2, \ldots n_i$.

The *expected utility* of alternative $a_i$ is:

$$u(a_i) = \sum_{j=1}^{n_i} p(c_{ij}) u(c_{ij}) \tag{13.1}$$

where $p(c_{ij})$ is the probability of obtaining consequence $c_{ij}$ and $u(c_{ij})$ is the utility of consequence $c_{ij}$.

It is suggested that the one alternative should be chosen, that maximizes expected utility. Frank Ramsey, Bruno De Finetti and Leonard Savage demonstrated that this is the only choice coherent with a set of postulates that they presented as self-evident.

Among these postulates, the following ones have the strongest intuitive appeal:

**Transitivity** Transitivity of preferences means that if $a_i \succ a_j$ and $a_j \succ a_k$, then $a_i \succ a_k$.

**Independence** Independence of irrelevant alternatives means that $a_i \succ a_j$ iff $a_i \cup a_k \succ a_j \cup a_k, \ \forall a_k$.

**Completeness** Completeness means that $\forall (a_i, a_j)$, a preference relation $\succ$ is defined.

Utility maximization is neither concerned with conceiving alternatives, nor with the formation of preferences, which are assumed to be given and subsumed by the utility function. Probabilities may eventually be updated by means of frequency measurement, but at least their initial values are supposed to be given as well. Thus, utility maximization takes as solved many of the problems with which its critics are concerned.

Utility maximization takes a gambler playing dice or roulette as its prototypical setting. In fact, in this setting the set of alternatives is given, utilities coincide with monetary prizes and probabilities can be assessed independently of utilities. By contrast, for some critics of utility maximization, gambling is not an adequate prototype of most real-life situations.

The interaction of several utility-maximizing decision-makers is covered by *Game Theory*. Game Theory assumes that collective decision-making is the combination of several individual decision processes, where each individual maximizes his utility depending on the alternatives selected by the other individuals. Since selecting an alternative implies considering what alternatives other players may select, alternatives are generally called *strategies* in this context.

Utility is called *payoff* in Game Theory. Games in which one player does better at another's expense are called *zero-sum games*. Games may be played once, or they may be repeated.

The bulk of Game Theory is concerned with equilibria. If each player knows the set of available strategies and no player can benefit by changing his or her strategy while the other players keep theirs unchanged, then the current choice of strategies and the corresponding payoffs constitute a *Nash equilibrium*. Since this implies stepping in another player's shoes in order to figure out what she would do if one selects a particular strategy, Nash-equilibria are fixed points in self-referential loops of the kind "I think that you think that I think ...".

Note that being at a Nash equilibrium neither implies that each player reaches the highest possible payoff that she can attain, nor that the sum of all payoffs of all players is the highest that can be attained. This is eventually a concern for economics, for it implies that individual interests may not produce the common good.

If a game is repeated, a Nash equilibrium may be either realized with *pure strategies*, meaning that the players choose consistently one single alternative, or *mixed strategies*, meaning that the players select one out of a set of available strategies according to a probability distribution. Accepting the idea of mixed strategies often allows to find Nash equilibria where there would be none if only pure strategies are allowed. However, the realism of random decision-makers choosing strategies according to a probability distribution, is at least questionable.

Most of the games analysed by Game Theory involve two, or in any case a very limited number of players. On the contrary, *evolutionary games* (Weibull 1997) concern large populations of players playing different strategies, that are subject to an evolutionary dynamics regulated by *replicator equations*. Successful strategies replicate and diffuse, unsuccessful strategies go extinct. Occasionally, new strategies may originate by random mutation.

The equilibrium concept of evolutionary games is that of *evolutionarily stable strategies*. An evolutionary stable strategy is such that, if almost every member of the

| | Opera | Football |
|---|---|---|
| Opera | 3, 2 | 0, 0 |
| Football | 0, 0 | 2, 3 |

| | | |
|---|---|---|
| | M, L | 0, 0 |
| | 0, 0 | L, M |

**Fig. 13.1** A payoff matrix for the Battle of the Sexes (*left*) and its generic representation (*right*). The *left number* is the payoff of the row player (wife), the *right number* is the payoff of the column player (husband). In this generic representation, *L* is the payoff of the least preferred alternative whereas *M* is the payoff of the most preferred alternative

population follows it, no mutant can successfully invade. Alternatively, evolutionary games may be played in order to observe typical dynamics, in which case they become akin to the influence games that will be handled in Sect. 13.3.

The following games propose prototypical modes of human interaction. Games used by experimental economics in order to evince human attitudes do not pertain to this list.

### *13.2.1 The Battle of Sexes*

Imagine a couple where the husband would like to go to the football game, whereas the wife would like to go to the opera. Both would prefer to go to the same place rather than different ones. They cannot communicate, so they cannot agree, e.g., to go alternatively to the opera and to the football game.

The payoff matrix in Fig. 13.1 is an example of the Battle of Sexes, where the wife chooses a row and the husband chooses a column. Aside, a generic representation of the game where $L < M$.

This representation does not account for the additional harm that might come from going to different locations and going to the wrong one, i.e., the husband goes to the opera while the wife goes to the football game, satisfying neither. Taking account of this effect, this game would bear some similarity to the Game of Chicken of Sect. 13.2.7.

This game has two pure-strategy Nash-equilibria, one where both go to the opera and another where both go to the football game. Furthermore, there is a Nash equilibrium in mixed strategies, where the players go to their preferred event more often than to the other one.

None of these equilibria is satisfactory. One possible resolution involves a commonly observed randomizing device, e.g., the couple may agree to flip a coin in order to decide where to go.

| | Stag | Hare |
|---|---|---|
| Stag | 3, 3 | 0, 1 |
| Hare | 1, 0 | 1, 1 |

| | |
|---|---|
| C, C | S, B |
| B, S | D, D |

**Fig. 13.2** A payoff matrix for the Stag Hunt (*left*) and its generic representation (*right*). The *left number* is the payoff of the row player, the *right number* is the payoff of the column player. In this generic representation, $C$ is the payoff that accrues to both players if they cooperate, $D$ is the payoff that accrues to both players if they defect from their agreement, $S$ is the sucker's payoff and $B$ is the betrayer's payoff

## *13.2.2 The Stag Hunt*

Rousseau described a situation where two individuals agree to hunt a stag, which none of them would be able to hunt alone. Each hunter may eventually notice a hare and shoot at it. This would destroy the stag hunt, so the other hunter would get nothing.

An example of the payoff matrix for the stag hunt is pictured in Fig. 13.2, along with its generic representation. The stag hunt requires that $C > B \geq D > S$.

This game has two pure-strategy Nash-equilibria, one where both hunters hunt the stag, the other one where both hunters hunt a hare. The first equilibrium maximizes payoff, but the second equilibrium minimizes risk. There exists also a mixed-strategy Nash-equilibrium, but no payoff matrix can make the hunters play "stag" with a probability higher than 1/2.

The Stag Hunt exemplifies the idea of society originating out of contracts between individuals. The examples of "social contract" provided by Hume are Stag Hunts:

- Two individuals must row a boat. If both choose to row they can successfully move the boat, but if one does not, the other wastes his effort.
- Two neighbours wish to drain a meadow. If they both work to drain it they will be successful, but if either fails to do his part the meadow will not be drained.

Several animal behaviours have been described as stag hunts. For instance, orcas corral large schools of fish to the surface and stun them by hitting them with their tails. This works only if fishes do not have ways to escape, so it requires that all orcas collaborate to kill all fishes they caught rather than catching a few of them.

## *13.2.3 The Prisoner's Dilemma*

The Prisoner's Dilemma is a central subject in economics, for it apparently contradicts its basic assumption that common good arises out of self-interested individuals. This difficulty is eventually overcome by repeating the game.

| | Cooperate | Defect |
|---|---|---|
| Cooperate | 3, 3 | 0, 5 |
| Defect | 5, 0 | 1, 1 |

| | |
|---|---|
| C, C | S, B |
| B, S | D, D |

**Fig. 13.3** A payoff matrix for the Prisoner's Dilemma (*left*) and its generic representation (*right*). The *left number* is the payoff of the row player, *right number* is the payoff of the column player. In this generic representation, $C$ is the payoff if both players cooperate, $D$ is the payoff if both defect from their agreement, $S$ is the sucker's payoff, $B$ is the betrayer's payoff

The basic formulation of the Prisoner's Dilemma is as follows. Two suspects, A and B, are arrested by the police. Having insufficient evidence for a conviction, the police visits them separately, offering each prisoner the following deal: if one testifies for the prosecution against the other and the other remains silent, the betrayer goes free and the silent accomplice receives the full 10-year sentence. If both stay silent, both prisoners are sentenced to only 6 months in jail for a minor charge. If each betrays the other, each receives a 5-year sentence. Each prisoner must make the choice of whether to betray the other or to remain silent; unfortunately, neither prisoner knows what choice the other prisoner made.

The Prisoner's Dilemma describes any situation where individuals have an interest to be selfish, though if everyone cooperates a better state would be attained. Examples may include unionising, paying taxes, not polluting the environment, or else. Figure 13.3 illustrates a payoff matrix for the Prisoner's Dilemma, as well as its generic representation. The Prisoner's Dilemma requires that $B > C > D > S$.

The Prisoner's Dilemma has only one Nash equilibrium at $(D, D)$. Notably, all individual incentives push towards this equilibrium. Nevertheless, this equilibrium is not socially optimal.

Eventually, the difficulty raised by the Prisoner's Dilemma can be overcome if players can repeat the game (which requires $2C > B + S$). In particular, by playing the Prisoner's Dilemma as an evolutionary game with large numbers of players and strategies it is possible that islands of cooperation sustain themselves in a sea of selfish choices. One possibility for islands of cooperation to emerge is to allow reciprocity, e.g., with a "tit-for-tat" strategy: start with cooperating whenever you meet a new player, but defect if the other does. Another possibility is that players cooperate when they meet players that exhibit a randomly selected tag – e.g., a tie may be worn in order to inspire confidence – so that islands of cooperation emerge even if agents have no memory.

### *13.2.4 The Traveller's Dilemma*

The Traveller's dilemma is a non-zero-sum game in which two players attempt to maximize their own payoff, without any concern for the other player's payoff. It is a game that aims at highlighting a paradox of rationality. It is a thought experiment on the following problem.

An airline loses two suitcases belonging to two different travellers. The suitcases contain identical antiques. An airline manager tasked to settle the claims of both travellers explains that the airline is liable for a maximum of $100 per suitcase, and in order to determine a honest appraised value of the antiques the manager separates both travellers and asks each of them to write down the amount of their value at no less than $2 and no more than $100. He also tells them that if both write down the same number, he will treat that number as the true value of both suitcases and reimburse both travellers that amount. However, if one writes down a smaller number than the other, this smaller number will be taken as the true value, and both travellers will receive that amount along with a bonus/malus: $2 extra will be paid to the traveller who wrote down the lower value and a $2 deduction will be taken from the person who wrote down the higher amount. The challenge is: what strategy should both travellers follow in order to decide what value they should write down?

If this game is actually played, nearly all the times everyone chooses $100 and gets it. However, rational players should behave differently.

Rational players should value the antique slightly less than their fellow traveller, in order to get the the bonus of $2. For instance, by pricing at $99 one would get $101, whereas the opponent would get $97. However, this triggers an infinite regression such that $2 is the only Nash-equilibrium of this game. Thus, being rational does not pay.

The Traveller's Dilemma suggests that in reality people may coordinate and collaborate because of their bounded rationality, rather than in spite of it. If they would be smarter than they are, they would obtain less.

### 13.2.5 The Dollar Auction

The dollar auction is a non-zero sum sequential game designed to illustrate a paradox brought about by rational choice theory. In this game, players with perfect information are compelled to make an ultimately irrational decision based on a sequence of rational choices.

The game involves an auctioneer who offers a one-dollar bill with the following rule: the dollar goes to the highest bidder, who pays the amount he bids. The second-highest bidder must also pay the highest amount that he bids, but gets nothing in return.

Suppose that the game begins with one of the players bidding 1 cent, hoping to make a $0.99 profit. He will be quickly outbid by another player bidding 2 cents, as a $0.98 profit is still desirable. Similarly, another bidder may bid 3 cents, making a $0.97 profit. At this point the first bidder may attempt to convert their loss of 1 cent into a gain of $0.97 by also bidding 3 cents. In this way, a series of bids is maintained.

One may expect that the bidders end up with offering $1.00 for a one-dollar bill, which is what the auction is for. However, a problem becomes evident as soon as the bidding reaches 99 cents. Suppose that one player had bid 98 cents. The other players now have the choice of losing 98 cents or bidding a dollar even, which would make their profit zero. After that, the original player has a choice of either losing 99 cents or bidding $1.01, losing only 1 cent. After this point these rational players continue to bid the value up well beyond the dollar, and neither makes a profit.

### *13.2.6 Pure Coordination Games*

Pure coordination games are an empirical puzzle for Game Theory. Pure coordination games are one-shot games where players face a set of alternatives knowing that a positive payoff will only accrue to them if they coordinate on the same choice. For instance, two subjects may be shown a city map and asked, independently of one another, to select a meeting point. Or, subjects may be asked to select a positive integer. In the first case they obtain a positive payoff if they select the same meeting point; in the second case, if they select the same integer.

The difficulty of pure coordination games derives from the fact that players cannot communicate and that the game is not repeated. The astonishing fact about pure coordination games is that, if they are actually played, players reach an agreement much more often than they would if they would play randomly.

The commonly held explanation is that pure coordination games generally entail cues that single out one choice as more "salient" than others. For instance, subjects asked to select a meeting point generally end up with the railway station, whereas the majority of those asked to name a positive integer select the number 1.

However, this suggests that coordination may eventually be attained because of conventions, habits or values that do not enter the description of decision settings. People may not even be aware of what makes them coordinate with one another.

### *13.2.7 The Game of Chicken*

The game of Chicken models two drivers, both headed for a single lane bridge from opposite directions. One must swerve, or both will die in the crash. However, if one driver swerves but the other does not, he will be called a "chicken". Figure 13.4 depicts a typical payoff matrix for the Chicken Game, as well as its generic form.

Chicken is an anti-coordination game with two pure-strategy Nash-equilibria where each player does the opposite of what the other does. Which equilibrium is selected depends very much on the effectiveness in signaling pre-commitment before the game is played. For instance, a driver who disables the brakes and the

| | Swerve | Straight | | |
|---|---|---|---|---|
| Swerve | 0, 0 | −1, +1 | V/2, V/2 | 0, V |
| Straight | +1, −1 | −10, −10 | V, 0 | $\frac{(V-C)}{2}, \frac{(V-C)}{2}$ |

**Fig. 13.4** A payoff matrix for the Game of Chicken (*left*) and its generic representation (*right*). The *left number* is the payoff of the row player, the *right number* is the payoff of the column player. In this generic representation, *V* is the value of power, prestige, or of the available resource to be obtained, *C* is the cost if both players choose "straight"

steering wheel of his car and makes it known to the other driver, may induce him to swerve.

Bertrand Russell remarked that the nuclear stalemate was much like the Game of Chicken[1]:

> As played by irresponsible boys, this game is considered decadent and immoral, though only the lives of the players are risked. But when the game is played by eminent statesmen, who risk not only their own lives but those of many hundreds of millions of human beings, it is thought on both sides that the statesmen on one side are displaying a high degree of wisdom and courage, and only the statesmen on the other side are reprehensible. This, of course, is absurd. Both are to blame for playing such an incredibly dangerous game. The game may be played without misfortune a few times, but sooner or later it will come to be felt that loss of face is more dreadful than nuclear annihilation. The moment will come when neither side can face the derisive cry of 'Chicken! ' from the other side. When that moment is come, the statesmen of both sides will plunge the world into destruction.

The Game of Chicken has been re-interpreted in the context of animal behaviour. It is known as *Hawk-Dove Game* among ethologists, where the Hawk-Dove game has the same payoff matrix as in Fig. 13.4. In the Hawk-Dove game, "swerve" and "straight" correspond to the following strategies, respectively:

**Dove** Retreat immediately if one's opponent initiates aggressive behaviour;
**Hawk** Initiate aggressive behaviour, not stopping until injured or until the opponent backs down.

Whilst the original Game of Chicken assumes $C > V$ and cannot be repeated, the Hawk-Dove game lacks this requirement and is generally conceived as an evolutionary game.

The strategy "Dove" is not evolutionary stable, because it can be invaded by a "Hawk" mutant. If $V > C$, then the strategy "Hawk" is evolutionarily stable. If $V < C$ there is no evolutionarily stable strategy if individuals are restricted to following pure strategies, although there exists an evolutionarily stable strategy if players may use mixed strategies.

[1] Russell BW (1959) Common Sense and Nuclear Warfare. George Allen/Unwin, London.

### 13.2.8 The War of Attrition

The war of attrition is a game of aggression where two contestants compete for a resource of value $V$ by persisting with their intentions while constantly accumulating costs. Equivalently, this game can be seen as an auction in which the prize goes to the player with the highest bid $B_h$, and each player pays the loser's low bid $B_l$.

The war of attrition cannot be properly solved using its payoff matrix. In fact, the players' available resources are the only limit to the maximum value of bids. Since bids can be any number, if available resources are ignored, then the payoff matrix has infinite size. Nevertheless, its logic can be analysed.

Since players may bid any number, they may even exceed the value $V$ that is contested over. In fact, if both players bid higher than $V$, the high bidder does not so much win as lose less, in the sense that $-B_l < V - B_h < 0$ – a Pyrrhic victory.

Since there is no value to bid which is beneficial in all cases, there is no dominant strategy. However, this does not preclude the existence of Nash-equilibria. Any pair of strategies such that one player bids zero and the other player bids any value equal to $V$ or higher, or mixes among any values $V$ or higher, is a Nash-equilibrium.

The War of Attrition is akin to a Chicken or Hawk-Dove game – see Sect. 13.2.7 – where if both players choose “swerve”/“Dove” they obtain 0 instead of $V/2$ as in Fig. 13.4.

The evolutionarily stable strategy when playing it as an evolutionary game is a probability density of random persistence times which cannot be predicted by the opponent in any particular contest. This result has led to the conclusion that, in this game, the optimal strategy is to behave in a completely unpredictable manner.

## 13.3 Influence Games

Contrary to those of Sect. 13.2, the games in this section are not concerned with Nash-equilibria. Players are not assumed to figure out which alternatives the other players might choose, originating infinite regressions that can only stop at equilibrium points.

Rather, boundedly rational players are assumed to follow certain rules, that may be quite simple but need not be necessarily so. The game then concerns what collective behaviours emerges out of mutual influence.

However, the games in this section are not so different from those of Sect. 13.2 when they are played as evolutionary games. Such is the case, for instance, of simulations where a large number of players iterate the Prisoner's Dilemma.

Two prototypical games will be expounded in this section. The *Ising model* (originally developed in physics, where it is also known as *spin glass* model) is concerned with imitation. The *minority game*, also known as the *El Farol Bar*

*Problem*, is concerned with the contrary of imitation. It is about doing the opposite of what others do.

### 13.3.1 The Ising Model

The Ising model was originally developed in physics in order to study the interaction between atoms in a ferromagnetic material. For this reason its agents can only take two states, or opinions in social applications, and are fixed in space.

The Ising model is an exceedingly stylized model of imitation dynamics. Clearly, many imitation models are more complex and more realistic than the Ising model. However, the closed-form solutions of the Ising model may guide the builder of more complex models in the process of understanding their behaviour.

In general, the Ising model is not presented as a game. It is done here in order to stress its symmetry with the minority game.

Let $N$ players be denoted by means of an index $i = 1,2,\ldots N$. Players must choose between an alternative $A = -1$ and an alternative $A = 1$.

The payoff of a player does not only depend on the alternative that she has chosen, but also on the average of the alternatives chosen by the other players. Let $m$ denote this average.

Since we want to reproduce situations where the individual follows the herd, the effect of $m$ should be the stronger, the more homogeneous the group. Since $A \in \{-1;1\}$ and consequently $m \in \{-1;1\}$, we can reach this goal by requiring that the payoff depends on a term $A\ m$. This term may eventually be multiplied by a coefficient $J > 0$.

A stochastic term $\varepsilon$ is necessary in order to understand our game as a system jumping between many equilibria. This term will disappear when expected values will be taken.

In the end, the following functional form is chosen for the payoff of a player:

$$u(A) = v(A) + J\ A\ m + \varepsilon \tag{13.2}$$

where $u(A)$ is the total payoff of a player and $v(A)$ is its individual component.

Furthermore, let us assume that this individual component takes the following form:

$$v(A) = \begin{cases} -h & \text{if A} = -1 \\ h & \text{if A} = 1. \end{cases} \tag{13.3}$$

where $h \in \Re,\ h > 0$.

By assuming that the stochastic terms $\varepsilon_i$ are Gumbel-distributed, we can apply the logit model. By combining Eqs. 13.2 and 13.3 we derive the following expressions for the probability that a player selects one of the two alternatives:

$$p\{A=-1\} = \frac{e^{\mu(-h-Jm+\varepsilon)}}{e^{\mu(-h-Jm+\varepsilon)} + e^{\mu(h+Jm+\varepsilon)}} \tag{13.4}$$

$$p\{A=1\} = \frac{e^{\mu(h+Jm+\varepsilon)}}{e^{\mu(-h-Jm+\varepsilon)} + e^{\mu(h+Jm+\varepsilon)}} \tag{13.5}$$

The expected value of the selected alternative is $E\{A\} = -1 \cdot p\{A=-1\} + 1 \cdot p\{A=1\}$. Since it is also $E\{A\} = m$ we obtain the following expression:

$$m = \tanh(\mu h + \mu Jm) \tag{13.6}$$

where $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$ is the hyperbolic tangent.

Equation 13.6 provides an analytic description of a game with herd behaviour on two alternatives described by means of a mean-field approximation. It admits a closed form solution that provides the following findings:

- If $\mu J < 1$ and $h = 0$ there exists one single solution at $m = 0$. Consider that this is a discrete-time system, so its attractors are stable if all eigenvalues of the state transition function are in $(-1,1)$. Intuitively, $\mu J < 1$ means that this system is globally stable. Furthermore, $h = 0$ means that the individual component of the payoff is zero so the players have no incentive to choose one of the two alternatives. Consequently, the stochastic term makes $m = 0$ the only solution.
- If $\mu J < 1$ and $h \neq 0$ there exists one single solution with the same sign as $h$. In fact, as in the previous case the system is globally stable so it admits one single solution. However, since in this case the players' payoff includes an individual component, it is this component that determines what equilibriun is reached. If most players prefer $A = -1$ the equilibrium will be $m \approx -1$; likewise, if most players prefer $A = 1$ then the equilibrium will be $m \approx 1$.
- If $\mu J \geq 1$ and $h = 0$ there exist three solutions: $m = 0$ and $m = \pm m(\mu J)$. In fact, the system is globally unstable but locally stable equilibria may exist. Since the individual component of the payoff is zero, the system may either tend towards $m = 0$, or $m \approx -1$, or $m \approx 1$.
- If $\mu J \geq 1$ and $h \neq 0$, the following subcases must be distinguished:
  - If, for any given $\mu$ and $J$, there exists a threshold $H(h) > 0$ such that $|h| \leq H$, then three solutions exist, one with the same sign as $h$ and the other two with opposite sign. Condition $|h| < H$ means that the individual component of the payoff is limited even if not zero. Therefore, results are similar to the previous case.
  - If, for any given $\mu$ and $J$, there exists a threshold $H(h) > 0$ such that $|h| > H$, then there exists one single solution with the same sign as $h$. In fact, if the individual component of the payoff can take any value, then the whole system is forced into its direction.

In the Ising model, each player observes the average behaviour of all other players. If each player observes only the behaviour of his neighbours, one obtains Schelling's model of racial segregation (According to Schelling's model of racial segregation, a city where Blacks and Whites are randomly distributed turns into a chessboard of homogeneous quarters if its inhabitants, although absolutely ready to accept the presence of the other colour, do not want to be a small minority in their own neighbourhoods).

### 13.3.2 The Minority Game

The minority game originates from a consideration inspired by the *El Farol* bar in Santa Fe, New Mexico (USA). The economist Brian Arthur remarked that people go to the bar in order to meet other people, but they do not want to go when all other people go, because the bar is too crowded on such occasions. Thus, they want to do the opposite of what most people do –go to the bar when most people stay at home, stay at home when most people go to the bar. This is interesting, because the "El Farol Bar Problem" cannot have a stable equilibrium. In fact, once the majority observed what the minority did, it wants to imitate it, which turns the minority into majority, and so on endlessly.

Physicists Damien Challet and Yi-Cheng Zhang remarked that this is the essence of stock market dynamics. In fact, in the stock market those traders gain, who buy when prices are low (because most traders are selling) and sell when prices are high (because most traders are buying). So all traders want to belong to the minority, which is clearly impossible, hence the inherent instability of this game. Among the physicists, the "El Farol Bar Problem" became the "Minority Game".[2]

Let us consider $N$ players who either belong to a group denoted 0 or a group denoted 1. Players belonging to the minority group receive a positive payoff. Players belonging to the majority receive zero.

Strategies are functions that predict which will be the minority group in the next step given the minority group in the $m$ previous steps. Thus, a strategy is a matrix with $2^m$ rows (dispositions with repetition of two elements of class $m$) and two columns. The first column entails all possible series of minority groups in the previous $m$ steps, henceforth *histories*. The second column entails the group suggested to be minority in the next step. As an example, Fig. 13.5 illustrates a strategy with $m = 2$.

Each player owns $s$ strategies. If $s = 1$, the game is trivial because the time series of the minority group is periodical.

If $s > 1$, players choose the strategy that cumulated the greatest amount of payoffs. Thus, a number of feedbacks may arise between what strategies are chosen

[2] The rest of this section has been extensively drawn from E. Moro, *The Minority Game: An Introductory Guide*, working paper available online.

| *History* | *Prediction* |
|---|---|
| 0 0 | 0 |
| 0 1 | 1 |
| 1 0 | 1 |
| 1 1 | 1 |

**Fig. 13.5** An example of a strategy based on the two previous steps of the minority game. The first column lists all possible stories. The second column, depending on past history, makes a prediction

and their capability to predict the minority. In fact, in this game players must adapt to an environment that they themselves create.

An important magnitude in this game is the variance of the time series of the number of players belonging to group 1 (or, equivalently, group 0). Henceforth, this magnitude will be denoted by $\sigma^2$.

The average of the number of players belonging to each group is generally close to $N/2$. If $\sigma^2$ is small, then the distribution of the number of players belonging to group 1 is concentrated around $N/2$. This implies that the minority is large, eventually close to its maximum ($N/2 - 1$). On the contrary, if $\sigma^2$ is large the number of players belonging to group 1 tends to be either much smaller or much larger than $N/2$, implying that the minority is often very small.

Let us consider $\sigma^2/N$ in order to normalize to the number of players. Let us define the *efficiency of coordination* $e_c = N/\sigma^2$ as the reciprocal of the extent to which players behave differently from one another.

Figure 13.6 depicts numerical simulations of $e_c$ as a function of the number of histories in a strategy $2^m/N$. Graphs are shown for different values of $s$. The horizontal line marks the value that $e_c$ attains if players would make a random choice among the strategies available to them.

With low $m$ the efficiency of coordination is low. This happens because if memory is short, then players have greater difficulties to adapt to the changing features of the game.

If only few strategies are available ($s = 2$, $s = 3$, $s = 4$), at intermediate values of $m$ many players guess the correct strategy so $e_c$ increases above the level that can be attained if strategies are chosen randomly. This threshold is marked by the dashed vertical line. However, this effect disappears if many strategies are available ($s = 8$, $s = 16$). In this case the decision process becomes similar to a random choice so even at intermediate values of $m$ the efficiency of coordination is close to the level attained when strategies are chosen randomly.

Independently of the number of available strategies, with increasing $m$ the value of $e_c$ tends to the level attained when strategies are chosen randomly. In fact, a history of length $m$ occurs again after $2^m$ steps on average, so a strategy that is successful with a particular history needs $2^m$ steps in order to be successful again. With very high values of $m$, no strategy can present itself as particularly successful; therefore, a nearly-random dynamics ensues.

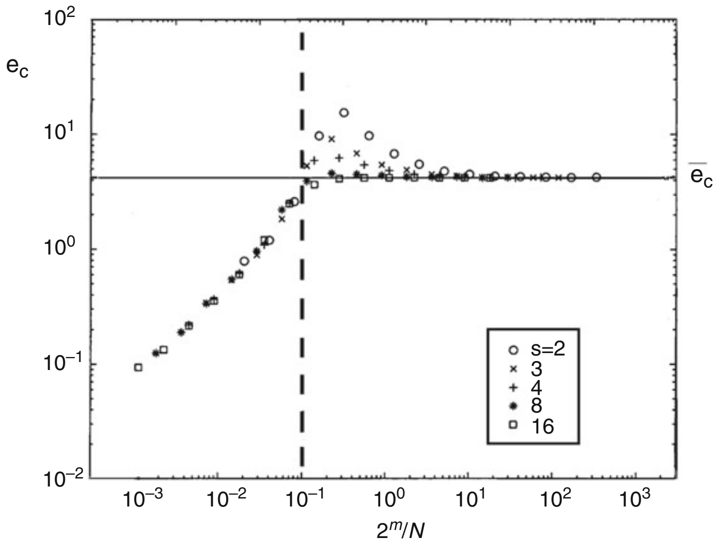



**Fig. 13.6** Efficiency of coordination $e_c$ as a function of the number of histories in a strategy $2^m/N$, for different values of the number of available strategies $s$. The *horizontal line* at $e_c = \overline{e_c}$ marks the efficiency level when player select a strategy at random. The *vertical dashed line* marks the point where $e_c$ can be greater than $\overline{e_c}$

Let us consider what information is available to players. The only information available to them is what group was the minority in previous time steps. Let this information be carried by a variable $W_t$, where $W_t = 0$ means that at time $t$ the group 0 has been minority, $W_t = 1$ otherwise. The issue is whether this information is used efficiently; if it is not, there may exist exist arbitrage possibilities for players who utilize information more efficiently than their peers.

Let us consider $W_t$ and $W_{t+1}$ as distinct signals. Let us compute their mean mutual information $I(W_t, W_{t+1})$.[3]

Mean mutual information measures whether the information entailed in the outcomes of two steps of the game, taken together, is greater than the sum of the information entailed in the outcomes of the two steps taken independently of one another. Thus, mean mutual information says whether a player, by observing the time series of the outcome of the game, could do better than his peers.

[3] Given a source of binary symbols $\{a_1, a_2, \ldots a_M\}$ issued with probabilities $p_1, p_2, \ldots p_M$, the average information that they convey is defined as $H(A) = \sum_{i=1}^{M} p(a_i) \lg_2 1/p(a_i)$ and it is called *information entropy*. Suppose that there is a second source issuing symbols $\{b_1, b_2, \ldots b_N\}$ with information entropy $H(B)$. Let $H(A,B)$ denote the information entropy of the whole system. *Mean mutual information* $H(A) + H(B) - H(A,B)$ measures to what extent the two sources interact to correlate their messages. Mean mutual information is zero if the two sources are independent of one another.

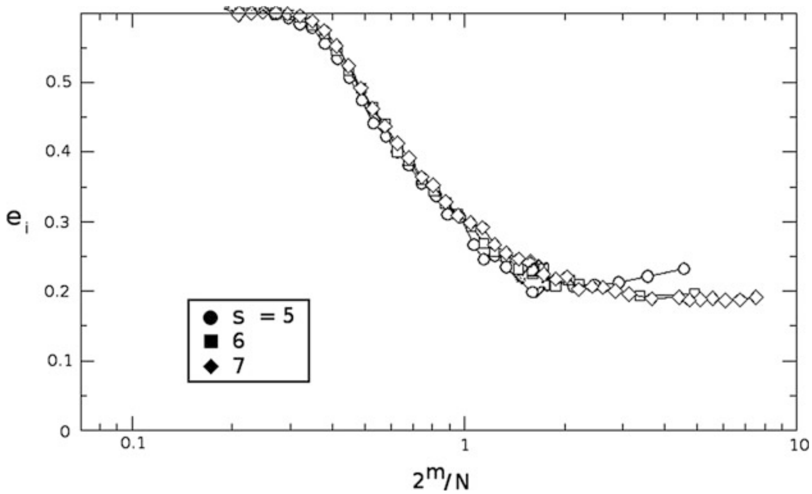



**Fig. 13.7** Efficiency of information exploitation as a function of the number of stories in a strategy, normalized to the number of players

Recalling the analogy with the stock market, $I(W_t, W_{t+1}) > 0$ means that a trader could gain from arbitrage.

Let us introduce *information efficiency* $e_i = 1/I(W_t, W_{t+1})$. Being the reciprocal of mean mutual information, information efficiency is high when mean mutual information is low, i.e., when information is efficiently exploited by the player so there is little room for arbitrage.

Figure 13.7 depicts numerical simulations of $e_i$ as a function of the number of stories in a strategy $2^m/N$. Graphs are shown for different values of $s$.

One may observe in Fig. 13.7 a sudden drop of $e_i$ in the [0.3,1] interval. This interval is entailed in the interval [0.1,1] where $e_c$ was observed to rise above the level corresponding to random choice in Fig. 13.6. Thus, we may subsume the behaviour of the minority game in the following Table 13.1:

Table 13.1 shows that the minority game has two large behaviour modes, one inefficient in coordination but efficient in the exploitation of information, the other one efficient in coordination but inefficient in the exploitation of information. In between, a tiny space where the efficiency of coordination and the efficiency of information exploitation may change dramatically depending on $s$ and $m$.

Since the minority game is a stylised representation of stock markets, we may ask in which region stock markets operate. It is well known that very many traders operate in stock markets, so we may assume that $N$ is very large. Human bounded rationality suggests that traders do not make use of complicated algorithms that take account of events far back in the past, so $m$ should be in the order of a few units. Consequently, $2^m/N$ is likely to be very small.

This suggests that financial markets are characterised by low coordination, which implies irregular oscillations where large majorities and small minorities may appear. At the same time, financial markets are efficient in exploiting information. Thus, the observation of its time series offers few possibilities to extrapolate future courses.

**Table 13.1** Efficiency of coordination and efficiency of information exploitation in the minority game

| $2^m/N < 0.1$ | $2^m/N > 1$ |
|---|---|
| Inefficient coordination | Efficient coordination |
| Low $e_c$ | High $e_c$ |
| Efficient information exploitation | Inefficient information exploitation |
| High $e_i$ | Low $e_i$ |

## 13.4 Some Pitfalls of Utility Maximization

Utility maximization strikes its adepts for its elegance, simplicity and beauty. Unfortunately, empirical tests have have shown that in many situations decision-makers do not follow its prescriptions. Furthermore, there are cases where maximizing utility leads to paradoxical decisions.

Some of these paradoxes can be reduced to utility maximization by means of special additions to the basic theory. Others cannot, thereby suggesting that utility maximization, besides poor descriptive strength, may have poor normative value as well. In this section the main paradoxes will be discussed, together with their eventual resolution within the utility maximization framework.

### *13.4.1 Ellsberg's Paradox and Sub-additive Probabilities*

Suppose that a decision-maker is placed in front of two urns, henceforth denoted A and B. The decision-maker is informed that urn A entails white and black balls in equal proportion, e.g., urn A may contain 10 white balls and 10 black balls. Regarding urn B, the decision-maker knows only that it entails white and black balls. Suppose to ask the decision-maker to evaluate the probability to extract a white ball from urn A and the probability to extract a white ball from urn B.

Since urn A entails white and black balls in equal proportions, the probability to extract a white ball from urn A is 0.5. On the contrary, nothing is known regarding the proportion of white to black balls in urn B. In cases like this, the so-called "principle of insufficient reason" – i.e., the fact that there is no reason to think otherwise – suggests to imagine that also urn B entails white and black balls in equal proportions. Thus, also in this case the probability to extract a white ball is assessed at 0.5. And yet, something is not in order: intuitively, urn B should be characterized by a greater uncertainty than urn A!

Ellsberg's paradox actually deals with the size of the sample on which probabilities are evaluated. In fact, Ellsberg's paradox places two extreme situations aside.

In the case of urn A, since we know that it entails white and black balls in equal proportions we are able to compute probability with infinite precision. It is just like extracting a ball (and replacing it afterwards) infinite times. We are measuring probability on a sample of infinite size.

In the case of urn B, lack of knowledge on the proportion of white to black balls is equivalent to estimating the probability of extracting a white ball prior to any extraction. It means that the probability must be measured on a sample of size zero. We guess its value at 0.5, but the reliability of our estimate is very low.

One possibility for overcoming Ellsberg's paradox is that of representing uncertainty by means of two magnitudes. The first one is probability, whereas the second one is sample size. In statistics, sample size is expressed by *precision indicators*.

Another possibility is to resort to the theory of sub-additive probabilities. While according to classical probability theory the sum of the probabilities of an exhaustive set of events must be equal to 1, according to the theory of sub-additive probabilities this holds only if probabilities are measured on a sample of infinite size. In all other cases, probabilities take values such that their sum is smaller than 1.

Let us consider the following example: We are playing dice in a clandestine gambling room. Since we fear that we are playing with an unfair die, we may not assign probability 1/6 to each face, but rather less, e.g. 1/8. Thus, the sum of the probabilities of all faces is $6 \times 1/8 = 3/4$, which is smaller than 1. Subsequently, if we have a possibility to throw the die many times –i.e. if we can increase the size of our sample –we may find out that the die is unfair in the sense that, e.g., face "2" comes out with probability 1/3 while the other faces come out with probability 2/15. The sum of all these probabilities is $5 \times 2/15 + 1/3 = 2/3 + 1/3 = 1$.

Let us return to Ellsberg's paradox. In the case of urn A, the probability to extract a white ball is 0.5 and the probability to draw a black ball is 0.5. The sum of these probabilities is 1. In the case of urn B, the decision-maker may judge that the probability to extract a white ball is, for instance, 0.4, and that the probability of extracting a black ball is also 0.4. The sum of these probabilities is 0.8, but this does not constitute a problem for the theory.

By employing sub-additive probabilities, utility maximization can be meaningfully applied to situations where probabilities are less-then perfectly known. Thus, utility maximization can safely deal with the difficulty raised by Ellsberg's paradox.

### *13.4.2 Allais' Paradox and Prospect Theory*

The following experiment was proposed by Maurice Allais. Subjects are asked to choose between the alternatives A and B reported on the rows of Table 13.2. It is empirically observed that most people choose alternative B.

**Table 13.2** The first choice in Allais' experiment

| | Consequence 1 | Consequence 2 | Consequence 3 |
|---|---|---|---|
| Alternative A | Receive \$2,500 with prob. 0.33 | Receive \$2,400 with prob. 0.66 | Receive nothing with prob. 0.01 |
| Alternative B | Receive \$2,400 with prob. 1.00 | | |

**Table 13.3** The second choice in Allais' experiment

| | Consequence 1 | Consequence 2 |
|---|---|---|
| Alternative C | Receive \$2,500 with probability 0.33 | Receive nothing with probability 0.67 |
| Alternative D | Receive \$2,400 with probability 0.34 | Receive nothing with probability 0.66 |

Subsequently, the same subjects are confronted with the alternatives C and D reported on the rows of Table 13.3. It is empirically observed that most people choose alternative C.

Let us now examine the expected utilities of these two pairs of alternatives, (A,B) and (C,D). Preferring (B) to (A) means that $u(2{,}400) > 0.33 \times u(2{,}500) + 0.66 \times u(2{,}400)$, which can be written as $0.34 \times u(2{,}400) > 0.33 \times u(2{,}500)$. Unfortunately, preferring (C) to (D) implies the opposite, i.e. that $0.33 \times u(2{,}500) > 0.34 \times u(2{,}400)$. So it turns out that most people either do not behave rationally, or do not maximise utility.

Allais' paradox is due to the presence of a tiny probability of not obtaining anything in alternative (A). Thus, it is due to aversion to risk.

Daniel Kahneman and Amos Tversky introduced non-linear transformations of utilities and probabilities in order to balance risk aversion. The transformed utilities and probabilities can describe the observed behaviour as expected utility maximization. This is called Prospect Theory.

A prospect is a set of pairs $\{(c_1, p_1), (c_2, p_2), \ldots\}$, where $c_j$ is a consequence that will obtain with probability $p_j$. As a preliminary step, prospects with identical consequences are summed, dominated prospects are eliminated and riskless components are ignored.

Prospects Theory prescribes that the utilities and the probabilities of the above prospects be transformed according to the following rules:

1. Utility is transformed by means of a non-linear function $v = f(u)$ such that $f'(u) > 0$ and $f''(u) < 0$ for $u > 0$, $f'(u) > 0$ and $f''(u) > 0$ for $u < 0$, with $|f''(u)|_{u<0} > |f''(u)|_{u>0}$.
2. Probabilities $p$ are transformed into "weights" $w$ by means of a non-linear function $w = g(p)$ such that $g(0) = 0$ and $g(1) = 1$ but $\exists\ \overline{p} \in (0, 1)$ such that $\forall p < \overline{p}$ it is $g(p) \geq p$ and $\forall p > \overline{p}$ it is $g(p) \leq p$.

3. Weights $w$ are transformed into coefficients $q$ by means of the following rules:

| | | | | |
|---|---|---|---|---|
| $q_{-h}^-$ | $=$ | $w^-(p_{-h})$ | for | $j=-h$ |
| $q_i^-$ | $=$ | $w^-(p_{-h}+\ldots+p_i)-w^-(p_{-h}{}^+\ldots{}^+p_{i-1})$ | for | $-h<j\le 0$ |
| $q_i^+$ | $=$ | $w^+(p_i{}^+\ldots+p_k)-w^+(p_{i+1}{}^+\ldots{}^+p_k)$ | for | $0\le j<k$ |
| $q_k^+$ | $=$ | $w^+(p_k)$ | for | $j=k$ |

where $w^-$ and $q^-$ refer to prospects with negative utility, denoted with an index $j\in[-h, 0]$, whereas $w^+$ and $q^+$ refer to prospects with positive utility, denoted with an index $j\in[0, k]$.

The $v$ and $q$ obtained at the end of this procedure can be used much like utilities and probabilities, respectively. Prospect Theory succeeds to eliminate the inconsistencies highlighted by Allais' paradox, but it does not explain why it works. It should be called a heuristic, rather than a theory. In the end, utility maximisation does succeed to cope with Allais' paradox, but at the cost of a patch that has a flavour akin to the epicycles that had to be added to the Ptolemaic system in order to support the idea that it was the Sun that was turning around the Earth.

### *13.4.3 Preference Reversal in Slovic's Paradox*

Let us consider a series of bets with different characteristics. For instance, a series of bets on different horses, or playing on a series of different slot machines, or a series of unfair dice different from one another. The game consists of choosing to bet on a specific horse, choosing to play on a specific slot machine or selecting a specific die to throw. In other words, the game consists of choosing one bet out of a series of bets.

In order to simplify matters, let us consider series composed by two bets. More specifically, let us consider the four pairs of bets in Table 13.4.

For any pair of bets, subjects are asked to select either bet A or bet B. On average, the number of subjects who prefer A to B is slightly greater than the number of subjects who prefer B to A.

At this point, a different game is played. Subjects are asked to imagine that they own a lottery ticket for each bet, and that they have a possibility to sell it. That is, they can either wait for the outcome of each bet, where they may win or loose with a certain probability, or they can sell the ticket. In order to compare the willingness to play to the willingness to sell the ticket, subjects are asked to fix a minimum selling price for each bet.

In general, it is empirically observed that most people ask a higher price for bets B than for bets A.

However, for each pair of bets, bet A has the same expected (utility) value than bet B. Thus, utility maximizers should be indifferent between A and B. On the contrary, most subjects have a slight preference for A if they are asked to play one of the two bets but they definitely prefer B if they are asked to fix a selling price.

**Table 13.4** Slovic's experiment

| | | |
|---|---|---|
| Pair of bets I | | |
| | **Consequence 1** | **Consequence 2** |
| **Bet** $\mathbf{A_{I}}$ | Win $4.00 | Loose $1.00 |
| | With probability 0.99 | With probability 0.01 |
| **Bet** $\mathbf{B_{I}}$ | Win $16.00 | Loose $2.00 |
| | With probability 0.33 | With probability 0.67 |
| Pair of bets II | | |
| | **Consequence 1** | **Consequence 2** |
| **Bet** $\mathbf{A_{II}}$ | Win $3.00 | Loose $2.00 |
| | With probability 0.95 | With probability 0.05 |
| | Win $6.50 | Loose $1.00 |
| | With probability 0.50 | With probability 0.50 |
| Pair of bets III | | |
| | **Consequence 1** | **Consequence 2** |
| **Bet** $\mathbf{A_{III}}$ | Win $2.00 | Loose $1.00 |
| | With probability 0.80 | With probability 0.20 |
| **Bet** $\mathbf{B_{III}}$ | Win $9.00 | Loose $0.50 |
| | With probability 0.20 | With probability 0.80 |
| Pair of bets IV | | |
| | **Consequence 1** | **Consequence 2** |
| **Bet** $\mathbf{A_{IV}}$ | Win $4.00 | Loose $0.50 |
| | With probability 0.80 | With probability 0.20 |
| **Bet** $B_{IV}$ | Win $40.00 | Loose $1.00 |
| | With probability 0.10 | With probability 0.90 |

The distinguishing feature of bets A is that the first consequence has a much higher probability than the second one. Thus, one may assume that it is this difference of probability values that orientates decision-making.

The distinguishing feature of bets B is that the first consequence concerns a much larger amount of money than the second one. Probabilities, on the contrary, are sometimes very similar, sometimes very different from one another. Thus, one may assume that it is this difference of money values that orientates decision-making.

If subjects are asked to bet their attention is caught by probabilities, so either they are indifferent or they prefer A. If subjects are asked to sell lottery tickets their attention is caught by money values, so they prefer B.

Slovic's paradox shows that preferences change if the decision-maker focuses on the probability of a consequence or, rather, on its utility (here, money value). This means that human beings are unable to evaluate probabilities and utilities independently of one another.

Slovic's paradox – often known as "preference reversal" – is destructive for utility maximization. In fact, it undermines the assumption that a utility function and a probability function can be defined, independently of one another. Ultimately, Slovic's paradox suggests that uncertain belief cannot be split into utilities and probabilities.

Obviously, several attempts to reconcile preference reversal with the theory of rational choice have been made. Preference reversal can be accommodated with the theory of rational choice if either violations of transitivity, or of independence, or of completeness of preferences are accepted. While the attempts to reconcile preference reversal with the theory of rational decision by relaxing transitivity or independence of preferences did not receive much attention because these properties are essential for our idea of rationality – see Sect. 13.2, the more recent idea of dropping completeness deserves some discussion. In fact, allowing preferences to be incomplete amounts to accept the idea that a utility function can be defined, at most, for only *some* alternatives. Possibly, just the simplest and most repetitive ones.

### *13.4.4 Arrow's Paradox*

The following paradox of social choice is due to Kenneth Arrow. Let A, B and C denote three alternatives, and let 1, 2 and 3 denote three individuals. Let us assume that:

- Individual 1 prefers alternative A to alternative B and alternative B to alternative C. Thus, he prefers alternative A to alternative C.
- Individual 2 prefers alternative B to alternative C and alternative C to alternative A. Thus, he prefers alternative B to alternative A.
- Individual 3 prefers alternative C to alternative A and alternative A to alternative B. Thus, he prefers alternative C to alternative B.

If these three individuals constitute a democratic community with a majority rule, then this community prefers A to B (individuals 1 and 3) and alternative B to alternative C (individuals 1 and 2). Thus, if the community wants to have transitive preferences, it must prefer A to C. But, the majority of its members (individuals 2 and 3) prefers C to A!

Arrow's paradox shows that there are conditions such that the aggregate outcome contradicts a basic assumption of utility maximization, even if individuals do not. It is not destructive for utility maximization as a theory of individual decision-making, but it impairs its extension to group or organizational decision-making.

Several proposals have been made in order to overcome Arrow's paradox. The most common way out is to allow individuals to have different preferences if all alternatives are presented to them, instead of being presented with pairs of alternatives. Or, one may limit voters to two alternatives presented in tournaments. In this way, Arrow's paradox would disappear, yet the final choice is not necessarily the one that would be preferred by the largest possible majority.

## 13.5 Logic of Consequence and Logic of Appropriateness

As we have seen in Sect. 4, utility maximization is not a good descriptor of decision processes. Its proponents have objected that utility maximization is not meant to be a faithful description of what people actually do, but rather a prescription of what they should do. It pretends to be a normative theory, although it is not a descriptive theory.

However, the preference reversals highlighted by Slovic point to such a huge distance between theory and reality, that the normativeness of utility maximization might be questioned. If utilities do not exist prior to decision-making, it may make little sense to tell decision-makers that they should maximize them. Furthermore, if evolution shaped human reasoning along patterns that are different from utility maximization, we ought to be careful to declare these patterns “illogical”, or “irrational”. Rather, it may make sense to observe how human beings actually make their decisions, understand their rationales, and eventually revise our theories.

Sociology knows a distinction between the “logic of consequences” that underlies utility maximization, and a “logic of appropriateness” where human beings behave according to what they deem appropriate depending on past experiences and social pressures to conformity in specific settings. Note that the logic of appropriateness does not separate an individualistic step (Utility) from social interaction (Game Theory).

According to the logic of appropriateness, human minds are viewed as coherence-seeking machines that make use of available information in order to construct a plausible interpretation of reality, be it social roles, scientific theories, or else. By drawing causal relationships and eliminating inconsistencies a decision-maker tells herself a story that explains why certain facts are the way they are and why certain people did what they did. This story, a founding story that suggests a decision-maker what it is appropriate to do, is called a *narrative*.

The construction of a narrative may require that issues that do not fit into the picture are ignored, downplayed or forgotten. It may require that opinions are changed even dramatically, and yet their purporters candidly claim that they have always been coherent throughout their lives, or that they have been coherent in spite of having changed their opinion, if their story is seen from a particular point of view.

Albeit disturbing for our idea of rationality, the extent and easiness with which human beings distort previous experiences is proven by a number of experiments in psychology. It is easy to induce the subjects of experiments to change opinion while they are still convinced to have been coherent throughout the whole experiment. Experiments show that people remember past events to the extent that they fit their narratives, and that they are ready to change their interpretation of the past if new evidence must be accommodated. On the whole, experimental evidence tells us that human beings are ready to lie to themselves in order to build coherent narratives.

This attitude is puzzling, because distorting reality in order to construct a coherent narrative is at odds with our idea of rationality. So either human nature is inherently irrational, or our idea of rationality is incorrect.

According to James March, re-inventing the past is a crucial ability that enables decision-makers to conceive new goals and figure out a strategy in an uncertain future. Later, a similar argument has been made by Karl Weick under the label of "sensemaking". Essentially, these authors suggest that in order to make decisions in the face of an uncertain future it is good to have a narrative that explains the past as if previous decisions had been made along a coherent line. This line guides the decision-maker into the future, providing a rationale for action even if certainties are very few.

So here comes a straightforward argument for normativeness. If seeking coherence has the purpose of constructing a narrative, and if narratives are useful, then a decision theory based on constructing narratives should be regarded as rational, and openly prescribed.

In business, politics and other fields, narratives may constitute the bulk of strategies. David Lane and Robert Maxfield have made a years-long field observation of the elaboration and modification of the narrative of a Silicon Valley firm. This study is worth reporting, because it is very clear in making us understand that narratives are useful precisely because they provide a guidance in the face of an uncertain future, and that their usefulness is not impaired by the fact that their coherence is based on an arbitrary interpretation of reality.

### *13.5.1 A Real Story*

In 1990, *Echelon*, a Silicon Valley company, launched *LonWorks*, an innovative technology for distributed control. Previously, control was centralized by one main processing unit that commanded several peripheral devices. With *LonWorks*, each device is endowed with a microprocessor and can communicate with all others, so the peripheral electromechanical devices control each other. Distributed control is more resilient than centralized control, and easily implements modular architectures to which additional devices can be added.

Distributed control is particularly suited to the automation of office spaces in large buildings, post-Fordist productive plants, as well as any setting where a large number of heterogeneous devices must coordinate their operations while retaining some flexibility. Thus, in its early days *Echelon* focused on partnerships with large producers of the devices to be automated, e.g., a producer in the field of heating and air conditioning was offered a possibility to integrate a microchip in their devices, as well on lifts, doors and windows in order to integrate all controls in a large building, from lighting to heating to theft protection.

With some disappointment, *Echelon* had to recognize that the *LonWorks* technology was not exploited in its full potentialities. In fact, each large producer was specialized in one tiny sector so it had neither the power nor the capability to implement *LonWorks* on all devices. For instance, a producer in the field of heating and air conditioning found it difficult to install *LonWorks* on doors, windows, lights and lifts, for the production of these devices was covered by other firms. Indeed, the difficulty was that *Echelon* was attempting to create a single market for automation where the marketplace was covered by producers of several devices.

*Echelon* was conscious of the enormous difficulties connected with the creation of a new market. Nevertheless, it deemed that long-term relations with a few specialized producers would pay in the long run. *Echelon* had a narrative, saying that large specialized producers would slowly but persistently adopt and impose *LonWorks*. Consequently, it invested all of its resources in these relations.

By 1994, *Echelon* was loosing confidence in this narrative. *Echelon* started to approach large system integrators of ICT, such as *Olivetti* and *Ameritech*. However, the crucial move was that of hiring a person for this job, who did not come from Silicon Valley as all other executives did. Through this employee *Echelon* approached smaller companies, that integrated devices they bought from different producers. Some people at *Echelon* conceived the idea of embedding *LonWorks* in a box that could be attached to any electromechanical device, of whatever producer.

Scholars of technological innovation know how difficult it is for visionary employees to convince their boss of the value of their idea. In the case of *Echelon* the CEO embraced enthusiastically the new idea, because it appeared to fit with his previous experience.

*Echelon*'s CEO had been the successful entrepreneur of a small firm that exploited digital technologies to produce private branch exchange systems (PBX) with innovative features. This firm had been able to displace giants such as *AT&T* by providing small independent installers with a superior product. When this CEO met small independent integrators of electromechanical devices, he mapped the idea proposed by his new employee onto his previous experience.

In 1996, and within a few months, *Echelon* changed its narrative. *Echelon* presented itself as a provider of an innovative microchip for independent system integrators, a microchip that could be installed on any electromechanical device, of whatever producer.

Most importantly, *Echelon* told itself that it had always pursued this strategy. Nobody in the firm seemed to be aware that the firm's strategy had changed. According to the narrative that they had developed, they had always done what they were doing.

Moreover, when faced with evidence that the firm did change its strategy, management wished that the final publication would not stress this aspect (Lane: personal communication). This makes sense, for according to our idea of rationality narratives should reflect "objective information", and decision-makers should stick to it. Thus, management did not want to appear irrational according to common wisdom.

However, the case of *Echelon* highlights that constructing a narrative by re-interpreting the past may be good and useful for decision-makers. In fact, the reported case reveals that precisely by re-interpreting its mission *Echelon* was able to direct its investments. If the future is uncertain, as it is often the case, interpreting the past in order to find a direction for the future is a sensible activity.

So the trouble may rather lie with our idea of rationality. Since re-interpreting the past is regarded as irrational, then it must be done in secrecy. However, if re-interpreting the past may have positive effects, then it should be prescribed.

## 13.6 Tools for the Logic of Appropriateness

Although the logic of appropriateness cannot propose itself with a ready-made and ready-to-use formula such as utility maximization, there exist some tools that can be used to reproduce its building blocks. These are essentially classification tools, i.e., tools that form concepts out of information, and coherence tools, i.e., tools that arrange concepts into coherent stories.

In particular, the following tools will be reviewed in this section:

1. Unsupervised neural networks;
2. Constraint Satisfaction Networks;
3. Evidence Theory, also called Belief Functions Theory.

Unsupervised neural networks reproduce the formation of mental categories out of a flow of information. Constrain Satisfaction Networks arrange concepts into coherent explanations. Finally, Evidence Theory assumes that an actor receives information on possibilities and arranges them into coherent hypotheses. Although these tools have not been integrated into one another, they all concern the process of selecting some items from the flow of experiences, arranging them in a coherent narrative, and deciding accordingly.

The logic of consequence makes sense in the restricted realm of games of chance, where it is possible to overview an exhaustive set of possibilities and enlist all of the consequences of any alternative. On the contrary, the logic of appropriateness makes sense precisely because, quite often, such conditions do not hold. Thus, this review does not cover tools concerned with classification in a *given* set of categories, such as Case-Based Decision Theory and supervised neural networks.

### *13.6.1 Unsupervised Neural Networks*

Human mental categories are not defined by pre-specified similarity criteria that the objects to be classified should fulfil. Rather, mental categories are continuously constructed and modified according to the similarity of a just-received piece of information to the pieces of information that have already been stored in existing categories. For instance, a child observing house chairs may start with an idea of "chair" as an object having four legs, then observe an office chair with only one leg and yet sufficiently similar to house chairs to be added to their category, which from this time on does not have the number of legs as a common property of all the objects that it entails.

This is a radically different process from that of defining criteria in order to classify objects into given categories. In the case of mental categories, categories do not exist prior to the beginning of the classification process. Categories form out of similarity of certain input information to some information that has been previously received, so these items are stored in the same "place" and concur to build up a mental category. There exist no criteria defined *ex-ante* that control the classification of input information; since the only rule is similarity of information items, categories form depending on what items are received. For instance, the aforementioned mental category "chair" may depend on what chairs have been observed in the course of a life spent in Manhattan, rural China or tropical Africa, in modern times or the Middle Age.

By contrast, logical reasoning works with objects that have been clearly defined. The point here is that these definitions are made *once* mental categories have been formed. Eventually, a mental category formed around similarity judgements may suggest a definition for all the objects that it entails if they share some common feature.

However, this is not necessarily the case. There are instances where definitions are not possible, simply because a mental category entails objects that do not have any common feature. For instance, the mental category expressed by the word "game" refers to children amusing themselves with toys, adults involved in a serious competition on a chess board and as well as a set of wild animals. One may speculate that man transposed the emotions involved in hunting into the more intellectual context of chess, and that the fact that chess was an amusement suggested some similarity to what children were supposed to do. So pairwise intersections of the meanings of the word "game" do exist, but this does not imply that all of their meanings have a common intersection. Nevertheless, human beings are perfectly at ease with this as well as many other concepts that cannot be defined.

Unsupervised neural networks (UNN) are able to reproduce the idea that mental categories arise out of adding examples. In fact, these networks construct categories around the most frequent input patterns, mimicking the idea that a child creates a category "chair" upon observation of many objects similar to one another.

It is important to stress once again that the ensuing account deals with *unsupervised* neural networks (UNN) only. Supervised neural networks (SNN), with their neurons arranged in layers and training phases to teach them in what categories they must classify input information, do not fit into the present account. Nor does Case-Based Decision Theory which, similarly to SNN, is concerned with classifying information into previously defined categories. Directions to these tools will be provided in the bibliography, but they do not pertain to this chapter.

Neural networks, of whatever kind, are composed by a set of *neurons* which produce an output $y \in \Re$ by summing inputs $x_1, x_2, \ldots x_N \in \Re$ by means of coefficients $a_1, a_2, \ldots a_N$:

$$y = \sum_{i=1}^{N} a_i x_i \tag{13.7}$$

For any set of coefficients $a_i$, this simple device is able to classify inputs in a category by yielding the same output $y$ for several input vectors $\mathbf{x}$. In fact, there exist several vectors $\mathbf{x}$ whose weighted sum yields the same $y$. For instance, if $\forall i$ it is $a_i = 1$, then e.g. $y = 10$ can arise out of $\mathbf{x}' = [9 \quad 1]$, $\mathbf{x}'' = [2.5 \quad 7.5]$, as well as many other vectors. In this sense, the neuron classifies the input vectors [9 1] and [2.5 7.5] in the same category.

Note that a neuron has no difficulty to classify input vectors that do not perfectly fit its categories. For instance, if there is a category $y = 10$ and a category $y = 11$, an input vector $\mathbf{x}''' = [2.1 \quad 8]$ is classified in the category $y = 10$ just as $\mathbf{x}'$ and $\mathbf{x}''$.

The shape of the categories implemented by a neuron depends on the coefficients $a_i$. For instance, if $a_1 = 0.5$ and $a_2 = 20$ the input vector $\mathbf{x}' = [9 \quad 1]$ yields $y = 24.5$ and may not lie in the same category as $\mathbf{x}'' = [2.5 \quad 7.5]$, which yields $y = 151.25$.

The coefficients $a_i$ may be chosen by the user of the network during a training phase, in which case we are dealing with a SNN. Alternatively, the coefficients $a_i$ may be initialised at random and subsequently changed by the network itself according to some endogenous mechanism. In this case we have a UNN, of which Kohonen networks are the most common instance.

In UNN, the ability of a neuron to change its categories stems from a feed-back from output $y$ and a feed-forward from input $\mathbf{x}$, towards coefficients $a_i$:

$$\frac{da_i}{dt} = \varphi(\mathbf{a}, y)x_i - \gamma(\mathbf{a}, y)a_i \forall i \tag{13.8}$$

where $\varphi(\mathbf{a},y)$ and $\gamma(\mathbf{a},y)$ may be linear or non-linear functions.

In Eq. 13.8, the term $\varphi(\mathbf{a}, y)x_i$ enables the neuron to learn input patterns. It entails both a feed-back (from $y$) and a feed-forward (from $x_i$). This *learning term* makes $a_i$ increase when *both* $y$ and $x_i$ take high values, thereby enhancing those coefficients that happened to yield a high $y$ when a particular $x_i$ was high. Thus, the structure of coefficients vector $\mathbf{a}$ ultimately depends on which vectors $\mathbf{x}$ appeared most often as inputs.

The learning term is such that the neuron learns the patterns that it receives most often. This is sufficient to make the network work, but makes it unable to construct different categories if different patterns appear. Furthermore, since the learning term works by multiplying inputs and outputs, it may produce an explosive output. This should be curbed in order to use the network.

For both reasons, a *forgetting term* that makes the coefficients $a_i$ decay towards zero is in order. In Eq. 13.8 the forgetting term is $\gamma(\mathbf{a}, y)a_i$. It depends on a feed-back from output $y$ and, most importantly, on coefficient $a_i$ itself.

Figure 13.8 illustrates the feed-backs and -forwards within a neuron of a UNN.

Simple, but non trivial examples of Eq. 13.8 are: $\dot{\mathbf{a}} = \mu y\mathbf{x} - \nu\mathbf{a}$, $\dot{\mathbf{a}} = \mu\mathbf{x} - \nu y\mathbf{a}$, $\dot{\mathbf{a}} = \mu y\mathbf{x} - \nu y\mathbf{a}$, $\dot{\mathbf{a}} = \mu y\mathbf{x} - \nu y^2\mathbf{a}$, where $\mu$ and $\nu$ are constants. Each functional form corresponds to different strengths of the learning and forgetting terms.

In general, a network of neurons is able to discriminate input information according to much finer categories than a single neuron can do. As a rule, the greater the number of neurons, the finer the categories that a network constructs. However, a neural network is useful precisely because it is able to classify a huge amount of information into a few broad categories. If categories are so fine that they track input information exactly, then a neural net becomes useless. Thus, the number of neurons

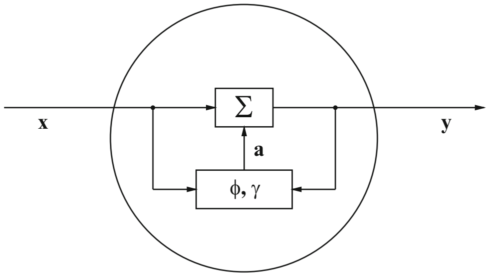



**Fig. 13.8** The neuron of a UNN. The feed-backs and -forwards are responsible for the most notable properties of UNN, including the absence of a training phase. In a sense, the "training phase" of SNN may be seen as a feed-back and -forward passing through a human operator

that a network should possess depends on the variability of the input as well as on user needs.

However, the behaviour of a neural network does not only depend on the number of its neurons, but also on the structure of the connections between them. In fact, just like the capabilities of neurons depend on feed-backs and -forwards, the capabilities of a neural network depend on linkages that eventually enable information to circulate in loops. If information can circulate within the network, then the whole network acquires a *memory*.

It is a *distributed* memory, fundamentally different in nature from the more usual *localised* memories. Localised memories such as books, disks, tapes etc., store information at a particular point in space. This information can only be retrieved if one knows where its support is (e.g. the position of a book in a library, or the address of a memory cell on a hard disk).

On the contrary, in a neural network each neuron may be part of a number of information circuits where information is "memorised" as long as it does not stop to circulate. Although this *is* a memory, one cannot say that information is stored at any particular place. Hence the name.

For obvious reasons, the information stored in a distributed memory cannot be retrieved by means of an address. However, a piece of information flowing in a particular loop can be retrieved by some other piece of information that is flowing close enough to it. Thus, in a distributed memory information can be retrieved by means of *associations* of concepts, with a procedure that reminds of human "intuition". Intuition, according to this interpretation, would consist of associations between concepts that occur when information flowing in different but neighbouring circuits comes in touch.

### *13.6.2 Constraint Satisfaction Networks*

Parallel constraint satisfaction networks (CSN) arrange concepts into coherent theories. Although they belong to the larger family of neural networks, they do not carry out any classification process.

Constraint satisfaction networks are characterized by:

- Excitatory and inhibitory connections;
- Feedbacks between neurons.

Neurons represent possibilities, or concepts, or propositions. Connections represent inferences: an excitatory connection from neuron A to neuron B means “A implies B”, whereas an inhibitory connection from neuron A to neuron B means “A implies ¬ B”.

Let $a_i$ denote the activation (the output) of neuron $i$, with $a_i \in \Re$. Let $w_{ij} \in \Re$ denote the weight by which neuron $i$ multiplies the input arriving from neuron $j$.

The net excitatory input to neuron $i$ is:

$$ene\,t_i = \sum_j w_{ij} a_j \quad if \;\; w_{ij} a_j \geq 0 \tag{13.9}$$

The net inhibitory input to neuron $i$ is:

$$ene\,t_i = \sum_j w_{ij} a_j \quad if \;\; w_{ij} a_j < 0 \tag{13.10}$$

At each time step, the activation of neuron $i$ is increased by its excitatory inputs and decreased by its inhibitory inputs:

$$\Delta a_i = enet_i(a_{max} - a_i) + inet_i(a_i - a_{min}) \tag{13.11}$$

where, in general, $a_{max} = 1$ and $a_{min} = -1$.

Feedbacks between neurons make the network maximize consonance:

$$C = \sum_i \sum_j w_{ij} a_i a_j \tag{13.12}$$

or, equivalently, minimize energy $E = -C$.

Consonance maximization means that those neurons are strengthened, that represent possibilities, concepts or propositions that are coherent with one another. Thus, constraint satisfaction networks can be used to model any cognitive process characterized by a search for coherence. In particular, researchers have emphasized the ability of CSN to construct narratives, much like humans actually do.

Notable applications of CSN are the elaboration of scientific theories, which amounts to arrange empirical findings in a network of coherent causal relations, as well as the evaluation of guilt or innocence in a trial, which amounts to fitting testimonies in a coherent frame. Under this respect, CSN share a common concern with Evidence Theory or, to be more precise, with its Dempster-Shafer’s combination rule reported in Eq. 13.16 of Sect. 13.6.3.

When modelling decision-making, CSN can be used to model the process of emphasizing the positive aspects of one alternative and the negative aspects of its competiting alternatives until a coherent frame is available and a decision can be made. This oscillation between competing explanations reproduce at least one important aspect of *Gestalt* theories, namely, the idea that the human mind may shift among alternative interpretations of reality, as exemplified by Rubin's vase and other images where at least two interpretations are possible.[4] Many cues suggest that this is a fundamental pattern in decision-making.

A clear limitation of CSN is that they work with given possibilities, concepts, or propositions. In other words, CSN can reproduce the arrangement of possibilities and concepts into narratives, but not their arousal. By contrast, UNN reproduce the arousal of concepts out of empirical experiences. Possibly, future models will be able to couple UNN to CSN in order to model both processes at a time.

### *13.6.3 Evidence Theory*

Evidence Theory is a branch of the mathematics of uncertain reasoning that, unlike Probability Theory, does not assume that a decision-maker knows the set of all possible events. Rather than defining a "residual event" for anything that cannot be clearly expressed, Evidence Theory leaves a decision-maker's possibility set open to novelties.

Evidence Theory makes use of a particular class of monotone uncertainty measures, Choquet capacities of infinite order. Furthermore, it assumes that no operation is attached to the possibility set, which frees a decision-maker to define a "residual event" by complementation. Novel possibilities can appear in the possibility set in the course of the calculations, and the possibility set is called *frame of discernment* in order to stress its cognitive nature.

Evidence Theory does not take a gambler as its prototypical subject, but a judge or a detective. The reason is that a gambler playing with dice or throwing a coin knows what possibilities can occur. On the contrary, judges and detectives know that unexpected proves and testimonies may open up novel possibilities. Possibly, managers making investments, politicians steering their countries, or just anyone in the important choices of her daily life is more akin to a judge or a detective looking for cues than a gambler looking for luck.

Let us consider a frame of discernment $\Theta$. Let us suppose that a person receives testimonies, or *bodies of evidence*, as numbers that to various extents support a set of possibilities $A_1, A_2, \ldots A_N$, where $A_1 \subseteq \Theta$, $A_2 \subseteq \Theta \ldots A_N \subseteq \Theta$ and where the

[4] The simplest picture of this kind is a cube depicted by its edges: it is up to the observer to choose which face stays in the front and which face stays in the rear. Rubin's vase is white and stands against a black background. The observer may see a white vase, or two black profiles in front of one another.

$A_i$s are not necessarily disjoint sets.[5] Let us denote these numbers $\{m(A_1), m(A_2), \ldots m(\Theta)\}$, where $m(A_i)$ measures the amount of empirical evidence that supports the possibility $A_i$.

Numbers $m$ are exogenous to the person (the judge, the detective) who owns the frame of discernment. They are not subjective measures for this person, though they may be subjective evaluations of those who provide the testimonies. Numbers $m$ are cardinal measures of the amount of empirical evidence supporting each possibility.

Since no operation is defined on the frame of discernment, the number $m$ that has been assigned to $\Theta$ does not concern any specific possibility. Rather, it indicates how small the evidence is, that supports the possibilities envisaged in the testimony, or, in other words, how strongly a person fears that the possibilities that she is envisaging are not exhaustive. The greater the ignorance of a person on which possibilities exist, the greater $m(\Theta)$.

Note that $m(\Theta)$ can be smaller than any $m$ of the $A_i$s that it entails. Indeed, this applies to the $A_i$s as well: if $A_i \supset A_j$, this does *not* imply that $m(A_i) > m(A_j)$.

Although it is not strictly essential for Evidence Theory, numbers $m$ are generally normalised by requiring that:

$$\sum_{i=1}^{N} m(A_i) + m(\Theta) = 1 \tag{13.13}$$

For instance, if the original format of the testimony is:

$$\{5, 32, 12, 3\}$$

by applying Eq. 13.13 we obtain:

$$\{0.096, 0.615, 0.231, 0.058\}$$

whose numbers sum up to one.

Let us suppose that the decision-maker wants to evaluate to what extent the available empirical evidence supports certain hypotheses that she is entertaining in her mind. Since a hypothesis concerns the truth of a possibility or a set of possibilities, hypotheses are subsets of the frame of discernment just as possibilities are. A body of evidence $\{m(A_1), m(A_2), \ldots m(\Theta)\}$ supports a hypothesis $H$ to the extent that some $A_i$s are included or at least intersect $H$.

Note that, whilst the possibilities $A_i$ entailed in the testimonies cannot be combined with one another (intersected, complemented, etc.) to form novel possibilities, a hypothesis $H$ represents a construct of the owner of a frame of discernment (the judge, the detective, etc.). This person is absolutely free to conceive any

[5] For simplicity, the theory is expounded with respect to a finite number of possibilities. No substantial change is needed if an infinite number of possibilities is considered.

hypothesis, as well as its opposite. Thus, although $\overline{A}_i$s are forbidden, $\overline{H}$ can be safely considered.

Given a testimony $\{m(A_1), m(A_2), \ldots m(\Theta)\}$, the belief in hypothesis $H$ is expressed by the following *belief function*:

$$Bel(\mathrm{H}) = \sum_{\mathrm{A_i} \subseteq \mathrm{H}} \mathrm{m}(\mathrm{A_i}) \tag{13.14}$$

By definition, $Bel(\emptyset) = 0$ and $Bel(\Theta) = 1$. However, this last condition does not imply that any of the possibilities included in the frame of discernment must necessarily realise. It simply means that any possibility must be conceived within the frame of discernment, independently of what possibilities are envisaged at a certain point in time.

The belief function takes account of all evidence included in $H$. The *plausibility function* takes account of all evidence that intersects $H$:

$$Pl(H) = \sum_{H \cap A_i \neq \emptyset} m(A_i) \tag{13.15}$$

It can be shown that belief and plausibility are linked by the relation $Pl(H) = 1 - Bel(\overline{H})$, where $\overline{H}$ denotes the hypothesis opposite to $H$. If $m(\Theta) > 0$ these two measures are not equivalent, so both of them need to be considered. In general, $Bel(H) \leq Pl(H)$.

Let us suppose that some unexpected facts occur, that are told by a new testimony. The new testimony must be combined with previous knowledge, confirming it to the extent that it is coherent with it. On the contrary, previous beliefs must be weakened if the new evidence disconfirms them.

Let $\{m(B_1), m(B_2), \ldots m(\Theta)\}$ be the new testimony, which must be combined with $\{m(A_1), m(A_2), \ldots m(\Theta)\}$. The new testimony may entail possibilities that are coherent with those of the previous testimony, possibilities that contradict those of the previous testimony, and possibilities that partially support, partially contradict the previous testimony. Figure 13.9 illustrates contradictory, coherent and partially coherent/contradictory possibilities on the frame of discernment. Contradictory possibilities appear as disjoint sets. A possibility is coherent with another if it is included in it. Finally, two possibilities that are partially coherent, partially contradictory, intersect one another.

Let us suppose that two testimonies

$$\{m(A_1), m(A_2), \ldots m(\Theta)\}$$

and

$$\{m(B_1), m(B_2), \ldots m(\Theta)\}$$

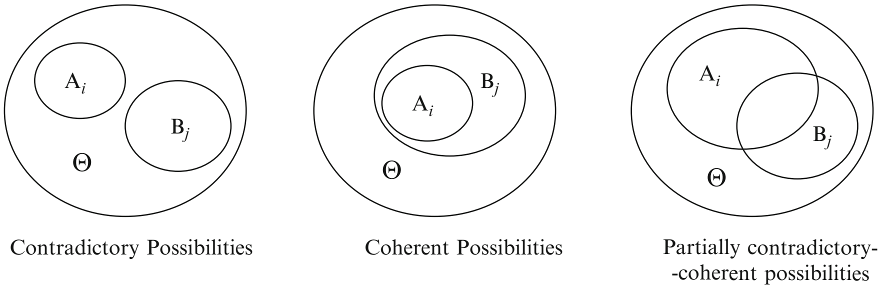



**Fig. 13.9** *Left*, two contradictory possibilities. *Centre*, two coherent possibilities. *Right*, two partially coherent, partially contradictory possibilities

that satisfy Eq. 13.13, must be combined into a testimony

$$\{m(C_1), m(C_2), \ldots m(\Theta)\}$$

that also satisfies Eq. 13.13. Dempster-Shafer's combination rule yields a combined testimony $\{m(C_k)\}$ where the coherent possibilities between $\{m(A_i)\}$ and $\{m(B_j)\}$ have been stressed.

According to Dempster-Shafer combination rule, possibilities $\{C_k\}$ are defined by all intersections of each possibility in $\{A_1, A_2, \ldots \Theta\}$ with each possibility in $\{B_1, B_2, \ldots \Theta\}$. For any possibility $C_k$, the amount of empirical evidence is:

$$m(C_k) = \frac{\sum\limits_{A_i \cap B_j = C_k} m(A_i) m(B_j)}{1 - \sum\limits_{A_i \cap B_j = \emptyset} m(A_i) m(B_j)} \tag{13.16}$$

The numerator of Eq. 13.16 measures the extent to which both the first and the second testimony support the possibility $C_k$. In fact, for each possible $C_k$ the sum extends to all pairs of possibilities from the two testimonies that are coherent on $C_k$ (see Fig. 13.9). The more the intersections between the $A_i$s and the $B_j$s that give rise to $C_k$, and the greater their amounts of evidence, the larger the numerator.

The denominator is the complement to one of those elements of the second testimony that contradict the first one. In fact, the complement to one is made on those $A_i$s and $B_j$s that are disjoint sets (see Fig. 13.9). The denominator represents a measure of the extent to which the two testimonies are coherent, in the sense that all evidence that supports contradictory possibilities is excluded.

Essentially, Dempster-Shafer combination rule says that the evidence supporting possibility $C_k$ is a fraction of the coherent evidence between $\{m(A_1), m(A_2), \ldots m(\Theta)\}$ and $\{m(B_1), m(B_2), \ldots m(\Theta)\}$. The amount of this fraction depends on the sum of all elements of the testimonies that support $C_k$.

Dempster-Shafer's rule can be iterated to combine any number of testimonies. The outcome of Dempster-Shafer combination rule is independent of the order in which two testimonies are combined.

The above description made clear that Evidence Theory provides an algorithm for handling an exogenous flow of new, unexpected possibilities. Indeed, the decision-maker of Evidence Theory is not supposed to conceive possibilities. She merely listens to exogenous testimonies that consist of possibilities and degrees of evidence supporting them, and combines these testimonies into a coherent whole by means of Dempster-Shafer theory. She does not conceive novel possibilities out of a creative effort. Rather, novel possibilities – the $\{C_k\}$ – arise out of combination of exogenous inputs.

Under this respect Evidence Theory bears some similarly to UNN. In fact, both tools are open to novel arrangements of input information. These dynamical, open arrangements may be variably called "possibilities" in Evidence Theory or "categories" in UNN, but they express a similar reasoning.

At the same time, Evidence Theory bears some similarity to CSN when it comes to the combination of evidence. In fact bot CSN and Dempster-Shafer's combination rule (13.16) seek to improve the coherence of available information.

While UNN and CSN take on the complementary tasks of modelling the arousal of new concepts and, respectively, their combination into coherent narratives, Evidence Theory is a mathematical theory that may be developed into a framework able to encompass both CSN and UNN as simulation tools. This is, at present, just pure speculation; however, it is also a possible narrative that may direct future research efforts.

## 13.7 Conclusions

This review presented tools to model decision-making according to two opposing paradigms, namely, the logic of consequence and the logic of appropriateness. The reader may feel unease because scientists do not provide a univocal answer to the demands of the modeller.

However, a pragmatic attitude may suggest that tools should be used depending on conditions. Utility maximization and Game Theory require that all available alternatives and all of their possible consequences can be listed. Thus, it may be sensible to make use of these tools when one such exhaustive list is available, eventually releasing the requirement of perfect rationality and the pursuit of Nash-equilibria while assuming some form of bounded rationality as influence games do. Unsupervised neural networks, constraint satisfaction networks and Evidence Theory, on the contrary, may be used when more challenging decision settings must be modelled. The modeller should remember that constructing narratives makes sense because decision-makers may be uncertain regarding what possibilities exist, so these tools become necessary only in this kind of decision settings.

The trouble, in this last case, is that the tools mentioned above have not been integrated into a unified framework. No simple formula is available to be readily used, so the modeller must resort to a higher degree of creativity and intuition. On the other hand, here is an exciting opportunity for modellers to participate to theory development.

## Further Reading

Game Theory is a huge subject. Relevant handbooks are Aumann and Hart (1992, 1994, 2002) and, at a more introductory level, Rasmusen (2007). However, agent-based modelers should keep in mind that a substantial part of Game Theory has been developed around equilibrium states, which are generally not a main concern for agent-based modelers. Evolutionary games, thoroughly discussed in the above handbooks, are possibly closest to agent-based modeling. For other evolutionary mechanism, see Chap. 18 in this volume (Chattoe-Brown and Edmonds 2013).

Neural networks are a huge subject as well. This field is currently split in two streams: On the one hand, research on neural networks as a model of cognitive processes in the brain. On the other hand, research on neural networks as an engineering tool for signal processing. A handbook oriented towards cognitive problems is Arbib (2002). Handbooks oriented towards engineering problems are Hu and Hwang (2002) and Graupe (2007). Specifically, unsupervised neural networks are often employed in pattern recognition. A comprehensive treatment of pattern recognition techniques is Ripley (1996).

All other tools and issues discussed in this chapter are in their infancy, so no generic reading can be mentioned. Interested scholars are better advised to start with the original papers mentioned in the bibliography, tracking developments on recent publications and working papers.

## Reasoned Bibliography

This chapter covered too many topics to be able to provide detailed references. Henceforth, a few basic publications will be listed, that may be used by interested readers as a first orientation to each of the topics mentioned in this chapter.

### *Utility and Games*

Utility maximization was pioneered by Frank Ramsay and Bruno De Finetti in the 1930s, and subsequently refined by Leonard Savage in the 1950s. Savage still provides the most comprehensive explanation of this approach to uncertain reasoning.

Game Theory was initiated by John Von Neumann and Oskar Morgenstern in the 1940s. It subsequently developed into a huge research field within economics, with several specialized journals. Today, game theory is a field characterized by extreme mathematical sophistication and intricate conceptual constructions.

This chapter did not focus on the assumptions and methods of Game Theory, but rather aimed at presenting the main prototypical games that have been devised hitherto. A classical treatise by Duncan Luce and Howard Raiffa may introduce the subject more easily than Von Neumann and Morgenstern did. Luce and Raiffa were first to present the Battle of the Sexes as well as the Prisoner's Dilemma, which they ascribed to Albert Tucker anyway.

Readers interested in evolutionary games may rather read the treatises written by Jörgen Weibull and Herbert Gintis, respectively. The former is more specific on evolutionary games, but also more technical than the second one.

Robert Axelrod is the main reference so far it regards simulations of the iterated Prisoner's Dilemma with retaliation strategies. The idea that the iterated Prisoner's Dilemma could yield cooperation simply relying on tags is due to Rick Riolo.

The Stag Hunt and the Game of Chicken are classical, somehow commonsensical games. The Game of Chicken has been turned into the Hawk-Dove Game by Maynard Smith and George Price. The Hawk-Dove game is not terribly different from The War of Attrition, conceived by Maynard Smith and improved by Timothy Bishop and Chris Cannings.

The Traveller's Dilemma and The Dollar Auction are recent games invented by Kaushik Basu and Martin Shubik, respectively. Pure Coordination games have been discovered by Thomas Schelling.

Axelrod RM (1984) The evolution of cooperation. Basic Books, New York

Basu K (1994) The traveller's dilemma: paradoxes of rationality in game theory. Am Econ Rev 84:391–395

Bishop DT, Cannings C, Smith JM (1978) The war of attrition with random rewards. J Theor Biol 74:377–389

Gintis H (2000) Game theory evolving. Princeton University Press, Princeton

Luce RD, Raiffa H (1957) Games and decision: introduction and critical survey. Wiley, New York

Riolo RL, Cohen MD, Axelrod RM (2001) Evolution of cooperation without reciprocity. Nature 414:441–443

Savage L (1954) The foundations of statistics. Wiley, New York

Schelling TC (1960) The strategy of conflict. Harvard University Press, Cambridge, MA

Shubik M (1971) The dollar auction game: a paradox in noncooperative behavior and escalation. J Conflict Resolut 15:109–111

Smith JM, Price GR (1973) The logic of animal conflict. Nature 246:15–18

Weibull J (1997) Evolutionary game theory. The MIT Press

## *Influence Games*

Ernst Ising introduced his model in the 1920s. Since then, a huge literature appeared.

The Ising model is taught in most Physics courses around the world, so a number of good introductions are available on the Internet. A printed introduction by Barry Cipra is mentioned here for completeness.

Schelling's model of racial segregation was developed independently of the Ising model. However, it may be considered a variation of it.

The *El Farol Bar Problem* was conceived by Brian Arthur. Renamed *The Minority Game* and properly formalized, it was introduced to physicists by Damien Challet and Yi-Cheng Zhang.

A huge literature on the Minority Game has appeared on Physics journals. Good introductions have been proposed, among others, by Esteban Moro and Chi-Ho Yeung and Yi-Cheng Zhang.

Arthur WB (1994) Inductive reasoning and bounded rationality. Am Econ Rev 84:406–411

Challet D, Zhang YC (1997) Emergence of cooperation and organization in an evolutionary game. Phys A 246:407–418

Cipra BA (1987) An introduction to the Ising model. The American Mathematical Monthly 94:937–959

Moro E (2004) The minority game: an introductory guide. In: Korutcheva E, Cuerno R (eds) Advances in condensed matter and statistical physics. Nova Science Publishers, New York, pp 263–286. (Also available online at http://arxiv.org/abs/cond-mat/0402651v1)

Schelling TC (1971) Dynamic models of segregation. J Math Sociol 1:143–186

Yeung CH, Zhang YC (2009) Minority games. In: Meyers RA (ed) Encyclopedia of complexity and systems science. Springer, Berlin. (Also available online at http://arxiv.org/abs/0811.1479v2)

## *Some Pitfalls of Utility Maximization*

The idea that probabilities measured on samples of size zero are somewhat awkward is quite old, and evidently linked to the frequentist view of probabilities. Daniel Ellsberg circulated this idea among economists, where in the meantime the subjectivist view of probability judgements had become dominant. Sub-additive probabilities were conceived by Bernard Koopman in the 1940s and popularized among economists by David Schmeidler in the 1980s.

Maurice Allais submitted his decision problem to Leonard Savage, who did not behave according to his own axioms of rational choice. Since then, Savage presented utility maximization as a normative, not as a descriptive theory. Prospect

Theory was advanced by Daniel Kahneman and Amos Tversky; it comes in a first version (1953), and a second version (1992).

The preference reversals highlighted by Paul Slovic have triggered a huge literature. A recent book edited by Sarah Lichtenstein and Paul Slovic gathers the most important contributions.

Kenneth Arrow originally devised his paradox as a logical difficulty to the idea of a Welfare State that would move the economy towards a socially desirable equilibrium. However, it may concern any form of group decision-making.

Michael Mandler is the main reference for a possible conciliation of Slovic's and Arrow's paradoxes with utility maximization, provided that preferences are incomplete.

Allais M (1953) Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école americaine. Econometrica 21:503–546

Arrow KJ (1950) A difficulty in the concept of social welfare. J Polit Econ 58:328–346. (Reprinted in The collected papers of Kenneth J. Arrow. Blackwell, Oxford, 1984)

Ellsberg D (1961) Risk, ambiguity, and the savage axioms. Quart J Econ 75:643–669

Kahneman D, Tversky A (1953) Prospect theory: an analysis of decision under risk. Econometrica 21:503–546

Koopman BO (1940) The axioms and algebra of intuitive probability. Annal Math 41:269–292

Lichtenstein S, Slovic P (2006) (eds) The construction of preference. Cambridge University Press, Cambridge

Mandler M (2005) Incomplete preferences and rational intransitivity of choice. Games Econ Behav 50:255–277

Savage L (1967) Difficulties in the theory of personal probability. Philos Sci 34:305–310

Schmeidler D (1989) Subjective probability and expected utility without additivity. Econometrica 57:571–587

Tversky A, Kahneman D (1992) Advances in prospect theory: cumulative representation of uncertainty. J Risk Uncertainty 5:297–323

## *Logic of Consequence and Logic of Appropriateness*

An introduction and thorough discussion of the differences between the logic of consequence and the logic of appropriateness can be found in a book by James March, *A Primer on Decision Making*. The same Author, in 1974 and 1976, has been the first to point to the fact that human beings distort their memories of the past in order to construct coherent stories that guide them into the future. This point has been also made by Karl Weick, who wrote a lengthy treatise on this subject a few decades later.

A number of psychological experiments confirm this idea. Interested readers may start with the works of Daryl Bem, Michael Conway, Michael Ross and a book edited by Ulric Neisser and Robyn Fivush.

However, the trouble with the idea of human beings reconstructing the past is that they are not willing to concede that they do so. Thus it is extremely difficult to find case-studies. The one by David Lane and Robert Maxfield is possibly the only exception, though they were not allowed to published all the material they gathered during their investigation (Lane, personal communication).

A final remark on lack of communication in this stream of research. James March, Karl Weick and David Lane worked independently, possibly unaware of one another, focusing on the same issue but employing different expressions.

Bem DJ (1966) Inducing belief in false confessions. J Pers Soc Psychol 3: 707–710

Bem DJ (1967) Self-perception: an alternative interpretation of cognitive dissonance phenomena. Psychol Rev 74:183–200

Cohen MD, March JG (1974) Leadership and ambiguity: the American College President. McGraw-Hill, New York

Conway M, Ross M (1984) Getting what you want by revising what you had. J Pers Soc Psychol 47:738–748

Greenwald A (1980) The totalitarian ego: fabrication and revision of personal history. Am Psychol 35:603–618

Lane DA, Maxfield RR (2005) Ontological uncertainty and innovation. J Evolut Econ 15:3–50

March JG (1994) A primer on decision making. The Free Press, New York

March JG, Olsen JP (1976) Organizational learning and the ambiguity of the past. In: March JG, Olsen JP (eds) Ambiguity and choice in organizations. Universitetsforlaget, Bergen

Neisser U, Fivush R (1994) (eds) The remembering self: construction and accuracy in the self-narrative. Cambridge University Press, Cambridge

Ross M, Newby-Clark IR (1998) Costructing the past and future. Soc Cognit 16:133–150

Weick KE (1979) The social psychology of organizing. Random House, New York

Weick KE (1995) Sensemaking in organizations. Sage, Thousand Oaks

## *Tools for the Logic of Appropriateness*

This chapter did not deal with tools where categories pre-exist to the information that is being received, namely, supervised neural networks and Case-Based Decision Theory. Readers interested in supervised neural networks may start with the classical handbook by Rumelhart, McClelland and the PDP Research Group. Readers interested in Case-Based Decision Theory may refer to a series of articles by Izhak Gilboa and David Schmeidler.

The earliest intuitions on the nature of mental categories date back to Ludwig Wittgenstein. A good explanation of the main features of mental categories, and why they are so different from our common idea of what a "category" is, is provided by George Lakoff in his *Women, Fire, and Dangerous Things*.

So far it regards unsupervised neural networks, the classic book by Teuvo Kohonen is still unrivaled for its combination of mathematical rigour and philosophical insight. Having been written at an early stage, it still keeps a strong link between artificial neural networks and the human brain.

Paul Thagard is the basic reference for constraint satisfaction networks. Constraint satisfaction networks appear in several contributions to the book *The Construction of Preference*, edited by Sarah Lichtenstein and Paul Slovic, mentioned in the section "Some Pitfalls of Utility Maximization". Regarding the importance of focussing on two alternatives in order to arrive at a decision, see a working paper by Guido Fioretti.

Evidence Theory started with a book by Glann Shafer in 1976, and triggered a small but continuous flow of mathematical works since then. An article by Guido Fioretti explains it to social scientists, along with examples of applications to decision problems.

Fioretti G (2009) Evidence theory as a procedure for handling novel events. Metroeconomica 60:283–301

Fioretti G (2011) Either, or: exploration of an emerging decision theory (Working paper). http://ssrn.com/abstract=1314352

Gilboa I, Schmeidler D (1995) Case based decision theory. Quart J Econ 110:605–639

Kohonen T (1989) Self-organization and associative memory. Springer, Berlin

Lakoff G (1987) Women, fire, and dangerous things. University of Chicago Press, Chicago

Rumelhart DE, McClelland JL, The PDP Research Group (1986) Parallel distributed processing: explorations in the microstructure of cognition. MIT Press, Cambridge, MA

Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton

Thagard P (2000) Coherence in thought and action. MIT Press, Cambridge, MA

## References

Arbib MA (ed) (2002) The handbook of brain theory and neural networks, 2nd edn. MIT Press, Cambridge, MA

Aumann R, Hart S (eds) (1992) Handbook of game theory with economic applications, vol 1. North-Holland, Amsterdam

Aumann R, Hart S (eds) (1994) Handbook of game theory with economic applications, vol 2. North-Holland, Amsterdam

Aumann R, Hart S (eds) (2002) Handbook of game theory with economic applications, vol 3. North-Holland, Amsterdam

Chattoe-Brown E, Edmonds B (2013) Evolutionary mechanisms. Chapter 18 in this volume
Graupe D (2007) Principles of artficial neural networks. World Scientific, Singapore
Hu YH, Hwang JN (eds) (2002) Handbook of neural network signal processing. CRC Press, Boca Raton
Rasmusen E (2007) Games and information: an introduction to game theory, 4th edn. Blackwell, Malden
Ripley BD (1996) Pattern recognition and neural networks. Cambridge University Press, Cambridge
Weibull JW (1997) Evolutionary game theory. MIT Press, Cambridge, MA

# Chapter 14
# Social Constraint

**Martin Neumann**

**Why Read This Chapter?** To understand social norms and their complexities, including: how they can operate, how they can effectively constrain action and how such processes have been represented within simulations. The chapter also helps the reader to acquire an integrated view of norms and become aware of some of the relevant work simulating them using this framework.

**Abstract** This chapter examines how a specific type of social constraint operates in Artificial Societies. The investigation concentrates on bottom-up behaviour regulation. Freedom of individual action selection is constraint by some kind of obligations that become operative in the individual decision making process. This is the concept of norms. The *two-way dynamics* of norms is investigated in two main sections of the chapter: the effect of norms on a social macro-scale and the operation of social constraints in the individual agent. While normative modelling is becoming useful for a number of practical purposes, this chapter specifically addresses the benefits of this expanding research field to understand the dynamics of human societies. For this reason, both sections begin with an elaboration of the problem situation, derived from the empirical sciences. This enables to specify questions to agent-based modelling. Both sections then proceed with an evaluation of the state of the art in agent-based modelling. In the first case, sociology is consulted. Agent-based modelling promises an integrated view on the conception of norms in role theoretic and individualistic theories of society. A sample of existing models is examined. In the second case, socialisation research is consulted. In the process of socialisation the obligatory force of norms become internalised by the individuals. A simulation of the feedback loop back into the mind of agents is only in the beginning. Research is predominantly on the level of the development of architectures. For this reason, a sample of architectures is evaluated.

M. Neumann (✉)
University of Koblenz-Landau, Institute of information systems in Business and Public Administration, Universitaetsstr. 1, 56070 Koblenz, Germany
e-mail: maneumann@uni-koblenz.de

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_14,

## 14.1 Introduction

Some kind of mechanism for action selection has to be implemented in the agents. The decision making process of isolated agents may be governed by BDI architectures. Developed by the philosopher Michael Bratman as a model of rational decision making – in particular to clarify the role of intentions in practical reasoning according to the norms of rationality – (Bratman 1987), the BDI framework has been adopted by Rao and Georgeff for the development of software technology (Rao and Georgeff 1991). BDI agents are data structures that represent beliefs about the environment and desires of the agents. Desires enable the goal-directed behaviour of an agent. Moreover, the decision-making process has to be managed, which is accomplished in two stages. To achieve a goal, a certain plan has to be selected, denoted as the intention of the agent. Secondly, the agent undertakes means-ends calculations, to estimate which actions are necessary to reach its goals. Intentions are crucial in mediating between the agent's beliefs and desires and the environment.

In groups, however, behaviour is more effective if agents orient their actions around other agents. The simplest example is of two robots moving towards each other. They have to decide how to pass. Hence, individual action selection has to be restricted by social constraints. In principle there exist two options to model social constraints on the individual agent's action selection: top down regulation by a central authority or bottom-up regulation of action selection. Top-down regulation can be exhibited computationally by a central processor, or – in human societies – by coercion of a central authority. An evaluation of modelling approaches to the former kind of social regulation can be found in the chapter on Power and Authority. In bottom-up approaches, freedom of individual action selection is constraint by some kind of obligatory forces that become operative in the individual decision making process even without a controlling authority. This is denoted by the concept of *norms*. Normative behaviour regulation can be enforced by some kind of generalised social influence such as sanctions and encouragement. Crucial for the normative behaviour regulation is that social constraints are internalised in the individual agent's decision making process. This can range from more or less automatically executed habits to processes of normative reasoning and balancing competing goals. In contrast to top-down regulation it is not based purely on coercion. Since bottom-up regulation is more flexible than pre-determined constraints by a central authority, the past decade(s) have witnessed a growing interest in the inclusion of norms in multi-agent simulation models. Numerous factors are responsible for attention being paid to norms in artificial societies, ranging from technical problems in the co-ordination of multi-agent system, e.g. with moving robots (Shoham and Tennenholtz 1992; Boman 1999), or practical problems such as e-commerce or electronic institutions (Lopez and Marquez 2004, Vazquez-Salceda et al. 2005) to philosophical interest in the foundation of morality (Axelrod 1986; Skyrms 1996, 2004) and the investigation of the wheels of social order (Conte and Castelfranchi 2001). The contribution of Artificial Societies to the latter problem is in the focus of this chapter: what is the potential contribution of Artificial Societies to the investigation of social order in human societies?

However, in the literature – not only, but *also* in the simulation literature – a great variety of concepts of norms exist. Typically, these differences are not much discussed in the AI literature. Nevertheless, some decisions are made, consciously or unconsciously. For this reason, this chapter aims to introduce the reader to the different concepts that are often only implicit in the different approaches to normative simulation models. This requires some background information about the different concepts of norms in the different empirical sciences. However, one reason for the variety of concepts of norms is that their investigation is scattered over a vast variety of different disciplines. For this reason also this chapter has to concentrate on some empirical disciplines that are of particular importance for agent-based modelling. A first restriction is motivated by the decision to concentrate on bottom-up behaviour regulation in this chapter. This suggests to exclude the literature on norms that can be found in the political sciences or the theory of law, since these norm concepts are more closely related to top-down approaches. Secondly, agent-based modelling is of particular relevance for the study of social mechanisms. These mechanisms, however, are of some kind of generality. Typically, Artificial Societies investigate stylised facts. This suggests focussing the examination by excluding historical narratives that can be found in anthropological studies or historical sociology such as the work of Michel Foucault or Norbert Elias. Instead, the survey of the empirical sciences concentrates on the literature that investigates the theoretical foundation of the specific dynamics between the micro and the macro level that is the particular focus of agent-based modelling. Namely, how norms influence the operations and effects of social systems and how such a normative structure of the social system recursively affects the generating agent level. This calls for a survey of sociology and socialisation research in social-psychology. They are most relevant sciences for an investigation of the contribution of norms to the wheels of social order.

Beside this introduction, the chapter contains three main sections. Sections 14.3 and 14.4 consist of two parts: one providing a broad overview of the empirical counterpart and a subsequent one about modelling. The sections can be read independently. A reader with a background in the empirical sciences who wants to get informed about simulation might concentrate on the parts evaluating simulation models. The sections provide the following information:

**Section** 14.2**:** First, a brief exposition of the *core concepts* of norms in the empirical sciences is provided. It is suggested to have a look at it, because here the core problems are exposed that are investigated in the following sections. These are the two main parts.

**Section** 14.3**:** an investigation of the dynamics of norm spreading from individuals to a social group and how this effects the operations of the social system is undertaken in this part. This refers to the sociological question of the operations and effects of normative behaviour regulation on the social *macro-level* (the emergence of a behaviour regularity). The section is divided into two parts: first the sociological questions to agent-based modelling are developed. Then a sample of models is examined. The sample is divided into two categories of models: models inspired by a game theoretic problem description and cognitive models in an AI tradition. This section is particular relevant for modelling normative regulation in *Artificial Societies*.

**Section** 14.4**:** in this part the recursive feedback loop is investigated. It is examined, how social behaviour regulation is executed in the *individual agent*. This refers to the socio-psychological problem of norm internalisation. To represent this process, cognitively rich agents are necessary. Also this section is divided into two parts: first, a problem exposition will be provided. This refers to theories of socialisation. Then a sample of architectures is examined with regard to the question how norms become effective within the agent. This section might be the most interesting one for readers interested *cognitive modelling*.

Finally, the main findings of the chapter are summarised in concluding remarks and further reading suggested for those who want to investigate further.

## 14.2 The Concept of Norms

Before turning the attention to an examination of existing modelling approaches, a brief exposition of the core concepts of norms will be provided. In some way, some of these aspects have to be represented in a normative simulation model. Dependent on the research question, the developer of a normative simulation model may concentrate on only some aspects. In fact no model includes *all* aspects. However, it might be useful to be aware of a more comprehensive perspective. Operationally, norms can be described as correlated structures. Agent behaviour exhibits regularities that can be found – to a certain degree – in a entire population. This can be described as a social constraint. From a social science perspective, norms are the most important concept of social constraints on behaviour regulation. Norms belong to the fundamental concepts in sociology to explain human behaviour. Summarising the individual and social aspects of norms, Gibbs provided the definition that "a norm is a belief shared to some extent by members of a social unit as to what conduct ought to be in particular situations or circumstances" (Gibbs 1981, p.7). Unfolding this concise definition reveals three essential components:

- An *individual* component: a belief
- A *social* component: the belief is shared by other members of a social unit
- A *deontic:* a conduct is obliged

1. The capability to understand the meaning of a deontic language game is acquired in the childhood. The deontic prescribes individual behaviour. The capacity to play language games that are centred around words such as *'you ought'* is the precondition for a normative orientation. It allows for the internalisation of norms. Although in the past decades consideration has been given to understanding processes of internalisation over an individual's entire life span, the central processes arguably take place during childhood. Indeed, the transmission of

cultural values has even been denoted as a ‘second birth’ (Claessens 1972). An individual may become accepted into a wider society through a variety of means. Norm internalisation is argued to represent one of the stronger mechanisms by which this process occurs. In contrast to compliance under coercion or to simply copying other behaviour patterns, internalisation, it is argued, is coupled with an individual’s intrinsic motivation and sense of identity. The mechanisms by which this occurs are the focus of socialisation research. Socialisation is the bridge between the individual and society (Krappmann 2006). Hence, socialisation research is at the border of psychology and sociology, and contributions from both disciplines can be found in the literature.

2. It is a central element of the concept of norms that it implies both a psychological component, namely the mental state of a belief, as well as a social component. This becomes apparent by the fact that this belief is shared by a number of individuals. Norms are thus essential for a comprehension of the relation between structure and agency. This is often denoted as the micro–macro link. While the aspect of agency can be found in the psychological aspect of norms, structure is entailed in the social prescription. This makes norms a fundamental building block of the wheels of social order. An understanding of norms is thus crucial for an understanding of the social world. They are a central mechanism in the *two way dynamics* creating of social order. On the one hand, individual interactions might produce social macro-patterns, namely a normative order. On the other hand, once the agents on the individual level recognise patterns on the resultant social level, these patterns might have an effect on the mind of the individual agent, namely the recognition of a deontic prescription. This complex dynamics can be embraced in the following schema:

This schema can be illustrated by an example: interactions of individual actors might lead to a certain aggregate result, for instance, social stratification. This macro-property, however, might have an effect on how the actor cognitively structures the perception of the world. This is an intra-agent process. For instance, actors might regard themselves as helpless losers or they might think that you can get what you want. This world perception, in turn, has an effect on how these actors behave in the interactions between the agents. Now the feedback loop is closed: the interactions result again in an aggregated macro-structure. The reader may think of other examples appropriate for his or her purpose. Presumably, it will turn out that the dynamics of the particular models can be described in terms of such a two way dynamics between inter – and intra agent processes (Fig. 14.1).

### 14.2.1 The Sociological Perspective

In this section process (1) is considered: the emergence and the properties of a normative macro structure. Hence, this section will concentrate mainly on the social aspect of norms, i.e. the one-way dynamics of *inter* agent processes. Before turning the attention to existing modelling approaches, the questions will be developed that

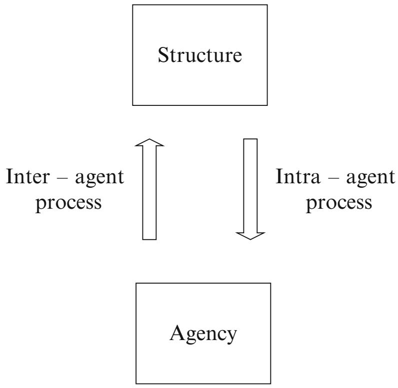



**Fig. 14.1** The two-way dynamics of norms

need to be posed to normative multi-agent models for a comprehension of the effects and operations of norms in a social system. For this purpose sociological theory will be utilised. Readers that are either familiar with this subject or with a purely practical interest might go on to Sect. 14.3. (2). However, a more or less implicit reference to the problems posed by social theory can be discerned in most of the simulation models.

The *function* of normative behaviour regulation for social integration (i.e. a social structure) has been particularly emphasised by the classical role theoretic account within the sociological theory (Neumann 2008b). It can be traced back to the functionalist sociology of Durkheim and Parsons. This theoretical account was decisively influenced by anthropological research of the late 1920s when cultural anthropologists discovered the variety of the organisation of the social life and the correlation of the structure of personality and society (Malinowski 1927; Mead 1928). *Role theory* claims that action is guided by a normative orientation, insofar as social roles are described by social norms. The concept of norms is introduced in this account as a *structural constraint* of individual actors. This theory of action is often paraphrased as the 'homo sociologicus'. For a long time this remained the dominant stream of sociological theory.

To specify where norms are placed in the theoretical architecture, let us consider the example of the famous Mr Smith, introduced by Ralf Dahrendorf (1956) in his analysis of the 'homo sociologicus' to characterise key elements of sociological *role theory*. We first meet him at a cocktail party (in the early 1950s) and want to learn more about him. What is there to find out?

Mr Smith is an adult male, ca. 35 years old. He holds a PhD, and is an academic. Since he wears a wedding ring, we know that he is married. He lives in a middle-sized town in Germany and is a German citizen. Moreover, we discover that he is Protestant and that he arrived as a refugee after the second World War in a town populated mostly by Catholics. We are told that this situation caused some difficulties for him. He is a Lecturer by profession and he has two kids. Finally, we learn that he is the third chairmen of the local section of a political party, Y, a passionate and skilful card player and a similarly passionate though not very good driver. This approximates to what his friends would tell us.

In fact, we may have the feeling that we know him rather better now. After all, we have some expectations as to how a lecturer is likely to behave. As a lecturer he stands in certain relations to colleagues and pupils. As a father he will love and care for his children, and card playing is also typically associated with certain habits. If we know the party Y, we will know a lot more about his political values. However, all that we have found out represent *social facts*. There are a lot more lecturers, fathers and German citizens beside Mr Smith. In fact, none of this information tell us anything about the unique identity of Mr Smith. We simply discovered information about social positions, which can, of course, be occupied by varying persons. However, social positions are associated with specific *social roles*. Roles are defined by the specific attributes, behaviour and social relations required of them. Demands of society determine—to a certain degree—individual behaviour. Individuals are faced with obligations and expectations. This social demand is transmitted to the individual by *norms*. Norms are the 'casting mould' (Durkheim 1895) of individual action. They regulate how lecturers, fathers and members of political parties ought to act to fulfil the role expectations of society. However nowadays, it is improbable that we will be told anything about driving competence. Thus, we have learned another lesson: Norms may *change* over the course of time. In particular, Talcott Parsons (1937) emphasised that the ends of individual actions are not arbitrary, but rather are prescribed by social norms. Thus, norms represent a key concept within sociological role theory.

To examine the explanatory account of this theoretical approach, the present investigation will abstract from a description of the content of concrete norms. We concentrate on the methodological characteristics of norms in general, rather than the content of specific norms. On closer inspection of this example, we find out that the concept of social norms is characterised by three *key elements:*

*First,* norms show some degree of generality. They are regarded as the 'casting mould' of individual action (Durkheim 1895). The very idea of role theory is that a social role must not be restricted to a unique individual. For instance, the roles of lecturer or chairman of a political party can be performed by different individuals. It might not, of course, be arbitrary as to who will play this role. In fact, it is a major focus of the empirical counterpart of role theory, statistical analysis of variables, to investigate the distribution of roles. For instance, monetary background might determine an individual's chances of securing an academic position. Nevertheless, academic positions are a feature of society, not the individual. In the classical account, the generality of norms is simply a given. However, agent-based models start from individual agents. Thus, in the individualistic approach norms have to *spread* in some way from one agent to another to gain generality. The explanation of norm spreading is essential for a reconstruction of social norms in terms of individual actors.

*Secondly,* the role set of father or lecturer encompasses a huge action repertoire. The choice of a concrete action cannot be determined only solely by an external force. The ends of an action have to be determined internally by the individual actor executing a specific role. This means that the ends of individual actions are (to a certain degree) determined by the society. For instance, it is a social norm how a caring father would look like and what kind of actions are to be undertaken. In fact, this varies between societies. This knowledge is often denoted as *internalisation,* even though the psychological mechanisms are not in the focus of sociological theory. Thus, already a comprehension of the inter agent processes that constitute a social role calls for a related subjective element at the level of the individual agent.

> *Thirdly,* this approach is characterised by a certain type of analysis: the normative integration of society (Parsons 1937; Davies and Moore 1945; Merton 1957). Hence, the question is to a lesser extent concerned with the origin of norms than with the *function* of norms for a society. For instance, the role of the father is to educate his child. The role of the lecturer is crucial for the socialisation of pupils. Thus, both roles are functionally relevant for the reproduction of the society.

However, in the past decades, this paradigm has been severely criticised: Firstly, role theory has been criticised for sketching an *oversocialised* picture of man (Wrong 1961). Already in the 1960s, Homans (1964) claimed to 'bring man back in'. In fact, the individual actors of the role theory have often been regarded as more or less social automata. If they have properly internalised the norms, they execute the program prescribed by their roles. Secondly, role theory is built on the claim that social phenomena should be explained with social factors (Durkheim 1895). Roles are a pre-given element in this theoretical architecture. Roles (and thus: norms) emanate from society. However, the origin of both roles and society is left unexplained. In this so-called functionalist methodology of the role theoretic account, it is argued that to perform a social role is to perform a social function and this social function is argued to be the source of the norm. However, this perspective lacks a description of a mechanism by which norms become effectively established. This deficit suggests to 'bring man back in'.

In fact, the explanatory deficit engendered an alternative. Others have suggested building sociology on the foundations of individual actors. This built on views advocated by John Stuart Mill. This is the programme of the so-called *methodological individualism* (Boudon 1981; Raub and Voss 1981; Coleman 1990; Esser 1993). A shift 'from factors to actors' (Macy and Willer 2002) can be observed in the foundations of sociology in the recent decades. This approach regards the appearance of norms as an *aggregate product* of a sum of individual actions. This approach to norms has been particularly stressed by Rational Choice Theories. The deficit of this approach, however, is a lack of a cognitive mechanism that could explain the deontic component of norms, i.e. how the social level may exhibit an obligatory force on the individual by norms. Agent-based modelling promises to overcome the complementary deficits of both theoretical accounts:

On the one hand, agent-based models contribute to the individualist theory building strategy: obviously, agents are the fundamental building block of agent-based models. In agent-based simulation models (Artificial Societies), structures emerge from individual interaction. On the other hand, however, the great advantage of this methodology is that it allows to explicitly consider the complexity generated by individual interactions. Compared to purely analytical models it is a great advantage of simulation approaches that they are not restricted to single or representative actors. This complexity generated by this particular feature enables the investigation of the feedback loop between individual interaction and collective dynamics. This has led to a growing awareness of its potential for investigating the building-blocks of social structure. For this reason, it is even claimed that agent-based simulation allows us to "discover the language in which the great book of social reality is written" (Deffuant et al. 2006), by constituting the promise to understand "how actors produce, and are at the same time a product of social reality" (Deffuant et al. 2006). Since structures on the macro level are a product of individual interaction, a causal understanding of the processes at work seems possible that might fill the gap between individual action and social structure with agent-based models.

This feature of agent-based modelling suggests that this methodology enables to bring together the role theoretic perspective that norms are a *structural constraints* and the individualistic perspective of norms as an *aggregated product* of a sum of individual actions. Hence, it might provide both a causal mechanism of how normative action constraints get established as well as how social forces exhibit cognitive power. For this reason the focus will now shift to an evaluation of simulation models.

### *14.2.2 Simulation Models of Norms*

After the theoretical problem exposition, the particular focus of this section is an investigation of the contribution of *simulation experiments* to sociological theory. The recourse to sociological theory in Sect. 14.3.(1) revealed the following *questions:*

- Can they provide insights into the normative regulation of society; that is, do they also reproduce the findings of a functional analysis of the effects and operations of norms in a society (*focus of contribution*)?
- Moreover, do they allow for a causal reconstruction of the mechanisms that generate the functional interconnectedness on the social level? This implies that two further questions have to be addressed:
- What transforms the agents in such a way that they factually follow norms? That is, what are the causal mechanisms at work that enable an internalisation of norms (*transformation problem*)?
- By what mechanisms in the model can norm-abiding behaviour spread to or decay from one agent to another (*transmission problem*)?

These questions will be examined in this section. It has to be emphasised that the investigation concentrates on methodology, not on the contents of norms governing concrete roles such as father or lecturer. However, existing models are clustered around various intuitions about norms, conventions or standards of behaviour. The concrete research question differs from model to model. Some models concentrate on the emergence or spreading of norms. Others concentrate on functional aspects or the feedback of norms on individual agent's behaviour. A *multiplicity of concepts* is at hand. Hence, a comprehensive review of all models and accounts that may be in some way related to the study of norms would go beyond the scope of this investigation.

The overwhelming mass of models, however, can be traced back to (or is at least influenced by) *two* traditions in particular: first, *game theory* and secondly an architecture of cognitive agents with some roots in *Artificial Intelligence*. Tradition and theoretical background has a direct impact on the terminology used. Depending on their background, the models tend to be communicated in different scientific communities. Additionally, references in articles tend to depend on their authors' background. Under the perspective of content, the models in the AI tradition typically contain references to conceptual articles relating to agent architectures. Articles with models in a more game theoretical tradition typically refer to game theoretic literature for the characterisation of the interaction structures in which the authors are interested. Of course, this tradition-influenced framing, publishing and referencing is a *tendency*. It does *not* constitute a clear-cut disjunction without any intersection. It has to be emphasised that this is neither a very precise nor a disjunctive categorisation. To some degree, the distinction between game theory and DAI is a distinction in the mode of speech employed by the authors. Some problems of game theoretic models could also be formulated in a DAI language and vice versa. The categorisation of models as following the DAI tradition shall only indicate that the agents employed by these models are in some way cognitively richer than those in the so-called game theoretic models.

Nevertheless, this distinction gives a rough sketch of the line of thought followed by the models, and also of the kind of problems, the concepts for their investigation, and the mode of speech in which the paper is presented. Moreover, this categorisation provide hints to other areas of research that are closely related to the models considered in this article: For instance, simulation is only a small sub-discipline of game theory in general and the distinction between analytical and simulation results is only gradual. Simulation models might describe problems in game theoretic terms, but the method of resolution is not that of analytical game theory (Binmore 1998). In fact, investigating norms with the means of *analytical* game theory is a highly active research field.

#### 14.2.2.1 Models Using a Game Theoretic Problem Description

First a closer examination of a sample of models applying a game theoretic problem description will be undertaken. The models investigated here build on the framework described in the chapter on games and utility in this handbook (Fioretti 2013).

> Axelrod (1986) studies the evolution of a standard not to cheat via a "norms game" and a "meta-norms game". In the "norms game" defectors may be punished by observers. In the "meta-norms game" it is also the case that observers of a defection that do not punish the defector may be punished. Only the latter game leads to a widespread standard of not defecting.
>
> Coleman (1987) investigates the effect of interaction structures on the evolution of co-operation in a prisoner's dilemma situation. Only small groups can prevent the exploitation of strangers.
>
> Macy and Sato (2002) examine the effect of mobility on the emergence of trust among strangers in a trust game. While agents with low mobility trust only their neighbours, high mobility supports the evolution of trust among strangers.

Vieth (2003) investigates the evolution of fair division of a commodity in an ultimatum game. Including the ability to signal emotions leads to a perfectly fair share. If detection of emotions is costly the proposals even exceed fair share.

Bicchieri et al. (2003) present a model of a trust game. It demonstrates how a trust and reciprocate norm emerges in interactions among strangers. This is realised by several different conditional strategies.

Savarimuthu and colleagues (2007) study the convergence of different norms in the interactions of two different societies. Both societies play an ultimatum game against each other. Two mechanisms are examined: a normative advisor and a role model agent.

In the model by Sen and Airiau (2007) a co-ordination and a social dilemma game are examined. Agents learn norms in repeated interactions with *different* agents. This is denoted as social learning to distinguish this interaction type from repeated games with the same player. The whole population converges to a consistent norm.

Obviously, all models have been developed for differing concrete purposes. To examine the extend to which these models capture the explanatory problems of the *contribution* problem, *transformation* problem and *transmission* problem, the various accounts of the different models will be outlined in a table. Moreover, a short hint to the concrete implementation is provided. This will enable an evaluation inasmuch normative agent-based models have so far reached the goal to discover 'the language in which social reality is written'.

| | Contribution | Transformation | Transmission | Implementation |
|---|---|---|---|---|
| Axelrod | Norm dynamics (*norms* broadly | conceived!) | Sanctions | Social learning; replicator dynamics |
| Dynamical propensities | | | | |
| Coleman | Norm dynamics | Punishment by defections (memory restrictions for identifying defections as sanctions) | (a) Group size (acquaintance) (b) Additionally: replicator dynamics | Conditional strategies |
| Macy & Sato | Norm dynamics | Losses by exclusion from interaction | Social learning | Dynamical propensities |
| Vieth | Norm dynamics | Losses by rejection | Social learning; replicator dynamics | Dynamical propensities |
| Bicchieri et al. | Norm dynamics | Sanctions by retaliating super game strategies | Strategy evolution; replicator dynamics | Conditional strategies |
| Savarimuthu et al. | Norm dynamics; functional analysis | Losses by rejection; advice | Advice updating based on collective experience | Dynamical propensities |
| Sen & Airiau | Norm dynamics | Experience | Social learning guiding behaviour convergence | Dynamical propensities |

Lessons

The classical model employing a game theoretical approach for the problem description is Axelrod's model. The main contribution of this approach is a clear understanding of the emergence of (commonly shared) normative behaviour constraints. The starting point of the models is a dilemma situation. This is a consequence of the game theoretic problem description. Simulation allows for an evolutionary perspective by analysing repeated games. Typically, in the long run and under specific conditions (which vary from model to model) it is possible that behaviour conventions emerge, that are in the benefit of all agents or that represent some intuitions about fairness. The diffusion of a behavioural regularity is then regarded as a norms. The subjective side of an obligatory force is not in the focus of this approach. Hence, what lessons can be gained from the investigation of this conception to model normative action constraints with regard to the specific questions?

1. Contribution
   Already Axelrod's classical model provides a causal explanation for *norm spreading*. This includes a designation of mechanisms of norm transmission and normative transformation. An investigation of the functional effect of norms on the society is left aside. This orientation remained predominant in this line of research. Typically models with a game theoretic background concentrate on the question of norm *dynamics*. They ask how a behaviour regularity emerges in an agent population. This is the problem of the Rational Choice tradition in sociological theory, namely the perspective of norms as an aggregated product of a sum of individual actions.
2. Transformation
   It is striking that, except for the model by Sen and Airiau, the transformation of individual behaviour in all models is driven by some kind of sanctions. However, also in the Sen & Airiau model agents react to losses of utility values. This is the causal mechanism of norm spreading. The great advantage of this account is to shed light on the process of norm *change*. As it has become apparent in discussing Mr Smith, this process can also be observed in human societies. However, norm change is only barely captured by the functional account of role theory. On the other hand, the models of this tradition only include a *restricted functional perspective:* On an individual level, the agents' choice of action is guided by the functional consideration of calculating the expected utility. However, a corresponding analysis on the social macro-level can be found only in Savarimuthu et al.'s model.
3. Transmission
   With regard to the transmission of norms it is striking that social learning is implemented in many game theoretic models by a *replicator dynamics*. Typically this is *interpreted* as social learning by imitation. If applied in a context where *no* real natural selection, rather than some kind of *learning* is at work, then using a replicator dynamics amounts to saying: Somehow the individuals learn in a way that – measured by the relative overall success of their type of behaviour – more successful types of behaviour become more frequent. As an effect this may

be true. However, no mechanism is indicated. In this dimension, the models struggle with the same kind of problem as functional analysis which the individualistic program tries to resolve; namely, the lack of a causal explanation.

4. Implementation
From the perspective of the role theory of action, a weakness of this approach becomes apparent, one immediately related to the game theoretic problem description. Agents are faced with a strategic (binary) decision situation. Thus, they have a fixed set of behaviour (Moss 2001). For this reason behaviour change can be implemented by dynamical propensities (i.e. the propensity to defect is dynamically updated). Faced with this situation, agents choose the alternative that maximises their expected utility. However, behaviour change goes not along with goal change. Agents can do no more than react to different environmental conditions. The agents' behaviour is guided strategic *adaptation.* An *active* element of normative orientation in the choice relating to the ends of action cannot be found in a game theoretic approach. This is simply due to the fact that agents do not possess any internal mechanism to reflect and eventually change their behaviour, other than the desire to maximise utility. This point has already been highlighted in Parsons' critique of 'utilitarian theories' of action (Parsons 1937): namely, that the ends of individual actions are in some way arbitrary. Even though the modelling of behaviour transformation is the strength of this kind of models, the ends of the action remain unchanged: the goal is to maximise utility. In this respect, the relation between the action and the ends of the action remains arbitrary.

   However, the very idea of role theory is to provide an answer to the question: Where do ends come from? Parsons' (and Durkheim's) answer was the internalisation of norms. A corresponding answer to this problem is not supplied in game theoretical models. This is due to the fact that agents do not act because they want to obey (or deviate from) a norm. They do not even 'know' norms. Even though the model provides a mechanism for the transformation of the agents, this is not identical with norm internalisation. This remains beyond the scope of this account. The agents' behaviour can only be *interpreted* as normative from the perspective of an external observer. Thus, transformation is not identical with internalisation. While the model provides a mechanism for behaviour transformation, it cannot capture the process of internalisation. Compared to the classical role theory, this is a principle limitation of a game theoretical description of the problem situation.

#### 14.2.2.2 Models Utilising Cognitive Agents

This shortcoming calls for cognitively richer agents. For this reason, a sample of models in the AI tradition will be examined more closely.

> Conte and Castelfranchi (1995b) investigate three different populations of food gathering agents: aggressive, strategic and normative agent populations. Aggressive agents attack 'eating' agents, strategic agents attack only weaker agents and normative agents obey a finder-keeper norm. The aggregated performance of the normative population is the best with regard to the degree of aggression, welfare and equality.

In an extension of the above model, Castelfranchi et al. (1998) study the interaction of the different agent populations. Interaction leads to a breakdown of the beneficent effects of norms, which can only be preserved with the introduction of normative reputation and communication among agents.[1]

Saam and Harrer (1999) present a further extension of Conte and Castelfranchi's model. They investigate the influence of social inequality and power relations on the effectiveness of a 'finder-keeper' norm.

Epstein (2000) examines the effect of norms on both the social macro and the individual micro level. On the macro level, the model generates patterns of local conformity and global diversity. At the level of the individual agents, norms have the effect of relieving agents from individual thinking.

Flentge et al. (2001) study the emergence and effects of a possession norm by processes of memetic contagion. The norm is beneficent for the society, but has short-term disadvantages for individual agents. Hence, the norm can only be retained in the presence of a sanctioning norm.

Verhagen (2001) tries to obtain predictability of social systems while preserving autonomy on the agent level through the introduction of norms. In the model, the degree of norm spreading and internalisation is studied.

Hales (2002) extends the Conte/Castelfranchi model by introducing stereotyping agents. Reputation is projected not on individual agents but on whole groups. This works effectively only when stereotyping is based on correct information. Even slight noise causes the norms to breakdown.

Burke and colleagues (2006) investigate the emergence of a spatial distribution of a binary norm. Patterns of local conformity and global diversity are generated by a decision process which is dependent on the local interactions with neighbouring agents.

The contribution of these models to the questions specified above can be summarised in the following table, which also includes a brief remark on the implementation specification.

| | Contribution | Transformation | Transmission | Implementation |
|---|---|---|---|---|
| Conte & Castelfranchi | Functional analysis | –/–[b] | –/–[b] | Conditional strategies |
| Castelfranchi et al. | Functional analysis | Updating conditionals (of strategies) through knowledge | Updating knowledge by experience (and communication[a]) | Conditional strategies |
| Saam & Harrer | Functional analysis | (a) –/–[b]<br>(b) Internalisation[a] | (a) –/–[b]<br>(b) Obligation[a] | Conditional strategies |
| Epstein | Norm dynamics; functional analysis | Observation | Social learning | Dynamical updating |
| Flentge et al. | Functional analysis | Memetic contagion | Contact | Conditional strategies |

(continued)

[1] For a more in-depth discussion of this model the interested reader is referred to the chapter on reputation (Giardini et al. 2013).

| | Contribution | Transformation | Transmission | Implementation |
|---|---|---|---|---|
| Verhagen | Norm dynamics | Internalisation | Communication | Decision tree |
| Hales | Functional analysis | Updating conditionals (of strategies) through knowledge | Updating knowledge by experience (and communication[a]) | Conditional strategies |
| Burke et al. | Norm dynamics | Signals | Social learning guiding behaviour convergence | Dynamical propensities (Threshold) |

[a]Only in a second experiment
[b]The agents are/are not already moral agents

## Lessons

The classical model of this kind of models is the one developed by Conte and Castelfranchi in (1995b). It was the starting point for several extensions. While the scope of these models has been significantly extended in the past decade, the most significant contribution of this approach still can be regarded as to enhance the understanding of the operation and effects of normative behaviour constraints.

Epstein, Burke et al. and Verhagen do not study specific norms but examine mechanisms related to the operations of norms. In particular, the spreading of norms is studied by these authors. In this respect, they recover and refine (by the notion of local conformity and global diversity, a pattern that cannot be found in game theoretic models) the findings of game theoretic models with the means of cognitive agents. The other models concentrate mainly on studying the effects and operations of specific norms in the society. Analysed are questions such as the effect of possession norms, for instance under the condition of social inequality and power relations. This is strongly influenced by Conte and Castelfranchi's (1995b) problem exposition.

What lessons can be gain from the investigation of this conception to model normative action constraints with regard to the specific questions?

1. Contribution
   A striking feature of these models is that they demonstrate how agent-based models are able to contribute to a *functional analysis* of norms. However, in contrast to the classical scheme of functional explanations in the social sciences, this result is reached by interactions of *individual agents*. Moreover, in these model a much stronger notion of norms is deployed than typically in game theoretic models. Norms are not just reached by mutual agreement, but are an explicitly prescribed action routine. The concept of norms in these models is in line with the role theoretical conception of norms as a *structural constraint* of individual actors. This conception of norms allows for a wider field of applications that could cover the role theoretic norm conception: these can be interpreted as internalised properties of the agents.

2. Transformation

   There exist a wide range of varieties how agents change their behaviour. Behaviour transformation is not as straightforward as in game theoretic models. However, it has to be emphasised that the very first model of Conte and Castelfranchi (1995b) did not include any behaviour transformation at all. Agents have no individual freedom in this model. As critics accuse the role theory, the action repertoire is also (depending on conditions) deterministic. Thus, even though the authors succeed in 'bringing man back in', the agents in the model are merely *normative automata*. Insofar as the norms are a pre-given element in the model, the approach can also be regarded as an 'over-socialised' conception of man.

   This limitation has been overcome by the subsequent developments. With regard to the transformation problem, the agents have become more flexible than in the very first model. However, a key difference to game theoretic models still remains: while game theoretic models mostly concentrate on sanctioning, in models of cognitive agents sanctions are only employed by Flentge et al. as the transformation mechanism.

   However, while the norms in these models can be interpreted as internalised properties of the agents, an investigation of the process of internalisation is only in the beginning. So far no commonly accepted mechanism of internalisation has been identified. Memetic contagion is a candidate. In Verhagen's model a quite sophisticated account is undertaken, including a self-model, a group model and a degree of autonomy. It is highly advanced in constructing a feedback loop between individual and collective dynamics. By the combination of a self- and a group-model a representation of the (presumed) beliefs held in the society is integrated in the belief system of individual agents. Conceptually, this is quite close to Mr Smith. However, it might be doubted whether the mechanisms applied are a theoretically valid representation of real processes.

3. Transmission

   Complementary to the wide range of different mechanisms of agent transformation, also a variety of different transmission mechanisms are applied. Basically, agents apply some kind of knowledge updating process, if agent transformation takes place at all. Up to date the transmission problem is no longer a blind spot of cognitive agents as it was the case in the Conte and Castelfranchi (1995b) model. By comparison, communication plays a much more important role than in game theoretic models and is much more explicitly modelled in models within the AI tradition. The processes utilised are more realistic mechanisms than the replicator dynamics of game theoretic models. However, no consensus has been reached, what an appropriate mechanism would be. This is also due to the fact, that a modelling of agent transformation and norm transmission is computationally more demanding than in game theoretic models. It has to be emphasised, however, that with regard to the transformation and the transmission problem, the borderlines of both approaches are no longer clear-cut. The models of Verhagen and Savarimuthu et al. include elements of the other line of thought.

4. Implementation
   Since actions performed by cognitive agents cannot be reduced to the binary decision to cooperate or defect, more complex behaviour rules than dynamic propensities have to be applied. The dominant approach for the implementation of normative behaviour constraints in cognitive agents is based on a straightforward intuition, namely to apply conditional strategies that are conditionally based on the agent's knowledge base. The strategy repertoire, however, depends on the concrete model.

   The overview of existing models has revealed that the focus of their contribution is mainly on the dynamics on the macro level. The questions of norm transmission and the focus of contribution concentrate on the social macro-scale. The analysis is focused on the one-way dynamics of effects and operations of norms on the *social level*. Hence, the analysis is focused on the emergence of social structure out of individual interaction rather than on the relation between structure and agency. This is one aspect of the full dynamics, namely the following process:

$$\text{inter-agent processes}: \ \textit{interaction} \ \Rightarrow \ \text{macro-property}: \ \textit{structure}$$

## 14.3 Socialisation: Norm Internalisation

It has been outlined, however, that the definition of a norm posses social and *psychological* components. Norms are essential for a comprehension of the relation of structure and agency. While processes of emergence of aggregated behaviour standards from interaction among individual agents has been extensively studied, a comprehension of the reverse process how the aggregate level gets back into agents' minds is not as yet fully reached. A full comprehension of normative behaviour regulation, however, has also to include the reverse dynamics of the effect of social structure on the *individual agency*. Already the problem of agent transformation refers to the effect of structure on the level of the individual agent. This is the most problematic aspect in the agents' design. It would include the following dynamics:

$$\textit{structure} \ \Rightarrow \ \text{Intra-agent processes}: \ \textit{agency}$$

This would be a step towards closing the feedback loop of the *two-way* dynamics. Obviously, intra agent processes are closely related to learning. The reader can find a more comprehensive examination of the state of the art in the chapter on evolution and learning. In particular, the effect of structure on (a transformation of) individual agency is particular relevant for studying the effects of norms, if agency is not restricted to a representative agent.

To represent such intra-agent processes, in particular the concept of *social learning* is well know in agent-based models. It is applied in a number of game theoretic models and can also be found in models of the AI tradition. The concepts of social learning, but also knowledge updating, can be traced back to *behaviouristic* psychological theories. Behaviourism is a theory of learning developed principally through experiments with animals. For instance, the conditioning experiments of Ivan Pavlov are well-known: he demonstrated that dogs can be trained to exhibit a specific reaction such as salivation by presenting a specific stimulus such as the sound of a bell together with food (Pavlov 1927). Bandura (1962, 1969) extended the behaviouristic approach with a social dimension by developing a theory of social learning through *imitation*. From a behaviouristic perspective, norms constitute a learned behaviour and thus have to be explained using these theories. The dynamical propensities of models inspired by game theoretical concepts are a straightforward implementation of such a view on intra agent processes. The propensity to co-operate or defect is updated in proportion to the propensity of sanctions. The propensity of sanctions, however, is a structural component resulting from inter-agent processes. Hence, agents learn to modify their behaviour according to structural conditions.

Here we find the feedback loop between social and individual components that are in fact essential for the concept of norms. However, the third component is missing: this approach does not include a concept of obligations. Deontics are out of the scope of this approach. This shortcoming can be traced back to the psychological theory that is represented in the agents: behaviourism is not capable of capturing mental processes. Indeed, it specifically avoids commenting on the mental processes involved. Under the influence of positivism, reference to unobservable entities such as the 'mind' has been regarded as not scientifically valid. Obligations are such unobservable entities. Hence, they cannot be represented by the means of behaviouristic learning theories that are applied in agent models.

In socialisation research, the complex cognitive processes necessary for grasping the meaning of an obligation is denoted as *internalisation*. It has already been shown that agent transformation is not the same as the internalisation of norms. This is also behaviourally important because normative behaviour, guided by deontics, need not be a statistical regularity, guided by propensities. In particular if moral reasoning is involved, deviant behaviour is not explained by chance variation, leading to some kind of normal distribution (where the mean value might be updated). There is a difference between norms and the normal.

To represent a complex cognitive concept such as norm internalisation calls for the cognitively rich agents of the AI tradition. However, the examination of current models has revealed that a comprehension of the cognitive mechanisms by which social behaviour regulation becomes effective in the individual mind is still in its fledgling stages. It has been shown that a multiplicity of concepts is at hand: While in the very beginning the agents where merely normative automata there exist conceptualisations of normative agent transformation ranging from updating conditionals (of strategies) through knowledge to signalling and memetic contagion. However, no consensus is reached what are the most relevant mechanisms. It can be suspected that they remain effect generating rather than process representation mechanisms. As an agenda for the next decade, a closer examination of the processes by which normative obligation become accepted by humans might be useful. For this purpose it is necessary to recall socialisation theory. This will help to clarify the problem situation. However, the results of socialisation research are not unequivocal. Hence, in the development of a normative architecture some fundamental decisions have to be made.

### *14.3.1 The Perspective of Socialisation Research on Norms*

To grasp an understanding of the decisions that have to be made in the cognitive design of an agent, first some fundamental aspects of theories of socialisation are briefly highlighted. Subsequently, current architectures of normative agents are evaluated with regard to the questions posed by the empirical science.

Broadly speaking a conceptual dichotomy of two main approaches can be identified Geulen (1991) in socialisation research concerning the relation between the individual and society: One position assumes a harmony between, or identity of, the individual and society. Philosophical precursors of this approach are Aristotle, Leibniz and Hegel. The second position stands in contrast and postulates an antagonism between the individual and society. Within this position two further standpoints can be distinguished already in the philosophical tradition: Hobbes, for example, is representative of the argument that society should tame the individual. By contrast, the position of Rousseau is paradigmatic of an approach that advocates the need for releasing the individual from society. Both philosophers share the assumption that an antagonism exists between the individual and society, although they disagree about the implications.

As it has been outlined, socialisation research sits at the border of psychology and sociology, and contributions from both disciplines can be found in the literature. From a sociological perspective, the beginning of investigating socialisation processes cumulated in the work of Emil Durkheim, founding father of sociology and professor of pedagogy. Starting from a clinical and psychological perspective, Sigmund Freud developed a theory of socialisation, which in many aspects is surprisingly akin to Durkheim's approach.

The early theories of Freud and Durkheim agree in that they assume an *antagonism* between individual and society. Durkheim asserted that the individual consists of two parts: first, a private domain that is egoistic and guided purely by basic drives. The egoistic domain corresponds to that of the newborn child. The original human is a 'tabula rasa' in which social norms have to be implemented. Only through the process of socialisation do humans become socially and morally responsible persons. This is the second 'part' of the individual. Durkheim claimed that the best of us is of a social nature. Society, however, is coercive (Durkheim 1895), and can even compel individuals to commit suicide (Durkheim 1897). Norms are finally internalised once the individual no longer perceives this coercion (Durkheim 1907). Yet for Durkheim coercion none the less remains. As with Durkheim, Freud assumed the existence of an antagonism between individuals and society. This assumption can be discerned in his distinction between Ego, Id and Super-Ego. The Id represents the drives of the child-like portion of the person. It is highly impulsive and takes into account only what it wants. It exclusively follows the pleasure principle (Freud 1932). The Super-Ego enables control of the primary drives: it represents the moral code of a society and involves feelings of shame and guilt (Freud 1955). It is the place where social norms can be found. It has been argued that the degree to which feelings of guilt are experienced is indicative of the degree of norm internalisation (Kohlberg 1996). Finally, the Ego is the controlling instance: it coordinates the demands of Id, Super-Ego and outer world. According to Freud, the Super-Ego is the mirror of the society. Freud's theory of Ego, Super-Ego and Id, then, parallels Durkheim's assumption that the internalisation of norms involves social coercion (Geulen 1991). From the perspective of both, society is in radical conflict with human nature. Norms are given as an external fact. Both Durkheim and Freud regard the individual as *passive* and internalisation as a unidirectional process.

Building on G.H. Mead (1934), and the theories of cognitive and moral development of Piaget (1932; 1947) and Kohlberg (1996), in recent times identity theories have become influential in socialisation research. In contrast to an orientation based solely either on the individual subject or the society, identity theories emphasise the interaction of culture and individuals in the development of a morally responsible person (Bosma and Kunner 2001; Fuhrer and Trautner 2005; Keupp 1999).

Mead developed a concept of identity that in contrast to Durkheim did not reduce the individual to a private subject guided purely by drives. Instead, he developed a theory of the social constitution of individual identity. A crucial mechanism in the development of personality is the capability of role taking: to regard oneself from the perspective of the other. This ability enables individuals to anticipate the perspectives and expectations of others and thereby to come to accept social norms. In the process of role taking, the individual develops a consciousness whereby the individual is itself a stimulus for the reaction of the other in situations of social interaction. This is the distinction between the spontaneous 'I' and self-reflected 'me'. Together they form what Mead denoted as identity: in other words, the 'self'. An abstraction of this process leads to the notion of the 'generalised other'. This is not a specific interaction partner but a placeholder for anybody. The notion of the 'generalised other' is the representation of society.

Identity theories follow Mead in seeing individual identity as the key link between person and culture. In contrast to the perspective to regard the social as constraining the individual, identity theories argue that socially embedded identity *enables action selection*. Action determination can be intrinsically or extrinsically motivated. The identity of individuals contributes to the development of their intrinsic motivation. There exist clear empirical evidence that sanctions and even incentives undermine intrinsic motivation (Deci and Ryan 2000). Norms, however, constitute a socially determined pattern of behaviour. Thus norm obedience is always extrinsically motivated. However, at this point *internalisation* comes into play. Extrinsic motivation can be internalised to different degrees, ranging from purely extrinsic behavioural regulation (e.g. sanctions) to motivations that are integrated into the self. Integration is attained when external guidelines have become part of personal identity. This is the highest degree of a transformation of external regulation into self-determination and is denoted as self-determined extrinsic motivation (Deci and Ryan 2000). In this case, a person is in line with itself if he or she orients behaviour around social norms. Integrated behaviour regulation is highly salient. Since norms are in full accordance with the personal values, action is regarded as autonomously motivated. Hence, the scale of internalisation from external regulation to integration is regarded as the scale from external control to *autonomy*. Internalisation of extrinsic motivation represents the bridge between psychological integrity and social cohesion. While intrinsic motivation is the paradigm of autonomous motivation, in the case of social behaviour regulation autonomy can only be reached through a process of internalisation.

How is the concept of identity described? With reference to William James (1890), criteria for identity are formulated as consistency, continuity and effectiveness. Identity consists of an *inner* and *outer perspective*. Moreover *personal* and *social identities* are differentiated (Tajfel 1970; Turner 1982; Turner and Onorato 1999). While the inner perspective is grounded on individual decisions, the outer perspective is based on ascription of others. Examples are ethnic or gender identity. However, the individual might decide to identify with these ascriptions. Then the ascription becomes part of the inner perspective. Examples can be found throughout the history. For instance, (beside other factors) elements of this psychological mechanism can be revealed in the black power movement in the 1960s, or the raise of ethnic conflicts since the 1990s. Personal identity is the self-construction of a personal biography. Social identity is determined by peer and reference groups. This refers to social networks. While peers are the group to which the individual factually belongs, the individual need not belong to the reference group. It is sufficient to identify with the values of this group. For instance, this identification might constitute sympathy for a political party. The social identity is decisively responsible for the process by which social norms and values become part of individual goals. This is particularly dependent on the salience of group membership. Norm internalisation, however, is not a unidirectional process of the transmission of a given norm. While embedded in a social environment, the individual has an *active* role in the social group.

### *14.3.2 Normative Architectures*

The brief overview of socialisation research suggests that for the design of normative agents in particular two main decisions have to be made:

> Is an antagonism or an identity (respective harmony) between individual and society presumed? Hence, does the Artificial Society represent the theories of Durkheim and Freud, or identity theories that follow G. H. Mead?

How is the effect of normative behaviour regulation on the individual agent represented? Does the individual agent play an active or a passive role, i.e. has the individual agent something comparable to a personal identity?

The second question leads to a follow-up question, namely:

If agents play an active role, if and how can this represent a process of identity formation? In particular, are agents embedded in social networks of peer or reference groups?

It will now be examined how current architectures can be assessed from an empirical perspective. As the examination of current simulation models has revealed, a comprehension of the effects of the social level on individual agents, is far from being sufficient so far. To provide an outlook of possible future modelling approaches of the effects of social norms on individual agents, a brief sample of normative agent *architectures* will be provided. In fact, the number of conceptually oriented articles on the architecture of normative agents exceeds the number of existing models. These architectures study how these processes could be modelled in principle. Typically, norms in concrete models are less sophisticated than concepts proposed in formal architectures Conte and Dignum (2001). The development of architectures is a kind of requirement analysis: it specifies the essential components of normative agents. It can be expected that future implementations will be guided by deliberations that can be found in these architectures. For this reason a sample of cases is selected (Neumann 2008a) for a closer examination with regard to the question of what decision are made how to represent effects of norms on individual agents.

Andrighetto et al. (2007) investigate the process of norm innovation. The behaviour of an agent may be interpreted by an observing agent as normative if it is marked as salient in the observer's normative board. Thus, norm instantiation is regarded as an inter-agent process.

Boella and van der Torre (2003) differentiate between three types of agents: agents who are the subject of norms, so-called defender agents, who are responsible for norm control, and a normative authority that has legislative power and that monitors defender agents.

Boella and van der Torre (2006) rely on John Searle's notion of institutional facts (so-called 'counts-as' conditionals) to represent social reality in the agent architecture. A normative base and a 'counts-as' component transforms brute facts into obligations and permissions.

The Belief-Obligation-Intentions-Desire (BOID) Architecture (Broersen et al. 2001) is the classical approach to represent norms in agent architectures. Obligations are added to the BDI Architecture to represent social norms while preserving the agent's autonomy. Principles of the resolution of conflicts between the different components are investigated in the paper.

Boman (1999) proposes the use of super-soft decision theory to characterise real-time decision-making in the presence of risk and uncertainty. Moreover, agents can communicate with a normative decision module to act in accordance with social demands. Norms act as global constraints on individual behaviour.

Castelfranchi et al. (2000) explore the principles of deliberative normative reasoning. Agents are able to receive information about norms and society. The data is processed in a multi-level cognitive architecture. On this basis, norms can be adopted and used as meta-goals in the agent decision process.

Conte and Castelfranchi (1999) distinguish between a conventionalist (in rational philosophy) and a prescriptive (in philosophy of law) perspective on norms. A logical framework is introduced to preserve a weak intuition of the prescriptive perspective which is capable of integrating the conventionalist intuition.

Conte and Dignum (2001) argue that imitation is not sufficient to establish a cognitive representation of norms in an agent. Agents infer abstract standards from observed behaviour. This allows for normative reasoning and normative influence in accepting (or defeating) and defending norms.

Dignum et al. (2002) investigate the relations and possible conflicts between different components in an agent's decision process. The decision-making process of so-called 'B-doing agents' is designed as a two-stage process, including norms as desires of society. The authors differentiate between abstract norms and concrete obligations.

Garcia-Camino et al. (2006) introduce norms as constraints for regulating the rules of interaction between agents in situations such as a Dutch auction protocol. These norms are regulated by an electronic institution (a virtual auctioneer) with an explicitly represented normative layer.

Lopez and Marquez (2004) explore the process of adopting or rejecting a normative goal in the BDI framework. Agents must recognise themselves as addressees of norms and must evaluate whether a normative goal has a higher or lower priority than those hindered by punishment for violating the norm.

Sadri et al. (2006) extend their concept of knowledge, goals and plan (KGP) agents by including norms based on the roles played by the agents. For this reason, the knowledge base KB of agents is upgraded by $KB_{soc}$, which caters for normative reasoning, and $KB_{rev}$, which resolves conflicts between personal and social goals.

Shoham and Tennenholtz (1992) propose building social laws into the action representation to guarantee the successful coexistence of multiple programs (i.e. agents) and programmers. Norms are constraints on individual freedom. The authors investigate the problem of automatically deriving social laws that enable the execution of each agent's action plans in the agent system.

Vazquez-Salceda et al. (2005) provide a framework for the normative regulation of electronic institutions. Norms are instantiated and controlled by a central institution, which must consist of a means to detect norm violation and a means to sanction norm violators and repair the system.

How can these examples be evaluated with regard to the design decision suggested by socialisation research? The existing approaches can be regarded as a hierarchy of increasingly sophisticated accounts, ranging from mere constraints to abstract concepts. Broadly speaking three concepts of norms can be differentiated: norms as constraints (the simplest choice), as obligations, or as abstract concepts (the most sophisticated choice). This can the summarised in the following table:

| Constraints | Obligations | Abstract concepts |
|---|---|---|
| Garcia-Camino et al. | Sadri et al. | Dignum et al. |
| Boman | Broersen et al. | Andrighetto et al. |
| Shoham & Tennenholtz | Boella and van der Torre (2006) | Conte & Dignum |
| | Lopez & Marquez | |
| | Boella and van der Torre (2003) | |
| | Conte & Castelfranchi | |
| | Vazquez-Salceda et al. | |
| | Castelfranchi et al. | |

(a) Constraints

The simplest and most straightforward way is to regard norms as mere constraints on the behaviour of individual agents. For example, the norm to drive on the right-hand side of the road restricts individual freedom. In this case, norms need not necessarily be recognised as such. They can be implemented off-line or can emerge in interaction processes. This may be sufficient for practical purposes. However, it follows that it is not possible to distinguish the norm from the normal. Hence, even though norms cannot be in contrast to individual desires in this account, the agents have no concept of obligations. They do not 'know' norms. Agents have a purely passive role. Since no decisions are possible they remain merely normative automata.

(b) Obligations

More sophisticated accounts treat norms as mental objects (Castelfranchi et al. 2000, Conte and Castelfranchi 1995b). This allows for deliberation about norms and, in particular, for the conscious violation of norms. Norms intervene in the process of goal generation, which might – or might not – lead to the revision of existing personal goals and the formation of normative goals. A number of accounts (such as the BOID architecture) rely on the notion of obligations. Obligations are explicit prescriptions that are always conditional to specific circumstances. One example of an obligation is not being permitted to smoke in restaurants. The rationale for including a separate obligation component next to a component of individual desires is geared towards ensuring an agent's *autonomy*: by explicitly separating individual and social desires, it is possible that the agent can deliberate over which component has priority. Conflicts may arise between different components. Compared to the literature on socialisation, a partial convergence with older theories can be observed. In particular, it is striking that Freud's architecture of the human psyche has some parallels to BOID agents: the Id, guided by egoistic drives taking into account only what it wants, can be found in the 'desires' component. Moreover, there is an obvious temptation to identify Freud's Super-Ego with the 'obligations' component. In fact, 'obligations' have been explicitly described as the desires of a society (Dignum et al. 2002). This conception is well supported by Freud's theory. With regard to current identity theories and the theory of self-determination, the situation is different: these theories emphasise that a full internalisation of norms is only realised when they have become part of one's identity. Thus, internalised norms form part of the person's own goals. According to identity theories, agents of this kind of architecture have not yet fully internalised norms. Norms, implemented in an 'obligations' component do not represent complete external regulation, but by the same token are not part of the agent's own desires. In fact, the dichotomy between obligations and desires becomes only effective once conflicts between both components arise. This is explicitly wanted: the 'obligations' component is added to the architecture to enable norm compliance as well as violation. It is claimed that this process preserves the agent's autonomy. Hence, the dichotomy of obligations and desires refers to an *antagonism* between an individual and society.

To represent the process of norm internalisation as described by modern theories, a dynamic relation between the components 'obligations' and 'desires' would be required: contingent on the salience of a norm, elements of the 'obligations' component should be imported to the 'desires' component.

(c) Abstract Concepts

Agents may face several obligations that may contradict one another. For this reason, some authors differentiate between norms and obligations. Norms are regarded as more stable and abstract concepts than mere obligations (Dignum et al. 2002; Conte and Dignum 2001). One example of such an abstract norm is 'being altruistic': further inference processes are needed for the formation of concrete action goals from this abstract norm. The striking feature of this approach is to allow for normative reasoning. This calls for an *active* role of the agent. This conception of norms is a *precondition* for a modelling approach of social behaviour regulation based on identity conceptions.

In particular the cognitive capacity of role taking constitutes a crucially important step in the development of goals from abstract concepts: that is, the ability to regard oneself from another's perspective. Interestingly, steps in this direction *can* be found in the AI literature. In Boella and van der Torre's architecture of a 'norm governed system' (2003), the agent's decision-making process is governed by the belief that they are observed by other agents and by the belief that the other agents have expectancies with regard to how they ought to behave. This can be regarded as a first step in simulating identity theory. However, from the perspective of socio-psychological identity theories, it is a shortcoming of this architecture that the agents regard themselves only in terms of the question concerning whether they fulfil their – externally given – social role. Identity consists of an inner and outer perspective. The inner perspective is dependent on one's personal decisions. This is not the case in this architecture, which consists solely of an outer perspective. It can be questioned if and how an inner perspective can be modelled: among other things, the development of an inner perspective is correlated to a *social identity*. This social identity, however, is correlated to peer groups and reference groups. Hence, it refers to social networks, which can be simulated. In principle, a propensity to take over group norms could be simulated, dependent on the salience of group membership. To model such cognitive development the agents thus need to be embedded in micro-social structures.

In conclusion, the concepts of norms as obligations and as an abstract concept are more closely related to concepts in empirical sciences than a mere constraint, which might be perfectly sufficient for practical purposes. It has to be noted, however, that they refer to *different* theories: the obligations concept of norms presumes an antagonism between the individual and the society, which is in line with Durkheim and Freud. The idea of norms as an abstract concept demands for a more active role of the agents. This is a precondition for modelling identity. There exist very first attempts that can be regarded as a modelling approach towards identity formation. Yet, it has to be emphasised that these are very first steps, and much is still not realised, such as, for instance, to implement a correlation between network structures and salience of normative orientation. However, one principle deficiency of current models and architectures in attempting to represent the process of norm internalisation remains; namely, that agents do not have a childhood (Guerin 2008). However, socialisation theory describes childhood as the most important site for the internalisation of norms. Since agents have no childhood, the process of human *cognitive development* cannot be represented.

## 14.4 Conclusion

In conclusion, it can be retained that the interaction processes, resulting in macro-structural constraints are quite well understood. In particular, the perspective to regard norms as an aggregated product of individual interactions is considerably elaborated. This is the view of sociological Rational Choice theories. In particular, the game theoretic paradigm has proved to be an effective means to study the dynamics of collective behaviour regularities. However, it lacks of an *active* element of normative orientation in the choice of the ends of action. The agents do not 'know' norms. Thus, these models do not capture the process of norm internalisation. Behaviour is merely guided by *adaptation* of agents to changing environmental conditions.

The role theoretic tradition emphasises that norms are structural constraints of individual behaviour. While models of cognitive agents in the AI tradition also have reached a substantial insights into norm dynamics, this aspect has been particularly studied these models. They have provided considerable insights into the effects of such structural constraints on a social macro-level. Hence, the inter-agent processes of interaction, leading to a macro-property of some kind of normatively structured social macro level, are relatively good understood. There is, however, still a lot to do with regard to achieving a comprehensive understanding of how actors produce, and are at the same time a product of social reality. While agent-based modelling *has reached* a substantial understanding of inter-agent processes, an investigation of the *recursive* impact on intra-agent processes is still in its fledgling stages.

This becomes apparent when considering socialisation theories of norm internalisation. An investigation of the effects of social behaviour regulation on individual agents is mostly at the level of conceptual considerations in the development of architectures. Here the development of new models has to be aware of the decisions to be made. The philosophical orientation that implicitly underlies the BOID architecture is inspired by the classical accounts of Durkheim and Freud: by opposing obligations and desires, an antagonism between individuals and society is assumed. However, it has to be (and – implicitly – is) decided whether an antagonism individual and society is assumed or not. This is the question, whether the social macro-level is perceived as action constraint or as enabling action selection. Empirical research suggests that a social embedding in networks of peer- and reference groups has a substantial impact on normative reasoning and thereby on the process of action selection (i.e. the agents' *desires*).

A comprehension of the two-way dynamics of the operations and effects of social behaviour regulation on a psychological as well as on the social level calls for interdisciplinary research. Agent-based modelling *is* an appropriate methodological tool for this effort. However, it has to be emphasised, that developmental processes in the socialisation process are only barely captured by current simulations.

## Further Reading

Even though they are quite old and some of their findings are out of date by now, it is still a good start (and not too much effort) to study the following two models to become familiar with the research field of normative agent-based models: Axelrod's (1986) evolutionary approach to norms, and Conte and Castelfranchi's (1995b) paper on understanding the functions of norms in social groups (using simulation).

As an introduction into the design and logical foundations of normative architectures the following anthologies are suggested: (Boella et al. 2005; Boella et al. 2007).

The relation of modelling and theory is particularly highlighted in the two anthologies (Conte and Dellarocas 2001; Lindemann et al. 2004). Here the reader will also find hints for further readings about the empirical and theoretical background.

For an overview of the theoretical background and developments in theorising norms it is suggested to refer to (Conte and Castelfranchi 1995a; Therborn 2002).

## References

Andrighetto G, Campennì M, Conte R, Paolucci M (2007) On the immergence of norms: a normative agent architecture. In: Trajkovski GP, Collins SG (eds) Emergent agents and socialities: social and organizational aspects of intelligence; proceedings of the 2007 fall AAAI symposium. AAAI Press, Menlo Park, pp 11–18

Axelrod R (1986) An evolutionary approach to norms. Am Polit Sci Rev 80:1095–1111

Bandura A (1962) Social learning through imitation. In: Jones M (ed) Nebraska symposium on motivation. University of Nebraska Press, Lincoln

Bandura A (1969) Principles of behaviour modification. Holt, New York

Bicchieri C, Duffy J, Tolle G (2003) Trust among strangers. Philos Sci 71:286–319

Binmore K (1998) Review of the book "The complexity of cooperation: agent-based models of competition and collaboration" by Axelrod R, Princeton, Princeton University Press. J Artif Soc Soc Simulat 1(1), http://jasss.soc.surrey.ac.uk/1/1/review1.html

Boella G, van der Torre L (2003) Norm Governed multiagent systems: the delegation of control to autonomous agents. In proceedings of the IEEE/WIC international conference on intelligent agent technology 2003 (IAT 2003). IEEE Press, pp 329–335

Boella G, van der Torre L (2006) An architecture of a normative system: counts-as conditionals, obligations, and permissions. In: *AAMAS '06:* Proceedings of the fifth international joint conference on autonomous agents and multiagent systems. ACM Press, New York, pp 229–231

Boella G, van der Torre L, Verhagen H (eds) Proceedings of the symposium on Normative Multi-Agent Systems (NORMAS 2005), AISB'05 convention: social intelligence and interaction in animals, robots and agents, 12–15 Apr 2005, University of Hertfordshire, Hatfield. Hatfield: society for the Study of Artificial Intelligence and the Simulation of Behaviour (SSAISB)
Boella G, van der Torre L, Verhagen H (2007) Introduction to normative multiagent systems. In: Boella G, van der Torre L, Verhagen H (eds) Normative multi-agent systems, Dagstuhl seminar proceedings 07122, vol I. IBFI, Schloss Dagstuhl, pp 19–25
Boman M (1999) Norms in artificial decision making. Artif Intell Law 7:17–35
Bosma H, Kunner E (2001) Determinants and mechanisms in ego development: a review and synthesis. Dev Rev 21(1):39–66
Boudon R (1981) The logic of social action. Routledge, London
Bratman M (1987) Intentions, plans and practical reasoning. CSLI Publications, Stanford
Broersen J, Dastani M, Huang Z, van der Torre L (2001) The BOID architecture: conflicts between beliefs, obligations, intentions, and desires. In: André E, Sen S, Frasson C, Müller JP (eds) Proceedings of the 5th international conference on autonomous agents. ACM Press, New York, pp 9–16
Burke M, Fournier G, Prasad K (2006) The emergence of local norms in networks. Complexity 11:65–83
Castelfranchi C, Conte R, Paolucci M (1998) Normative reputation and the costs of compliance. J Artif Soc Soc Simulat 1(3), http://www.soc.surrey.ac.uk/JASSS/1/3/3.html
Castelfranchi C, Dignum F,Treur J (2000) Deliberative normative agents: principles and architecture. In: Jennings NR, Lesperance Y (eds) Intelligent agents VI, agent theories, architectures, and languages (ATAL), 6th international workshop, ATAL '99, Orlando, July 15–17, 1999, proceedings (LNCS, 1757). Springer, Berlin, pp 364–378
Claessens D (1972) Familie und wertsystem: eine studie zur zweiten sozio-kulturellen geburt des menschen. Duncker & Humblot, Berlin
Coleman J (1987) The emergence of norms in varying social structures. Angew Sozialforschung 14:17–30
Coleman J (1990) Foundations of social theory. Harvard, Belknap
Conte R, Castelfranchi C (1995a) Cognitive and social action. UCL Press, London
Conte R, Castelfranchi C (1995b) Understanding the functions of norms in social groups through simulation. In: Gilbert N, Conte R (eds) Artificial societies: the computer simulation of social life. UCL Press, London, pp 252–267
Conte R, Castelfranchi C (1999) From conventions to prescriptions: towards an integrated view of norms. Artif Intell Law 7:323–340
Conte R, Castelfranchi C (2001) Are incentives good enough to achieve (info) social order? In: Conte R, Dellarocas C (eds) Social order in multiagent systems. Kluwer, Dordrecht
Conte R, Dellarocas C (eds) (2001) Social order in multiagent systems. Kluwer, Dordrecht
Conte R, Dignum F (2001) From social monitoring to normative influence. J Artif Soc Soc Simulat 4(2), http://www.soc.surrey.ac.uk/JASSS/4/2/7.html
Dahrendorf R (1956) Homo sociologicus. Ein Versuch zu Geschichte, Bedeutung und Kritik der Kategorie der sozialen Rolle. Westdeutscher Verlag, Opladen
Davies K, Moore W (1945) Some principles of stratification. Am Sociol Rev 10:242–249
Deci E, Ryan R (2000) The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 11(4):227–268
Deffuant G, Moss S, Jager W (2006) Dialogues concerning a (Possibly) new science. J Artif Soc Soc Simulat 9(1), http://jasss.soc.surrey.ac.uk/9/1/1.html
Dignum F, Kinny D, Sonenberg L (2002) From desires, obligations and norms to goals. Cognitive science quarterly, 2, http://www.cs.uu.nl/people/dignum/papers/CSQ.pdf
Durkheim E ([1895] 1970) Regeln der Soziologischen Methode. Neuwied: Luchterhand
Durkheim E ([1897] 2006) Der Selbstmord. Frankfurt a.M.: Suhrkamp
Durkheim E ([1907] 1972) Erziehung und Soziologie. Düsseldorf: Schwann
Epstein J (2000) Learning to be thoughtless: social norms and individual computation (working paper, no. 6). Center on social and economic dynamics, http://www.santafe.edu/media/workingpapers/00-03-022.pdf

Esser H (1993) Soziologie: allgemeine grundlagen. Campus, Frankfurt a. M
Fioretti G (2013) Utility, games, and narratives. Chapter 13 in this volume
Flentge F, Polani D,Uthmann T (2001) Modelling the emergence of possession norms using memes. J Artif Soc Soc Simulat 4(4), http://jasss.soc.surrey.ac.uk/4/4/3.html
Freud S (1932) Neue Vorlesungen zur Einführung in die Psychoanalyse. In: Freud S, Gesammelte Werke, vol. 15. London: Imago
Freud S (1955) Abriss der psychoanalyse. Fischer, Frankfurt
Fuhrer U, Trautner N (2005) Entwicklung und identität. In: Asendorpf J (ed) Enzyklopädie der pschologie – entwicklungspsychologie, vol 3. Göttingen, Hofgrebe
Garcia-Camino A, Rodriguez-Aguilar JA, Sierra C, Vasconcelos W (2006) Norm-oriented programming of electronic institutions: a rule-based approach. In: Nakashima H, Wellman M, Weiss G, Stone P, Proceedings of the 5th international joint conference on autonomous agents and multi-agent systems 2006, Hakodate, 08–12 May 2006. ACM Press, New York, pp 670–672
Geulen D (1991) Die historische entwicklung sozialisationstheoretischer ansätze. In: Hurrelmann K, Ulich D (eds) Neues handbuch der sozialisationsforschung. Beltz, Weinheim
Giardini F, Conte R, Paolucci M (2013) Reputation. Chapter 15 in this volume
Gibbs JP (1981) Norms, deviance and social control: conceptual matters. Elsevier, New York
Guerin F (2008) Constructivism in AI: prospects, progress and challenges. In proceedings of the AISB convention 2008, vol 12: computing and philosophy, Aberdeen, Scotland, 1–4 Apr 2008. London: The society for the study of artificial intelligence and simulation of behaviour, pp 20–27
Hales D (2002) Group reputation supports beneficent norms. J Artif Soc Soc Simulat 5(4), http://jasss.soc.surrey.ac.uk/5/4/4.html
Homans G (1964) Bringing man back in. Am Sociol Rev 29:809–818
James W (1890) The principles of psychology. Holt, New York
Keupp H (1999) Identitätskontruktionen. Rowohlt, Hamburg
Kohlberg L (1996) Die psychologie der moralentwicklung. Suhrkamp, Frankfurt
Krappmann L (2006) Sozialisationsforschung im spannungsfeld zwischen gesellschaftlicher reproduktion und entstehender handlungsfähigkeit. In: Schneider W, Wilkening F (eds) Enzyklopädie der psychologie – entwicklungspsychologie, vol 1. Göttingen, Hofgrebe
Lindemann G, Moldt D, Paolucci M (eds) (2004) Regulated agent-based social systems: first international workshop, RASTA 2002, vol 2934, Lecture notes in artificial intelligence. Springer, Berlin
Lopez F, Marquez A (2004) An architecture for autonomous normative agents. In: proceedings of the 5th Mexican international conference in computer science, *ENC'04*. IEEE computer society, Los Alamitos, pp 96–103
Macy M, Sato Y (2002) Trust, cooperation, and market formation in the U.S. and Japan. PNAS, 99, pp 7214–7220
Macy M, Willer R (2002) From factors to actors: computational sociology and agent-based modelling. Am Rev Sociol 28:143–166
Malinowski B (1927) Sex and repression in primitive society. Humanities Press, New York
Mead M (1928) Coming of age in Samoa. William Morrow & Co., New York
Mead GH ([1934] 1968) Geist, Identität und Gesellschaft. Suhrkamp, Franfurt
Merton R (1957) Social theory and social structure, 2nd edn. The Free Press, Glencoe
Moss S (2001) Game theory: limitations and an alternative. J Artif Soc Soc Simulat 4(2), http://www.soc.surrey.ac.uk/JASSS/4/2/2.html
Neumann M (2008a) A classification of normative architectures. In: Deffuant G, Cioffi-Revilla C (eds) Proceedings of the second world congress of social simulation (WCSS'08), Fairfax, 14–17 July 2008
Neumann M (2008b) Homo socionicus: a case study of normative agents. J Artif Soc Soc Simulat 11(4), http://jasss.soc.surrey.ac.uk/11/4/6.html
Parsons T ([1937] 1968) The structure of social action: a study in social theory with special reference to a group of recent European writers. New York: Free Press

Pawlow IP (1927) Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. Oxford University Press, London

Piaget J ([1932] 1983) Das moralische Urteil beim Kinde. Stuttgart: Klett-Cotta

Piaget J ([1947] 1955) Psychologie der Intelligenz. Rascher, Zürich

Rao AM, Georgeff M (1991) Modeling rational agents within a BDI architecture. In: Allen JF, Fikes R, Sandewall E (eds) Proceedings of the 2nd international conference on principles of knowledge representation and reasoning (KR'91), Cambridge, 22–25 Apr 1991. Morgan Kaufmann, San Francisco, pp 473–484

Raub W, Voss T (1981) Individuelles handeln und gesellschaftliche folgen: das individualistische programm in den sozialwissenschaften. Luchterhand, Darmstadt

Saam N, Harrer A (1999) Simulating norms, social inequality, and functional change in artificial societies. J Artif Soc Soc Simulat 2(1), http://jasss.soc.surrey.ac.uk/2/1/2.html

Sadri F, Stathis K, Toni F (2006) Normative KGP agents. Comput Math Organ Theor 12:101–126

Savarimuthu B, Purvis M, Cranefield S, Purvis MA (2007) How do norms emerge in multi-agent societies? Mechanism design (The information science discussion paper series, 2007/01). University of Otago, Dunedin

Sen S, Airiau S (2007) Emergence of norms through social learning. In: Veloso M (ed) Proceedings of the twentieth joint conference on artificial intelligence (IJCAI-07), Hyderabad, 6–12 Jan 2007. pp 1507–1512

Shoham Y, Tennenholtz M (1992) On the synthesis of useful social laws for artificial agent societies (preliminary report). In: Swartout WR (ed) Proceedings of the 10th National conference on artificial intelligence (AAAI'92), 12–16 July 1992 AAAI Press/The MIT Press, San Jose, pp 276–281

Skyrms B (1996) Evolution of the social contract. Cambridge University Press, Cambridge

Skyrms B (2004) The stage hunt and the evolution of social structure. Cambridge University Press, Cambridge

Tajfel H (1970) Experiments in intergroup discrimination. Sci Am 223(5):96–102

Therborn G (2002) Back to norms! on the scope and dynamics of normative action. Curr Sociol 50 (6):863–880

Turner JC (1982) Towards a cognitive redefinition of the social group. In: Tajfel H (ed) Social identity and intergroup relations. Cambridge University Press, Cambridge

Turner JC, Onorato RS (1999) Social identity, personality, and the self-concept: a self categorising perspective. In: Tylor T et al (eds) The psychology of the social group. Erlbaum, Mahwah

Vazquez-Salceda J, Aldewereld H, Dignum F (2005) Norms in multiagent systems: from theory to practice. Int J Comput Syst Engin 20:225–236

Verhagen H (2001) Simulation of the learning of norms. Soc Sci Comput Rev 19:296–306

Vieth M (2003) Die evolution von fairnessnormen im ultimatumspiel: eine spieltheoretische modellierung. Z Soziol 32:346–367

Wrong D (1961) The oversocialised conception of man. Am Sociol Rev 26:183–193

# Chapter 15
# Reputation

**Francesca Giardini, Rosaria Conte, and Mario Paolucci**

**Why Read This Chapter?** To understand the different conceptions underlying reputation in simulations up to the current time and to get to know some of the approaches to implementing reputation mechanisms, which are more cognitively sophisticated.

**Abstract** In this chapter, the role of reputation as a distributed instrument for social order is addressed. A short review of the state of the art will show the role of reputation in promoting (a) social control in cooperative contexts – like social groups and subgroups – and (b) partner selection in competitive contexts, like (e-) markets and industrial districts. In the initial section, current mechanisms of reputation – be they applied to electronic markets or MAS – will be shown to have poor theoretical backgrounds, missing almost completely the cognitive and social properties of the phenomenon under study. In the rest of the chapter a social cognitive model of reputation developed in the last decade by some of the authors will be presented. Its simulation-based applications to the theoretical study of norm-abiding behaviour, partner selection and to the refinement and improvement of current reputation mechanisms will be discussed. Final remarks and ideas for future research will conclude the chapter.

## 15.1 Reputation in Social Systems: A General Introduction

Ever since hominid settlements started to grow, human societies needed to cope with the problem of social order (Axelrod, 1984): how to avoid fraud and cheating in wider, unfamiliar groups? How to choose trustworthy partners when the likelihood of re-encounter is low? How to isolate cheaters and establish worthwhile alliances with the good guys?

F. Giardini (✉) • R. Conte • M. Paolucci
Laboratory for Agent-Based Social Simulation (LABSS), Institute of Cognitive Sciences and Technologies, National Research Council, Via Palestro, 32 – 00185, Rome
e-mail: francesca.giardini@istc.cnr.it; rosaria.conte@istc.cnr.it; mario.paolucci@istc.cnr.it

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_15,

Social knowledge like reputation and its transmission, i.e. gossip, plays a fundamental role in social order, adding at the same time cohesiveness to social groups and allowing for distributed social control and sanctioning (plus a number of other functionalities, see Boehm 1999). Reputation is a property that unwilling and unaware individuals derive from the generation, transmission and manipulation of a special type of social beliefs, namely social evaluations, and that contributes to regulate natural societies from the morning of mankind (Dunbar 1996). People use reputational information to make decisions about possible interactions, to evaluate candidate partners, to understand and predict their behaviours, and so on.

That reputation is a fundamental generator, vehicle and manipulator of social knowledge for enforcing reciprocity and other social norms is known since long (see a review in Conte and Paolucci 2002). In particular, in the study of cooperation and social dilemmas, the role of reputation as a partner selection mechanism started to be appreciated in the early 1980s (Kreps and Wilson 1982). However, little understanding of its dynamic and cognitive underpinnings was achieved at that stage. Despite its critical role in the enforcement of altruism, cooperation and social exchange, the social cognitive study of reputation is relatively new. Hence, it has not yet been fully clarified how this critical type of knowledge is manipulated in the minds of agents, how social structures and infrastructures generate, transmit and transform it, and consequently how it affects agents' behaviour.

The aim of this chapter is to guide the reader through the multiplicity of computational approaches concerned with the impact of reputation and its dynamics. Reputation is a complex social phenomenon that cannot be treated as a static attribute of agenthood, with no regard for the underlying process of transmission. We claim that reputation is both the process and the effect of transmitting information, and that further specifications about the process and its mechanisms are needed. Following these premises, we will first review some applications of reputation in computational simulation, highlighting problems and open questions, and then we will propose a theoretical social cognitive model of reputation. Moreover, we will present three different ways of applying the cognitive theory of reputation to model social phenomena: the Sim-Norm model, the SOCRATE framework and the REPAGE architecture.

This brief introduction will be followed by an outline of reputation research in different domains (social psychology, management and experimental economics, agent-based simulation), in order to show how many different viewpoints can be used to describe and explore this complex phenomenon. We will then focus on some of the results in electronic markets and multi-agent simulations. Electronic markets are a typical example of a complex environment where centralized control is not possible and decentralized solutions are far from being effective. In recent years, the Internet contributed to a growth of auction sites facilitating the exchange of goods between individual consumers, without guaranteeing transparency and safety of transactions. On the other hand, multi-agent applications are concerned with the problem of assessing the reliability of single agents and of social networks. In Sect. 15.2 we will propose a cognitive model of reputation, which aims to solve some of the problems left open by existing systems, moving from a theoretical analysis of cognitive underpinnings of reputation formation and spreading. This model will be tested in the following section, where a description of three different implementations will be provided. Finally, we will draw some conclusions about future work and directions.

### *15.1.1 State of the Art*

According to Frith and Frith (2006), there are three ways to learn about other people: through direct experience, through observation and through “cultural information”. When the first two modalities are not available, reputational information becomes essential in order to obtain some knowledge about one’s potential partner(s) in an interaction, and thus to predict their behaviour. Reputation allows people to predict, at least partially or approximately, what kind of social interaction they can expect and how that interaction may possibly evolve. Reputation is therefore a coordination device whose predictive power is essential in social interactions (Paolucci and Conte 2009). Furthermore, reputation has a strategic value and can be used to pursuit self-interest (Paine 1967; Noon and Delbridge 1993).

Reputation and its transmission (gossip) has an extraordinary preventive power: it substitutes personal experience in (a) identifying cheaters and isolating them, and in (b) easily finding trustful partners. It makes available all the benefits of evaluating someone, without implying the costs of direct interaction.

Furthermore, in human societies gossip facilitates the formation of groups (Gluckman 1963): gossipers share and transmit relevant social information about group members within the group (Barkow 1996), at the same time isolating out-groups. Besides, gossip contributes to stratification and social control, since it works as a tool for sanctioning deviant behaviours and for promoting, even through learning, those behaviours that are functional with respect to the group’s goals and objectives. Reputation is also considered as a means for sustaining and promoting the diffusion of norms and norm conformity (Wilson et al. 2000).

The theories of indirect reciprocity and costly signals show how cooperation in large groups can emerge when the agents are endowed with, or can build, a reputation (Nowak and Sigmund 1998a; 1998b; Gintis et al. 2001). As Alexander (1987) pointed out, “indirect reciprocity involves reputation and status, and results in everyone in the group continually being assessed and reassessed”. According to this theory large-scale human cooperation can be explained by individuals helping others in order to uphold a reputation and thus be included in future cooperation (Panchanathan and Boyd 2004).

Despite important advances in the study of reputation as a means to support cooperation (Sommerfeld et al. 2008), no explicit theory of the cognitive ingredients and processes which reputation is made of was provided. More recently, reputation and gossip has started to become crucial in other fields of the social sciences like management and organisation science, governance, or business ethics, where the importance of branding became apparent. The economic interest in the subject matter implied an extension of reputation to the super-individual level: Corporate reputation is considered as an external and intangible asset tied to the history of a firm and coming from stakeholders' and consumers' perceptions. Rose and Thomsen (2004) claim that good reputation and financial performance are mutually dependent, hence a good reputation may influence the financial asset of a firm and vice versa. Several researchers have tried to create a corporate reputation index containing the most relevant dimensions to take into account when dealing with corporate reputation. Cravens et al. (2003) interviewed 650 CEO in order to create a reliable index, but their index has so many entries, ranging from global strategy to employees' attributes, that it is not easy to foresee how such a tool could be used. Gray and Balmer (1998) distinguish between corporate image and corporate reputation. Corporate image is the mental picture consumers hold about a firm, therefore is similar to an individual perception, whereas the reputation results from the firm's communication and long-term strategy. Generally speaking, corporate reputation is treated as an aggregate evaluation stakeholders, consumers, managers, employees, and institutions form about a firm, but the mechanisms leading to the final result are still only vaguely defined.

Over the last 10 years several studies in experimental economics have investigated reputational dynamics through standard experimental settings, such as trust games, public good games, and moral hazard problems (see Fehr and Gachter 2000, for an introduction). The aim of these studies is to explain mechanisms underlying reciprocity, altruism and cooperation in humans, in order to answer the puzzling question: "Why do humans cooperate?" Axelrod (1984). According to Nowak and Sigmund (1998a, b), reputation sustains the emergence of indirect reciprocity, which gets people to cooperate in order to receive cooperation even from strangers. Following this hypothesis, holding a good reputation in your social group makes it more probable that someone else would help you when you will need help.

If social order is a constant of human evolution, it becomes particularly crucial in the e-society where the boundaries of interaction are extensively widened. The portentous development pace of ICT technologies dramatically enlarges the range of interaction among users, generating new types of aggregation, from civic communities to electronic markets, from professional networking to e-citizenship, etc. What is the effect of this widening of social boundaries? Communication and interaction technologies modify the range, structures and modalities of interaction, with consequences that are only partially explored, often only to resume the stereotype of technological unfriendliness (see the negative impact of computer terminals, as opposed to face-to-face interaction, on subjects' cooperativeness in experimental studies of collective and social dilemmas, Sell and Wilson 1991; Rocco and Warglien 1995). A detailed approach to the effects of technological infrastructures on interaction styles and modes has never been adopted. Perhaps, an exception to this is represented by the research on the effects of asymmetry of information on the markets. Known to encourage fraud and low-quality production, this phenomenon appears as an intrinsic feature of e-markets, but in fact it goes back to eleventh century Maghribi traders moving along the coast of the Mediterranean sea (Greif 1993). As exemplified by Akerlof (1970), asymmetry of information drives honest traders and high quality goods out of the market. The result is a market where only "lemons", or fraudulent commodities, are available – often to the detriment of both sellers and buyers. The classical example of a lemon market is the used car market, where only sellers have information about problems with the cars they are selling, and most consumers are incapable of discerning these problems. Contemporary online traders such as users of Internet auction sites face the same problem: online buyers can learn about the quality (or condition) of the good only once they have already paid for it.

Auction sites may be very generic concerning the products being offered and operate on a global scale (e.g., e-Bay), or may focus on specific products on a national scale (many car auction sites). At the moment virtually all consumer products are being auctioned on the Internet, ranging from used toys and CDs to cars and houses Utz et al. (2009). Compared to buying through online retailers, buying through auction sites is even less controllable, as the sellers are not visible and have not made major investments. Consumers who purchase through auction sites must rely on the accuracy and reliability of the seller (Melnik and Alm 2002). Sellers on the Internet may actively try to communicate their reputation to potential buyers, inflating common expectations on the effects of reputation. Melnik and Alm (2002) investigated whether an e-seller's reputation matters. Their results indicated that reputation had a positive – albeit relatively small – impact on the price levels consumers were willing to pay. Moreover, Yamagishi et al. (2009) show that reputation has significant positive effects on the quality of products.

Despite the role of reputation in economic transactions, online reputation reporting systems are only moderately efficient (Resnick and Zeckhauser 2002; Bolton et al. 2002). eBay, one of the best (if not the best) known examples, is also one of the world's largest online marketplaces with a community of over 50 million users registered. On eBay, most items are sold through English-type auctions, and the reputation mechanism used is based on the ratings that users perform after the completion of transactions. The user can give three possible values: positive (1), negative ($-1$) or neutral (0). The "reputation" value is the sum of the last 6 months' ratings, weighted by the relevance of the transaction.

In all of these models, the notion of reputation is weak and essentially reduced to centralized image: no direct exchange of information takes place among participants but only reports to a central authority, which calculates the final reputation value. The actual utility of this mechanism is debatable. For example, when forums are available, real reputation exchanges are performed in parallel, ignoring the centrally calculated reputation rating. Moreover, people are likely to not provide reputational feedback (under-provision) and if they do, they lean on providing only good reports (over-scoring). In sum, the reputation system does not seem to perform efficiently.

However, from an economic point of view, eBay prospers. How is this possible? What would happen if reputation worked more efficiently? Which type of reputation system should be used for which objective? As current online reputation systems are not theory-driven instruments, based upon an understanding of what reputation is, how it is generated and distributed, and how it works, all of these questions are still open.

However, the effects of reputation are at least partially known, generally restricted to partner selection. Indeed, agent-based social simulation has taught us some lessons: (1) what matters about reputation is its transmission (Castelfranchi et al. 1998), since by this means agents acquire zero-cost relevant information; (2) reputation has more impact than image: if agents transmitted only their own evaluations about one another (image), the circulation of social knowledge would stop soon (Pinyol et al. 2008). To exchange information about reputation, agents need to participate in circulating reputation whether they believe it or not (gossip) and, to preserve their autonomy, they must decide how, when and about whom to gossip. What is missing in the study of reputation is the merging of these separate directions in an interdisciplinary integrated approach, which accounts for both its social cognitive mechanisms and structures.

### *15.1.2 Simulating Reputation: Current Systems*

So far, the simulation-based study of reputation has been undertaken for the sake of social theory, namely in the account of pro-social behaviour – be it cooperative, altruistic, or norm-abiding – among autonomous, i.e. self-interested agents.

In this chapter, we will concentrate instead on two specific functionalities of reputation:

1. To promote norm-abiding behaviour in cooperative settings (e.g. social groups).
2. To favour partner selection in electronic markets and agentized environments.

The second aspect has been dealt with in the literature to some extent and the remainder of this section will give an overview of existing approaches; for a treatment of the first aspect we refer the reader to Sect. 15.2, where we mainly discuss our own models and studies of reputation.

Several attempts have been made to model and use reputation in artificial societies, especially in two sub-fields of information technologies: computerized interaction (with a special reference to electronic marketplaces) and agent-mediated interaction. It is worth emphasizing that in these domains trust and reputation are actually treated as the same phenomenon, and often the fundamentals of reputation mechanisms are derived from trust algorithms. Moreover, several authors (Moukas et al. 1999; Zacharia 1999; Zacharia et al. 1999) explain reputation in terms of trust and vice versa, continuously mixing up these two phenomena. We will review some of the main contributions in online reputation reporting systems and in multi-agent systems, in order to achieve a better understanding of the complex issue of implementing and effectively using reputation in artificial societies.

#### 15.1.2.1 Online Reputation Reporting Systems

The continuously growing volume of transactions on the World Wide Web and the growing number of frauds this appears to entail[1] led scholars from different disciplines to develop new online reputation reporting systems. These systems are meant to provide a reliable way to deal with reputation scores or feedbacks, allowing agents to find cooperative partners and avoid cheaters.

The existing systems can be roughly divided into two sub-sets: Agent-oriented individual approaches and agent-oriented social approaches, depending on how agents acquire reputational information about other agents.

The *agent-oriented individual approach* has been dominated by Marsh's ideas on trust (Marsh 1992, 1994a, b), on which many further developments and algorithms are based. This kind of approach is characterized by two attributes: (1) any agent may seek potential cooperation partners, and (2) the agent only relies on its experiences from earlier transactions. When a potential partner proposes a transaction, the recipient calculates the "situational reputation" by weighing the reputation of his potential trading partner with further factors, such as potential output and the importance of the transaction. If the resulting value is higher than a certain "cooperation threshold", the transaction takes place and the agent updates the reputation value according to the outcomes of the transaction. If the threshold is not reached, the agent rejects the transaction offer, which may be punished by a "reputation decline". These individual-based models (Bachmann 1998; Marsh 1994a; Ripperger 1998) differ with regard to the length of memory span they apply. Agents may forget their experiences slowly, fast, or never.

In *agent-oriented social approaches* agents not only rely on their direct experience, but are also allowed to consider third-party information (Abdul-Rahman and Hailes 1997a; Rasmusson 1996; Rasmusson and Janson 1996; Schillo 1999; Yu and Singh 2000). Although these approaches share the same basic idea, i.e. experiences of other agents in the network can be used when searching for the right transaction partner, they rely upon different solutions when it comes to weigh the third-party information and to deal with "friends of friends". Thus the question arises: how to react to information from agents who do not seem to be very trustworthy?

Another problem lies in the storage and distribution of information. To form a complete picture of its potential trading partners, each agent needs both direct (its own) and indirect (third-party) evaluations in order to be able to estimate the validity and the informational content of such a picture.

[1] The US-based Internet Crime Complaint Center (IC3) received 231,493 complaints for the year 2005, 62.7 % of which were related to electronic auctioning (IC3 2005).

Regan and Cohen (2005) propose a system for computing indirect and direct reputation in a computer mediated market. Buyers rely on reputation information about sellers when choosing from whom to buy a product. If they do not have direct experience from previous transactions with a particular seller they take indirect reputation into account by asking other buyers for their evaluations of the potential sellers. The received information is then combined to mitigate effects of deception. The objective of this system is to propose a mechanism which reduces the "undesirable practices" on actual reputation in online applications, especially on the part of sellers, and to prevent the market from turning into a "lemons market" where only low quality goods are listed for sale.

One serious problem with this and similar models concerns the reputation transmission. Agents only react to reputation requests, while proactive, spontaneous delivery of reputation information to selected recipients is not considered. On the other hand, despite its simplicity, this type of model tackles the problem of collusion between rating agents by keeping secret the evaluation of sellers amongst buyers, i.e. not disclosing it to the sellers.

As to electronic marketplaces, classic systems like eBay show a characteristic bias to positive evaluations (Resnick and Zeckhauser 2002), suggesting that factual cooperation among users at the information level may lead to a "courtesy" equilibrium (Conte and Paolucci 2003). As Cabral and Hortaçsu (2006) formally prove, initial negative feedbacks trigger a decline in sale price that drives the targeted sellers out of the market. Good sellers, however, can gain from 'buying a reputation' by building up a record of favourable feedback through purchases rather than sales. Thus those who suffer a bad reputation stay out – at least until they decide to change identity – while those who stay in can take advantage of a good reputation: after a good start, they will hardly receive negative feedback and even if they do, negative feedbacks will not get to the point of spoiling their good name. Under such conditions, even good sellers may have an incentive to sell lemons.

Intuitively, the courtesy equilibrium reduces the deterrent effect of reputation. If a reputation system is meant to impede frauds and improve the quality of products, it needs to be constructed in such a way as to avoid the emergence of a courtesy equilibrium. It is not by chance that among the possible remedies to ameliorate eBay, Dellarocas (2003) suggested a short-memory system, erasing all feedbacks but the very last one. While this might work for eBay, we believe that such remedies based on fragmented models and tailored to a particular application are not the way forward. Instead, a general theory of how reputation and its transmission work needs to be developed. On top of such a theory, different systems for different objectives can then be constructed. We will pursue this further in Section 15.3.

#### 15.1.2.2 MAS Applications

Models of trust and reputation for multi agent systems applications (e.g. Yu and Singh 2000; Carbo et al. 2002; Sabater and Sierra 2002; Schillo et al. 2000; Huynh et al. 2004; for exhaustive reviews see Ramchurn et al. 2004a; Sabater and Sierra 2005) present interesting new ideas and advances over conventional online reputation systems, with their notion of centralized global reputation.

Yu and Singh (2000) proposed an agent-oriented model for social reputation and trust management which focuses on electronic societies and MAS. Their model introduces a gossip mechanism for informing neighbours of defective transaction partners, in which the gossip is transferred incrementally through the network of agents. It also provides a mechanism to include other agents' testimonies in an agent's reputation calculation. Agents store information about the outcome of every transaction they had and recall this information in case they are planning to bargain with the same agent again (direct evaluation). If the agent meets an agent it has not traded with before, the reputation mechanism comes into play. In this mechanism, so-called referral chains are generated that can make third-party information available across several intermediate stations. An agent is thus able to gain reputation information with the help of other agents in the network. Since a referral chain represents only a small part of the whole network, the information delivered will most likely be partial instead of global as in centralized systems like eBay.

In the context of several extensive experiments, Yu and Singh showed that the implementation of their mechanism results in a stable system, in which the reputation of cheaters decreases rapidly while the cooperating agents experienced a slow, almost linear increase in reputation. Still, some problems remain. The model does not combine direct and indirect reputation, i.e. if an agent has already traded with another agent he has to rely on his own experience and cannot use the network information anymore. Thus it might take unnecessarily long to react to a suddenly defecting agent that cooperated before. In addition, Singh and Yu do not give an explanation of how their agent-centred storage of social knowledge (for example the referral chains) is supposed to be organized. Consequently, no analysis of network load and storage capacity can be done.

As this example shows, the "agentized environment" is likely to produce interesting solutions that may apply also to online communities. This is so for two main reasons. First, in this environment two problems of order arise: to meet the users' expectations (external efficiency), and to control agents' performance (internal efficiency). Internal efficiency is instrumental to the external one, but it re-proposes the problem of social control at the level of the agent system. In order to promote the former, agents must control, evaluate, and act upon each other. Reliability of agents is a proxy for reliability of users. Secondly, and consequently, the agent system plays a double role: it is both a tool and a simulator. In it one can perceive the consequences of given premises, which may be transferred to the level of user interactions. In a sense, implemented agent systems for agent-mediated interaction represent both parallel and nested sub-communities.

As a consequence, solutions applied to the problems encountered in this environment are validated more strictly, against both external and internal criteria. Their effects are observable at the level of the virtual community, with a procedure essentially equivalent to agent-based simulation and with the related advantages. Moreover, solutions may not (only) be implemented between agents, but (also) within agents, which greatly expands the space for modelling. So far, however, these potentials have not been fully exploited. Models have mainly been aimed at ameliorating existing tools implemented for computerized markets. We suggest that agent systems can do much more than this: they can be applied to answer the question as to (a) what type of agent, (b) what type of beliefs, and (c) what type of processes among agents are required to achieve useful social control. More specifically, what type of agents and processes are needed for which desirable result: better efficiency, encouraging equity and hence users' trust, discouraging either positive or negative discrimination (or both), foster collaboration at the information level or at the object level (or both), etc.

### *15.1.3 Concluding Remarks*

This review has given an overview of how reputation have been discussed and modelled in studies regarding online markets and multi-agent systems.

In the Internet-based models, the notion of reputation is weak and essentially reduced to centralized image: participants do not exchange information directly but only report their evaluations to a central authority, which calculates the final reputation value. The actual utility of this mechanism is debatable.

The solutions proposed for MAS systems are interesting but so far insufficient to meet the problems left open by online systems. There is a tendency to consider reputation as an external attribute of agents without taking into account the process of creation and transmission of that reputation.

We argue that a more theory-driven approach is needed, based upon a conceptual analysis of the differences and analogies among notions concerning social evaluation. In the next section we will therefore introduce our social cognitive approach towards reputation.

## 15.2 An Alternative Approach: The Social Cognitive Theory of Reputation

In this section we will present a social cognitive model of reputation, we will define the difference between image and reputation, introduce the roles different agents play when evaluating someone and transmitting this evaluation and, finally, we will explain the decision processes underlying reputation.

Let us first clarify the term "social cognitive". A cognitive process involves symbolic mental representations (such as goals and beliefs) and is effectuated by means of the mental operations that agents perform upon these representations (reasoning, decision-making, etc.). A social cognitive process is a process that involves social beliefs and goals, and that is effectuated by means of the operations that agents perform upon social beliefs and goals (e.g. social reasoning). A belief or a goal is social when it mentions another agent and possibly one or more of his or her mental states (for an in-depth discussion of these notions, see Conte and Castelfranchi 1995; Conte 1999).

Social cognitive processes are receiving growing attention within several subfields of the Sciences of the Artificial, in particular intelligent software agents, multi-agent systems, and artificial societies. Unlike the theory of mind (cf. Leslie 1992), which focuses upon one, albeit important aspect of social agency, namely social beliefs, this approach aims at modelling and possibly implementing systems acting in a social – be it natural or artificial – environment. Thus, it is aimed at modelling the variety of mental states (including social goals, motivations, obligations) and operations (such as social reasoning and decision-making) necessary for an intelligent social system to act in some domain and influence other agents (social learning, influence, and control).

### *15.2.1 Image and Reputation*

The social cognitive model presented here is a dynamic approach that considers reputation as the output of a social process of transmission of information. The input to this process is the evaluation that agents directly form about a given agent during interaction or observation. We will call this evaluation the social *image* of the agent. An agent's *reputation* is argued to be distinct from, although strictly interrelated with, its image. Image consists of a set of evaluative beliefs (Miceli and Castelfranchi 2000) about the characteristics of the target, i.e. it is an assessment of her positive or negative qualities with regard to a norm, a competence, etc. Reputation is both the process and the effect of transmission of a target image. The image relevant for social reputation may concern a subset of the target's characteristics, e.g. its willingness to comply with socially accepted norms and customs. More precisely, we define reputation to consist of three distinct but interrelated objects: (1) a cognitive representation, i.e. a believed evaluation; (2) a population object, i.e. a propagating believed evaluation; (3) an objective emergent property at the agent level, i.e. what the agent is believed to be.

Reputation is a highly dynamic phenomenon in two distinct senses: it is subject to change, especially as an effect of corruption, errors, deception, etc.; and it emerges as an effect of a multi-level bidirectional process (Conte and Paolucci 2002). In particular, it proceeds from the level of individual cognition to the level of social propagation (population level) and from there back to individual cognition. What is even more interesting, once it reaches the population level it gives rise to an additional property at the agent level. From the very moment an agent is targeted by the community, their life will change whether they want or believe it or not. Reputation has become the immaterial, more powerful equivalent of a scarlet letter sewn to one's clothes. It is more powerful because it may not even be perceived by the individual to whom it is attached, and therefore it is not in the individual's power to control and manipulate. Reputation is an objective social property that emerges from a propagating cognitive representation. This lack of an identified source, i.e. its impersonality, is the distinctive feature of reputation, whereas image always requires identifying the individual who made the evaluation.

To formalise these concepts we will begin by defining the building block of image. An agent has made an evaluation when he or she believes that a given entity is good for, or can achieve a given goal. An agent has made a social evaluation when his or her belief concerns another agent as a means for achieving this goal. A given social evaluation includes three sets of agents:

1. A nonempty set $E$ of agents who share the evaluation (evaluators)
2. A nonempty set $T$ of evaluation targets
3. A nonempty set $B$ of beneficiaries, i.e., the agents sharing the goal with regard to which the elements of $T$ are evaluated.

Often, the sets of evaluators and beneficiaries overlap but this is not necessarily the case. A given agent $t$ is a target of a social evaluation when $t$ is believed to be a good/bad means for a given goal of $B$, which may or may not include the evaluator. Evaluations may concern physical, mental, and social properties of targets; agents may evaluate a target with regard to both capacity and willingness to achieve a shared goal. The latter, willingness to achieve a goal or interest, is particular to social evaluations. Formally, $e$ (with $e \in E$) may evaluate $t$ ($t \in T$) with regard to a state of the world that is in $b$'s ($b \in B$) interest, but of which $b$ may not be aware.

The interest/goal with regard to which $t$ is evaluated may be a distributed or collective advantage. It is an advantage for the individual members of $B$, or it may favour a supra-individual entity, that results from the interactions among the members of $B$ (for example, if $B$'s members form a team).

To make this analysis more concrete, we will start with an example drawn from the Sim-Norm model, which will be described in more detail in Section 15.3.1. Let us consider a classical multi-agent situation, a set of agents fighting for access to a scarce resource (food). Assume that a norm of precedence (a proscription against attacking agents who are consuming their "own" resources) is applied to reduce conflicts. The norm is disadvantageous for the norm follower in the short run, but is advantageous for the community, and thus eventually for the individual followers. We will call $N$ the set of norm followers, or normative agents, and $C$ the set of cheaters, or violators of the norm. With regard to social evaluations (image), the targets coincide with the set of all agents; $T = N \cup C$ (all are evaluated). The agents carrying out the evaluation are restricted to the norm followers: $E = N$, because only the normative find purpose in evaluating. Finally, $B = N \cup C$: indeed, if normative agents benefit globally from the presence of the norm, cheaters in this simple setting benefit even more by exploiting the norm; they can attack the weaker while they themselves are safe from attacks by the gullible normative.

It is very easy to find examples where all three sets coincide. General behaviour norms, such as "Do not commit murder" apply to, benefit, and are evaluated by the whole universe of agents. There are situations in which beneficiaries, targets, and evaluators are separate, for example, when norms safeguard the interests of a subset of the population. Consider the quality of TV programs for children broadcast in the afternoon. Here, we can identify three more or less distinct sets. Children are the beneficiaries, while adults entrusted with taking care of children are the evaluators. It could be argued that *B* and *E* still overlap, since *E* may be said to adopt *B*'s interests. The targets of evaluation are the writers of programs and the decision-makers at the broadcast stations. There may be a non-empty intersection between *E* and *T* but no full overlap. In case the target of evaluation is the broadcaster itself, a supra-individual entity, the intersection can be considered to be empty: $\mathrm{E} \cap \mathrm{T} = \emptyset$.

Extending this formalisation to include reputation, we have to differentiate further. To assume that a target *t* is assigned a given reputation implies assuming that *t* is believed to be "good" or "bad," but it does not imply sharing either evaluation. Reputation therefore involves four sets of agents:

1. A nonempty set *E* of agents who share the evaluation;
2. A nonempty set T of evaluation targets;
3. A nonempty set *B* of beneficiaries, i.e. the agents sharing the goal with regard to which the elements of *T* are evaluated;
4. A nonempty set *M* of agents who share the meta-belief that members of *E* share the evaluation; this is the set of all agents aware of the effect of reputation (as stated above, effect is only one component of it; awareness of the process is not implied).

Often, *E* can be taken as a subset of *M*; the evaluators are aware of the effect of evaluation. In most situations, the intersection between the two sets is at least nonempty, but exceptions exist. *M* in substance is the set of reputation transmitters, or third parties. Third parties share a meta-belief about a given target, whether they share the underlying belief or not.

## 15.2.2 Reputational Roles

Agents may play more than one role simultaneously: evaluator, beneficiary, target, and third party. In the following, we will examine the characteristics of the four roles in more detail.

### 15.2.2.1 Evaluator

Any autonomous agent is a potential evaluator Conte et al. 1998. Social agents are likely to form evaluative beliefs about one another (see Castelfranchi 1998) as an effect of interaction and social perception. These may be positive or negative, depending on an agent's experiences. When agents evaluate one another with regard to their individual goals they obtain social evaluations. Image, the result of such social evaluations, serves to identify friends and partners and to avoid enemies. Who should the agent resort to for help? Who should he or she cooperate with? And who should be avoided due to being dangerous or ill willed?

Furthermore, agents may not only evaluate one another with regard to their own goals, but with regard to the goals or interests of a given set of agents (the beneficiaries), to which the evaluators may belong. A negative evaluation may be formed about agents violating others' rights or behaving in an apparently malevolent and hostile manner, whether or not the evaluators consider themselves potential victims of such actions. Information thus obtained may be used to infer that the target could violate other rights in the future, namely, those of the evaluator. In addition, evaluators may be concerned with one another's power to achieve the goals or interests of abstract social entities or institutions, as when we judge others' attitudes towards norms, the church, the government or a political party.

To sum up, agents evaluate one another with regard to their own goals and the goals they adopt from either other individual agents (e.g. their children) or supra-individual agents, such as groups, organisations, or abstract social entities.

#### 15.2.2.2 Beneficiary

A beneficiary is the entity that benefits from the action with regard to which targets are evaluated. Beneficiaries can either be individual agents, groups and organisations or even abstract social entities like social values and institutions. Beneficiaries may be aware of their goals and interests, and of the evaluations, but this is not necessarily the case. In principle, their goals might simply be adopted by the evaluators – as it happens, for example, when members of the majority support norms protecting minorities. Evaluators often are a subset of the beneficiaries.

Beneficiaries may be implicit in the evaluation. This is particularly the case when it refers to a social value (honesty, altruism, etc.); the benefit itself and those who take advantage of it are left implicit, and may coincide with the whole society. The beneficiary of the behaviour under evaluation is also a beneficiary of this evaluation: the more an (accurate) evaluation spreads, the likelier the execution of the positively evaluated behaviour.

#### 15.2.2.3 Target

The target of any social evaluation is the evaluated entity. While targets may even be objects or artefacts to be used by others when the evaluation pertains to image, for reputation mental and moral components are necessarily involved. Holders of reputation (targets) are endowed with the following important characteristics:

- Agency, in particular autonomous agency and sociality (the target is evaluated with regard to a given behaviour)
- Mental states, specifically willingness to perform the respective behaviour
- Decision making or deliberative capacity, i.e. the ability to choose a desirable behaviour from a set of options
- Social responsibility, i.e. the power to prevent social harm and possibly to respond for it, in case any damage occurs.

Intuitively, targets are complementary to beneficiaries, but this is not necessarily the case. Targets may be evaluated for their willingness to exhibit a behaviour from which they are expected to benefit themselves.

Other than beneficiaries, targets are always explicit. They may be individual entities or supra-individual like a group, a collective, an abstract entity, or a social artefact, such as an institution, provided this can be attributed the capacity to make decisions, achieve goals, and perform actions. A further distinguishing characteristic of targets is that they may inherit the reputation of the set they belong to. Offspring may inherit their parents' reputation; employees may suffer from the bad reputation of the firm they work for; members may inherit the reputation of their social groups, whether they have had a chance to decide to join those groups or not. In the latter case, the attitude under evaluation is considered to be a built-in tendency shared by the targets.

Two dynamic processes characterise reputation. The first concerns the propagation of a social cognitive representation from one agent to another. We call this *transmission of reputation* (or gossip). As we have seen so far, it is intrinsic to reputation spreading. The second process concerns the extension of a given agent's reputation to other agents who are related to the former by affiliation, proximity, similarity, etc. We will call it *contagion of reputation* (or inheritance). This dynamic is not intrinsic to reputation, although it may empirically co-occur with reputation spreading. It corresponds to what is usually called prejudice.

The transmission process spreads reputation as a cognitive representation, whereas the contagious process spreads reputation as a property. Both may proceed vertically and horizontally, within social groups and hierarchies and from one agent to his or her neighbours in social space. In this sense, reputation may be perceived as a highly fertile phenomenon, whose propagation is favoured by social membership and proximity.

#### 15.2.2.4 Third Party or Gossiper

An agent is a (potential) third party if she transmits (is in position to transmit) reputation information about a target to another agent or set of agents. Although sharing awareness of a given target reputation, third parties do not necessarily share the corresponding image (social evaluation) of the target. That is, they do not necessarily believe it to be true. Actually, an agent could share a given evaluation with others without being aware of others' existence. This would put him in a poor position to transmit reputation. If there is no belief about *E*, it is very difficult to justify (that is, to assign a goal to) the act of reputation transmission. Instead, the meta-belief about the sharing of the evaluation is a powerful tool and paves the way to complex (and potentially deceiving) new behaviours.

Third parties (if they are also targets) may deserve a negative evaluation; they may actually deceive their fellow third parties, the beneficiaries, the targets, or the society as a whole, by conveying information that they hold to be false. A third party may be bluffing: he or she may pretend to be benevolent with regard to beneficiaries, in order to (1) enjoy the advantages of sharing reputation information, (2) be considered as part of the in-group by other evaluators, and therefore (3) gain a good reputation without sustaining the costs of its acquisition (as would be implied by performing the socially desirable behaviour), and (4) avoid the consequences of a bad reputation. Agents may also spread a false reputation, i.e., pretend that a target has a given reputation when this is not the case. Agents do this in order to achieve the aforementioned benefits without taking responsibility for spreading a given social evaluation.

Agents may have incomplete and inaccurate information – a situation that can result in ill-reputed agents being taken as well-reputed and vice versa. One reason for incomplete or inaccurate information is the lack of personal experience and familiarity with the target. In such a case, an agent might have received information from other third parties or from agents who have had a direct contact with the target.

To demonstrate how the social categories so far identified may help in the analysis of reputation spreading we will return to the example of the quality of TV programs for children. Transmission of evaluations involves two distinct sets of agents, children (beneficiaries) and adults with the children's welfare in mind (evaluators), while the targets – operators and decision-makers at broadcast stations – are a subset of the evaluators. The set of third parties comprises the whole universe of adults, thereby including evaluators and targets but not the beneficiaries. The targets (broadcaster) offer an interesting example of the interplay between institutional and personal roles. While no assumption of benevolence towards *B* can be made on the part of the institution (broadcast stations) itself, whose purpose is profit, things are more intriguing in the case of their employees. These belong to *E* and are thus benevolent with regard to B, but are also members of *T* and thus committed to their employer.

### *15.2.3 Reputation-Based Decisions*

After identifying the different roles agents can take with regard to image, reputation and its transmission let us now examine the relevant decision-making processes. To understand the difference between image and reputation, we will categorise the mental decisions based upon them into three levels:

1. The *epistemic level* is concerned with decisions about accepting or rejecting an evaluative belief.
2. The *pragmatic–strategic level* concerns decisions about interacting with another agent (target).
3. The *memetic level* concerns decisions about transmitting evaluative beliefs about a given target to others

#### 15.2.3.1 Epistemic Level

When an agent decides whether or not to accept a particular belief that forms either a given image or acknowledges a given reputation he makes an epistemic decision. This involves direct evaluation of the belief in question. To acknowledge a given reputation implies two more specific beliefs: (a) that a nonempty set *E* of agents within the same population shares the given image about the target (reputation effect), and/or (b) that information about the target's image has been circulated (reputation process) among those agents.

Image and reputation about one and the same agent do not necessarily overlap. From an agent *X*'s meta-belief that agent *Y* is said to be e.g. a womaniser we are not entitled to derive *X*'s belief that *Y* is in fact a womaniser (although *X* may draw such a more or less arbitrary conclusion). Image and reputation about the same *Y* may be consistent (for example, *X* believes that *Y* is and is believed to be a womaniser) or inconsistent (*X* believes that *Y* either suffers from an undeserved bad reputation or enjoys an unworthy good one).

To accept one does not imply acceptance of the other. The two processes of acceptance are different not only in their respective outputs (evaluations and meta-beliefs), but also in the operations involved. To accept a given image implies coming to share it. The acceptance may be based, for example, upon supporting evidence and first-hand experience with the image target, consistent pre-existing evaluations (concerning, for example, the class of objects to which the target belongs), or trust in the source of the given evaluative belief.

Acknowledging a given reputation, on the other hand, does not necessarily lead to sharing others' evaluations but rather to believing that these evaluations are held or at least circulated by others. To assess the value of such a meta-belief is a rather straightforward operation. For the recipient to be relatively confident about this meta-belief, it is probably sufficient for him or her to hear some rumours.

#### 15.2.3.2 Pragmatic-Strategic Level

Agents resort to their evaluative beliefs in order to achieve their goals (Miceli and Castelfranchi 2000). In general, evaluations are guidelines for planning; therefore social evaluations (evaluations about other agents) are guidelines for social action and social planning. The image an agent *X* has about a target agent *Y* will guide his or her action with regard to *Y*, will suggest whether it is convenient to interact with *Y* or not, and also will suggest what type of interaction to establish with *Y*. Image may also be conveyed to others in order to guide their actions towards the target in a positive or negative sense. To do so, the agent must (pretend to) be committed to their evaluation and take responsibility for its truth value.

While image dominates direct pragmatic-strategic decisions reputation may be used when it is consistent with image or when no image of the target has been formed. To influence others' social decisions, however, agents tend to transmit information about the target's reputation rather than their image of the target. Two main reasons explain this inverse pattern:

Agents expect that a general opinion is more credible and acceptable than an individual one.

Agents reporting on reputation do not need to commit to its truth value, and do not have to take responsibility for it; consequently, they may influence others at a lower personal cost.

#### 15.2.3.3 Memetic Level

A memetic decision can be roughly described as the decision to spread reputation. Consequently, communication about reputation is communication about a meta-belief, i.e. about others' mental attitudes. To spread news about someone's reputation does not bind the speaker to commit himself to the truth value of the evaluation conveyed but only to the existence of rumours about it. In other words, communication about reputation does neither imply any personal commitment of the speaker with regard to the content of the information delivered – if an agent *X* reports on *Y*'s bad reputation he is by no means stating that *Y* deserved it – nor any responsibility with regard to the credibility of (the source of) information ("I was told that *Y* is a bad guy"). This is due to the source of the meta-belief being implicit ("I was told...") and the set of agents to whom the belief is attributed being undefined ("*Y* is ill/well reputed").

Communication about reputation is not always sincere. On the contrary, the lack of commitment and responsibility can (and often does) lead to deceptive reputation transmission. To deceive other agents about *Y*'s reputation agent *X* only needs to report it as a rumour independent of or even opposite to his own beliefs.

Several factors may affect an agent's decision to transmit reputation information (see Conte and Paolucci 2002):

- Certainty and acceptance of the evaluation,
- Reputation of the source from which information was received,
- Responsibility and accountability for the effects of distributing this evaluation to others,
- Benevolence toward the beneficiary as opposed to benevolence toward the target, or no benevolence at all.

Social responsibility can be defined as the power attributed to intelligent autonomous social systems to predict and prevent harm to others and/or themselves. A memetic agent may be said to have harmed another agent by transmitting bad reputation information about her. Responsibility comes from accountability, i.e. the power and obligation to respond and repair harm that one has been found accountable for. A memetic agent may be asked to take responsibility and possibly repair the harm. However, responsibility is much less crucial in a memetic than in a pragmatic decision concerning reputation. The decision to transmit reputation will thus depend on the extent to which the memetic agent is aware of potentially serious effects of reputation transmission on the target and his perceived contribution to these effects. The latter is influenced by the perceived “distance” from the target. It is easier to convey potentially dangerous evaluations when the target is far away, absent, unknown, etc., not only because the memetic agent is less likely to perceive the effective occurrence of harm but also because he or she will not be a direct cause of it. Harm will actually depend on a number of intermediate events, others’ decisions, etc. The greater the distance and the smaller the memetic agent’s (perceived) responsibility, the less cautious the memetic agent is likely to be.

Benevolence means whether and to what extent a candidate third party adopts the goals or interests of either the target or the beneficiary of reputation transmission. When circulating reputation, memetic agents may follow different strategies, according to the direction of their benevolence.

In case of benevolence towards the set of beneficiaries $B$, gossiping may follow some prudence rule like “pass on negative evaluation even if uncertain, pass on positive evaluation only if certain”. This may give way to circulation of a reputation which is overly critical and worse than the real characteristics of the target.

When the benevolence is target-oriented (set $T$), it is possible to expect the application of some courtesy rule like “pass on positive evaluation even if uncertain, negative evaluation only if certain”. This may lead to a courtesy equilibrium where no-one expresses critique anymore, especially fearing retaliation.

If memetic agents are not benevolent towards any of the two sets $B$ and $T$, Conte and Paolucci (2002) predicts scarcity of reputation transmission. In this case production of information may be induced by institutional reinforcement of it, e.g. a university requesting students to academically evaluate their lecturers.

Systematic application of a courtesy or prudence rule in reputation spreading may result in selective transmission of either positive or negative evaluations. We expect general adoption of such rules as a consequence of self-interested decisions of single memetic agents. We may reasonably suppose that a lower responsibility of memetic agents increments the quantity of circulating information, while benevolence direction is a consequence – all other factors being equal – of the overlapping of reputational roles of agents. Role overlapping means that the same people (or people having the same norm-related goals) are involved in more than one reputational role. Such agents then may consider it useful to be prudent or generous in their information spreading, or they may have not enough motivation to circulate information of any sort.

As we have seen in this and the previous section, operationalization of cognitive properties of reputation and dynamics of reputational groups allows us to express testable hypotheses, which can be investigated both by experimenting with human subjects and by designing software reputation agents.

## 15.3 Simulating Reputation

Social simulation models have been successfully employed to investigate the effects of reputation in different contexts. In this section we will present an overview of three different ways of designing agent-based systems based on the cognitive theory of reputation presented previously. These models can be distinguished according to:

- Different levels of cognitive complexity of the agents,
- The kind of setting (purely cooperative, competitive or both)
- The scenario.

The first, Sim-Norm, based on a very simple concept of reputation, has been applied to observe the impact of reputation in social control. The second, REPAGE, based on a more complex agent architecture and on the concepts of reputation and image introduced in the previous section, has been applied to explore the impact of social evaluation on the market. Finally, in the third model, called SOCRATE, relatively simple agents interact in a complex market in which they exchange both goods and information.

### *15.3.1 Sim-Norm*

This model was developed to examine the effect of reputation on the efficiency of a norm of precedence (Castelfranchi, Conte, Paolucci 1998; Conte and Paolucci 1999; Paolucci 2000) in reducing aggression, measured both at the global (i.e. societal) and local (i.e. individual) level. In particular, Sim-Norm was designed to explore why self-interested agents exercise social control. Albeit far from reaching a final conclusion on this issue, the studies based on Sim-Norm confirmed a positive impact of reputation on social control.

More precisely, while individually acquired evaluation of other agents gave norm executors no significant advantage, the transmission of these evaluations among norm executors proved decisive in levelling the outcomes of norm-abiders and cheaters (if numerically balanced).

#### 15.3.1.1 Hypotheses

Sim-Norm revolved around the question of which ingredients are necessary for social order to be established in a society of agents. The role of norms as aggression controllers in artificial populations living under conditions of resource scarcity was addressed. We set out to explore two hypotheses:

Norm-based social order can be maintained and its costs reduced via distributed social control.

Social cognitive mechanisms are needed to account for distributed social control. In particular, the propagation of social beliefs plays a decisive role in distributing social control at low or zero individual costs and high global benefit.

#### 15.3.1.2 Model Description and Experimental Conditions

The model defines agents as objects moving in a two-dimensional environment (a 10 $\times$ 10 grid) with randomly scattered food. At the beginning of a run, agents and food items are assigned locations at random. A location is a cell in the grid. The same cell cannot contain more than one object at a time (except when an agent is eating). The agents move through the grid in search of food, stopping to eat to build up their strength when they find it. The agents can be attacked only when eating; no other type of aggression is allowed.

At the beginning of each step, every agent selects an action from the six available routines: *eat*, *move-to-food-seen*, *move-to-food-smelled*, *attack*, *move-random*, and *pause*. Actions are supposed to be simultaneous and time consuming.

To investigate the role of norms in the control of aggression, we compared scenarios in which agents follow a norm – implemented as a restriction on attacks – with identical scenarios, in which they follow utilitarian rules. In all scenarios, each agent can perform only one of three strategies:

Blind aggression, or control condition, in which aggression is not constrained. If the agent can perform no better move (eating, moving to food seen or smelled), then it will attack without further considerations. Blind agents have access to neither their own strength nor the eater's strength; these parameters never enter their decision-making process.

Utilitarian, in which aggression is constrained by strategic reasoning. Agents will only attack those eaters whose strength is lower than their own. An eater's strength is "visible" one step away from the agent's current location. While blind agents observe no rule at all, utilitarian agents observe a rule of personal utility, which does not qualify as a norm.

Normative (N), in which aggression is constrained by a norm. We introduced a finder-keeper precept, assigning a "moral right" to food items to finders, who become possessors of the food. Possession of food is ascribed to an agent on the grounds of spatial vicinity; food owned is flagged, and every player knows to whom it belongs. Each food unit may have up to five owners, decided on the basis of proximity at the time of creation. The norm then prescribes that agents cannot attack other agents who are eating their own food.

The strategies can also be characterised by the kind of agents they allow to attack: while blind agents attack anybody, the utilitarian agents attack only the weaker, and the normative agents, respecting a norm of private property, will not attack agents who are eating their own food.

These strategies were compared (Castelfranchi et al. 1998) using an efficiency measure – the average strength of the population after *n* periods of simulation – and a fairness measure, the individual deviation from the average strength.

#### 15.3.1.3 Findings

The first two series of experiments showed that normative agents perform less well than non-normative agents in mixed populations, as they alone bear the costs of social control and are exploited by utilitarian agents. In a following series of experiments, "image" was added to the preceding experimental picture. Now, each normative agent collects information about the behaviour of other agents in an image vector. This information is binary and discriminates between the "respectful," who will abide with the norm and "cheaters," who will not respect the principle of finders–keepers. The vector is initialised to "all respectful" (presumption of innocence), but every time a normative agent is attacked while eating its own food, the attacker is recorded as a cheater. Moreover, the normative algorithm is modified so that the agents respect the norm only when facing agents known as respectful, while they behave with known cheaters according to one of the retaliation strategies listed above.

The results from another set of experimental runs on a mixed population equally composed of normative and utilitarian agents show that in this model useful knowledge can be drawn from personal experience, but therefore still at one's own cost. To reduce cost differences among subpopulations, image is insufficient.

Henceforth, we provided the respectful agents with the capacity to exchange with their (believed-to-be) respectful neighbours at distance one images of other agents. With the implementation of a mechanism of transmission of information, we can speak of a reputation system. We ran the experiments again with normative agents exchanging information about cheaters. The results suggest that circulating knowledge about others' behaviours significantly improves normative agents' outcomes in a mixed population.

The spreading of reputation can then be interpreted as a mechanism of cost redistribution for the normative population. Communication allows compliant agents to easily acquire preventive information, sparing them the costs of direct confrontations with cheaters. By spreading the news that some guys cheat, the good guys (1) protect themselves, (2) at the same time punish the cheaters, and possibly (3) exercise an indirect influence on the bad guys to obey the norm. Social control is therefore explained as an indirect effect of a "reciprocal altruism" of knowledge.

#### 15.3.1.4 Concluding Remarks

The Sim-Norm model presented in this section was studied with the purpose of clarifying the role of norms controlling aggression in simple multi-agent systems. The model shows that a simple norm of precedence is not as efficient as a utilitarian rule when utilitarian agents are present; the normative agents must resort to retaliation against cheaters. The addition of image is not enough to defend the norm, but image coupled with a mechanism of information transmission is. The necessity of information transmission points out the relevance of our distinction between image and reputation.

The model inspired further research in the social simulation community: Saam and Harrer (1999) used the same model to explore the interaction between normative control and power, whereas Hales (2002) applied an extended version of SimNorm to investigate the effects of group reputation. In his model agents are given the cognitive capacity to categorise other agents as members of a group and project reputation onto whole groups instead of individual agents (a form of stereotyping).

### 15.3.2 The REPAGE Cognitive Architecture

REPAGE is a computational system for partner selection in a competitive setting (marketplace), although in principle it can also be used in cooperative contexts (organizations). Based on a model of REPutation, imAGE and their interplay, REPAGE provides evaluations of potential partners and is fed with information from others plus outcomes from direct experience. This is fundamental to account for (and to design) limited autonomous agents as exchange partners. To select good partners, agents need to form and update own social evaluations; hence, they must exchange evaluations with one another.

But in order to preserve their autonomy, agents need to *decide* whether or not to share others' evaluations of a given target. If agents would automatically accept reported evaluations and transmit them as their own, they would not be autonomous anymore. In addition, in order to exchange information about reputation, agents need to participate in circulating it, whether they believe it or not (gossip); but again to preserve their autonomy, they must *decide* how, when and about whom to gossip.

In sum, the distinction between image and reputation suggests a way out from the paradox of sociality, i.e. the trade-off between agents' autonomy and their need to adapt to social environment. On one hand, agents are autonomous if they select partners based on their social evaluations (images). On the other, they need to update evaluations by taking into account others' evaluations. Hence, social evaluations must circulate and be represented as "reported evaluations" (reputation), before and in order for agents to decide whether to accept them or not. To represent this level of cognitive detail in artificial agents' design, there is a need for a specialised subsystem. This is what REPAGE provides.

In the following we briefly describe the architecture of the REPAGE model and how it is integrated with the other elements that compose a typical deliberative agent (cf. Sabater et al. 2006 for a more detailed description). REPAGE is composed of three main elements: a memory, a set of components called detectors and the analyzer. An implementation of REPAGE in Java has been published as a Sourceforge project.

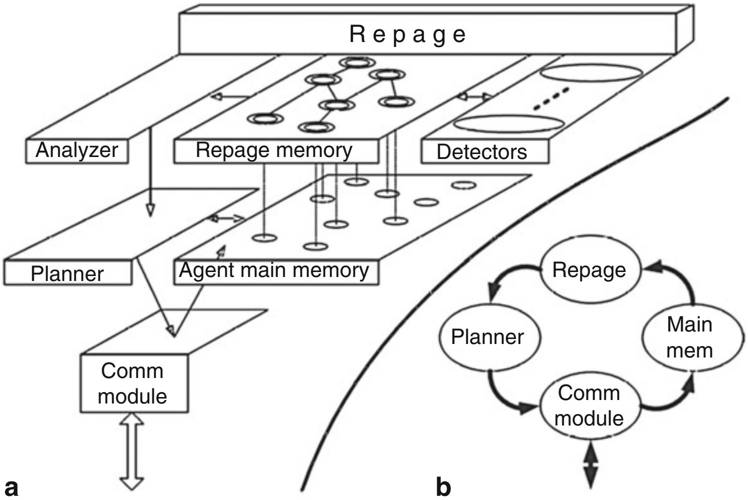



**Fig. 15.1** REPAGE and its environment. For the sake of clarity only agent components that interact with REPAGE are depicted

#### 15.3.2.1 Memory

In the implementation, to support the specialized nature of REPAGE, memory is actually composed by a set of references to the predicates in the agent general-purpose memory (see section A of Fig. 15.1). Only those predicates that are relevant for dealing with image and reputation are considered. Therefore, of all the predicates in the main memory, only a subset is also part of the REPAGE memory. A change in a predicate is immediately visible in both memories.

To mirror their dependence connections, in the REPAGE memory predicates are conceptually organized in different levels and inter-connected. Each predicate reference is wrapped by a component that adds connection capabilities to it.

Predicates contain a fuzzy evaluation, consisting of three aspects: the type of the evaluation (e.g. personal experience, image, third party image), the role of the target (e.g. informant or seller) and the actual content. To store the content, a simple number is used in e-Bay and in most reputation systems. This sharp representation, however, is quite implausible in inter-agent communication, which is one of the central aspects of REPAGE; in real life no-one is told that people are saying Jane is 0.234 good. To capture the lack of precision coming (a) from vague utterances, e.g. "I believe that agent *X* is good, I mean, very good – good, that is", and (b) from noise in the communication or in the recollection from memory, we decided to model the actual value of an evaluation with a tuple of positive real values that sum to one. Aggregation of these evaluations was detailed in Sabater and Paolucci (2007).

Finally, each predicate has a strength value associated to it. This value is a function of (1) the strength of its antecedents and of (2) some special characteristics intrinsic to that type of predicate. The network of dependencies specifies which predicates contribute to the values of other predicates. Each predicate in the REPAGE memory has a set of antecedents and a set of consequents. If an antecedent changes its value or is removed, the predicate is notified. Then the predicate recalculates its value and notifies the change to its consequents.

#### 15.3.2.2 Detectors

The detectors are inference units specialized in certain predicates. They populate the REPAGE memory (and consequently the main memory of the agent) with new predicates inferred from those already in the memory. They are also responsible for removing predicates that are no longer useful and, more importantly, for creating the network of dependencies among the predicates.

Each time a new predicate is added to or removed from the main memory (either by the action of another agent module – planner, communication module, etc. – or by the action of a detector) the REPAGE memory notifies all detectors 'interested' in that type of predicate. This starts a cascading process where several detectors are activated one after the other. At the same time, the dependency network ensures the correct update of the predicate values according to the new additions and subtractions.

#### 15.3.2.3 Analyzer

The main task of the analyzer is to propose actions that (1) can improve the accuracy of the predicates in the REPAGE memory and (2) can solve cognitive dissonances trying to produce a situation of certainty. The analyzer can propose one or more suggestions to the planner, which then decides whether to execute them or not.

#### 15.3.2.4 Integration with Deliberative Agent Architecture

One of the key points of the REPAGE design is its easy integration with the other elements that compose a deliberative agent. REPAGE is not only a passive module the agent can query to obtain information about image and reputation of another agent. The aim of REPAGE is also to provide the agent (or more specifically, the planner module of the agent) with a set of possibilities that can be followed to improve the reliability of the provided information.

The communication module connects the agent with the rest of the world. After a possible process of filtering and/or transformation of the received information, new predicates are added to the main memory.

The REPAGE memory contains references to those predicates in the main memory of the agent that are relevant to deal with image and reputation. The actions of the detectors over the REPAGE memory result in addition/removal of predicates as well as the creation of the dependence network. While the addition or removal of predicates has again an immediate effect on the main memory, the dependence network is present only in the REPAGE memory.

The planner uses the information in the main memory to produce plans. This information includes the information generated by REPAGE. By means of the analyzer, REPAGE always suggests new actions to the planner in order to improve the accuracy of existing images and reputations. It is a task of the planner to decide which actions are worth being performed. These actions (usually asking informers or interacting with other agents) will hopefully provide new information that will feed REPAGE and improve its accuracy. This cycle is illustrated in Fig. 15.1(B).

#### 15.3.2.5 Demonstration

To illustrate some of the main points in the behaviour of REPAGE let us consider two particular situations that are quite common in the application area of markets. The general scenario is the following: Agent *X* is a buyer who knows that agent *Y* sells what he needs but knows nothing about the quality of agent *Y* (the target of the evaluations) as a seller. Therefore, he turns to other agents in search for information – the kind of behaviour that can be found, for example, in Internet fora, auctions, and in most agent systems.

In the first situation, agent *X* receives a communication from agent *Z* saying that his image of agent *Y* as a seller is very good. Since agent *X* does not yet have an image about agent *Z* as informer he resorts to a default image that is usually quite low. The uncertain image as an informer adds uncertainty to the value of the communication.

Later on, agent *X* has received six communications from different agents containing their image of agent *Z* as an informer. Three of them give a good report and three a bad one. This information is enough for agent *X* now to build an image about agent *Z* as an informer so this new image substitutes the default candidate image that was used so far. However, the newly formed image is insufficient to take any strategic decision – the target seems to show an irregular behaviour.

At this point, agent *X* decides to try a direct interaction with agent *Y*. Because he is not sure about agent *Y* he resorts to a low risk interaction. The result of this interaction is completely satisfactory and has important effects in the REPAGE memory. The candidate image about agent *Y* as a seller becomes a full image, in this case a positive one.

Moreover, this positive image is compared (via a fuzzy metric) with the information provided by agent *Z* (which was a positive evaluation of agent *Y* as a seller); since the comparison shows that the evaluations are similar, a positive confirmation of the image of agent *Z* as an informer is generated. This reinforcement of the image of agent *Z* as a good informer at the same time reinforces the image of agent *Y* as a good seller. As a consequence, there is a positive feedback between the image of agent *Y* as a good seller and the image of agent *Z* as a good informer. This feedback is a necessary and relevant part of the REPAGE model.

The purpose of the second situation is to show how REPAGE differentiates between image and reputation. In this case agent *X*, after a couple of successful interactions with agent *Y*, receives four communications from different informants. Each informant communicates the reputation of agent *Y* as a seller, which happens to be negative. This contradicts agent *X*'s own positive evaluations of Y stemming from their direct interactions. However, it is not a problem in REPAGE because there is a clear distinction between image and reputation. In addition, unlike communicated images (see first situation) communications about reputation do not generate confirmations that reinforce or weaken the image of the informant.

#### 15.3.2.6 Concluding Remarks

REPAGE is a cognitive architecture to be used by artificial autonomous agents in different scenarios, be they MAS applications, artificial markets or teamwork. Its main objective is to give agents access to a fundamental module for social reasoning and decision-making on the first two levels described in Sect. 15.2.3.1 and 15.2.3.2, epistemic and strategic. Thus, the main outputs of REPAGE consist of "advice" about what to think of a given target and how to interact with it.

However, there are still some aspects of reputation-based decisions, which need further theoretical elaboration and inclusion into REPAGE. First, the interplay between image and reputation in the epistemic decision: which received evaluations will be transformed into own evaluations of a given target and why? Secondly, the grounds on which memetic decisions are made: what is said to whom, and moreover, why? Both these issues will prove particularly relevant in testing this architecture by means of simulation.

### *15.3.3 SOCRATE*

SOCRATE is an attempt to test the cognitive theory of reputation in a ideal-typical economic setting, modeled after an industrial district in which firms exchange goods and information (Giardini et al. 2008; Di Tosto et al. 2010). In this model, the focus is on social relationships among agents and their impact on the focusing on social links and on the resulting social structure, usually informal, which is a defining feature of industrial clusters (Porter 1998; Fioretti 2005; Squazzoni and Boero 2002). Social evaluations are the building blocks of social and economic relationships inside the cluster; they are used to select trustworthy partners, to create and enlarge the social network (Giardini and Cecconi 2010), and to exert social control on cheaters.

#### 15.3.3.1 Agents and Environment

We implemented an artificial environment in which agents can choose among several potential suppliers by relying either on their own evaluations, or on other agents' evaluations. In the latter case, the availability of truthful information could help agents to find reliable partners without bearing the costs of potentially negative, i.e. harmful, interactions with bad suppliers. Moreover, evaluations can be transmitted either as image (with an explicit source and the consequent risk of retaliation) or as reputation.

We tried to answer the following questions:

- How relevant is image when firms need to select suppliers, service providers and so on?
- Does transmission of image promote the improvement of quality in a cluster?
- How does false information affect the quality of the cluster?
- What are the effects of image and reputation, respectively, on the economic performance of firms?

Our model is characterized by the existence of two different kinds of interactions among agents: material exchange and evaluation exchange. The former refers to the exchange of products between leader firms and their suppliers, and it leads to the creation of a supply chain network. On the other hand, the flows of social evaluations among the firms create a social network. In this setting, agents can transmit true or false evaluations in order to either help or hamper their fellows searching for a good partner.

Agents are firms organized into different layers, in line with their role in the production cycle. The number of layers can vary according to the characteristics of the cluster, but a minimum of two layers is required. Here, we have three layers, but n possible layers can be added, in order to develop a more complex production process:

Layer 0 (L0) is represented by leader firms that supply the final product
Layer 1 (L1) is represented by suppliers of L0
Layer 2 (L2) are firms providing raw material to firms in L1.

When image transmission is allowed, both leader firms and suppliers exchange information with their fellows, thus creating and taking part in a social network. This process works only horizontally: L0 and L1 are not allowed to talk each other. Agents in both layers can play two possible roles:

*Questioner* – asks an Informer, i.e. another firm of the same layer, to suggest a good supplier;

*Informer* – provides her own image of a good supplier. Honest informers suggest their best rated supplier, whereas cheaters transmit the image of their worse supplier (as if it was a good one).

Agents in L0 have to select suppliers that produce with a quality above the average among all L1 agents. Suppliers can be directly tested or they can be chosen thanks to the information received by other L0 firms acting as Informers. Buying products from L1 and asking for information to L0 fellows are competing activities that can not be performed contemporaneously. In turn, once received an order for a product, L1 firms should select a good supplier (above the average quality) among those in L2. After each interaction with a supplier, both L0 and L1 agents create an evaluation, i.e. an image, of it, comparing the quality of the product they bought with the quality threshold value. Agents are endowed with an "Image Table" in which all the values of the tested partners are recorded and stored for future selections. In the Reputation condition, evaluations are exchanged without revealing their source, thus injecting the cluster with untested information. In this condition, retaliation against untrustful informers is unattainable.

At each simulation cycle, firms attempt to interact with the best known suppliers. Every time the best known supplier is unavailable they query their fellows about other high quality suppliers, which will be tested and integrated into the Image-Table.

#### 15.3.3.2 Results

The exchange of true and valuable information led to a significant increase in quality of exchanged goods, making the cluster exploration faster and permitting firms to obtain higher profits. However, in the Image condition average quality was negatively affected by the presence of cheaters, because false information triggered a mechanism of reciprocal retaliation with detrimental effects on the cluster as a whole. In the Reputation Condition, the cluster could absorb relatively high percentages of cheaters, without compromising its economic performance.

SOCRATE results provided further support to the hypotheses about the importance of reputation for social control, showing again that social evaluations and their features have consequences also in economic terms.

#### 15.3.3.3 Concluding Remarks

Given the assumption that in this "small-world", as in the real world, evaluations are crucial to selecting trustworthy partners and to isolating cheaters, we tried to demonstrate how useful this exchange is, especially in terms of global cluster quality and profits. Firms receiving reliable information about potential partners found good suppliers in a faster and more efficient way, compared to firms that were systematically cheated by their fellows. More interesting results are expected after we include an enriched economic structure and the implementation of reputation.

## 15.4 Conclusion and Future Work

In the last decade there has been a significant increase in research on reputation and gossip. There is growing evidence on the fact that the presence of reputation strongly promotes cooperation and represents an effective way to maintain social control. Exercising social control roughly means to isolate and punish the cheaters. However, punishment is costly and it inevitably implies the problem of second-order cooperation.

In this chapter, we discussed current studies of reputation as a distributed instrument for social order. After a critical review of current technologies of reputation in electronic institutions and agentized environments, a theory of reputation as a social cognitive artefact was presented. In this view, reputation allows agents to cooperate at a social meta-level, exchanging information (a) for partner selection in competitive settings like markets and (b) for cheater isolation and punishment in cooperative settings like teamwork and grouping.

To exemplify both functionalities, we introduced three simulation models of reputation in artificial societies developed within our research group during the last decade. Both have been used mainly as a theory-building tool.

The first, Sim-Norm, is a reputation-based model for norm compliance. The main findings from simulations show that, if circulated among norm-abiders only, reputation allows for the costs of compliance to be redistributed between two balanced subpopulations of norm-abiders and cheaters. In such a way, it contributes to the fitness of the former, neutralising the advantage of cheaters. However, results also show that as soon as the latter start to bluff and optimistic errors begin to spread in the population, things worsen for norm-abiders, to the point that the advantage produced by reputation is nullified.

REPAGE, a much more complex computational model than SimNorm, was developed to test the impact of image, reputation and their interaction on the market. Based on our social cognitive theory, it allows the distinction between image and reputation to be made, and the trade-off between agents' autonomy and their liability to social influence to be coped with. REPAGE allows the circulation of reputation whether or not third parties accept it as true.

Finally, SOCRATE is an attempt to combine fairly complex agents (endowed with a memory and able to manage different kinds of evaluations) with a market in which agents must protect themselves from both informational and material cheating. In this context, reputation has been proven to be useful to punish cheaters but it also prevented the social network from collapse.

These results clearly show that differentiating image from reputation provides a means for coping with informational cheating and that further work is needed to achieve a better understanding of this complex phenomenon. The long term results of these studies are expected to (a) answer the question on how to cope with informational cheating (by testing the above hypothesis), (b) provide guidelines about how to realize technologies of reputation that achieve specified objectives (e.g. promoting respect of contracts vs. increasing volume of transactions), and finally (c) show the impact of reputation on the competitiveness of firms within and between districts.

**Acknowledgements** The authors would like to thank Jordi Sabater and Samuele Marmo for their helpful collaboration. This work was partially supported by the Italian Ministry of University and Scientific Research under the Firb programme (SOCRATE project, contract number RBNE03Y338), by the European Community under the FP6 programme (eRep project, contract number CIT5-028575; EMIL project, contract number IST-FP6-33841).

## Further Reading

For a more in-depth treatment of the contents of this chapter we refer the reader to the monograph *Reputation in Artificial Societies* (Conte and Paolucci 2002). A review of this book was published in JASSS (Squazzoni 2004). For more on the same line of research, with an easier presentation aimed to dissemination, we suggest the booklet published as the result of the eRep project (Paolucci et al. 2009).

A conference aiming to propose a scientific approach to Reputation has been organized in 2009: the first International Conference on Reputation, ICORE 2009. Its proceedings (Paolucci 2009), available online, contain a collection of papers that give an idea of the range of approaches and ideas on Reputation from several academic disciplines.

Due to the focus on the theoretical background of reputation only a narrow selection of simulation models of reputation could be discussed in this chapter. Sabater and Sierra (2005) give a detailed and well-informed overview of current models of trust and reputation using a variety of mechanisms. Another good starting point for the reader interested in different models and mechanisms is the review by Ramchurn and colleagues (Ramchurn et al. 2004a).

Further advanced issues for specialised reputation subfields can be found in (Jøsang et al. 2007), a review of online trust and reputation systems, and in (Koenig et al. 2008), regarding the Internet of Services approach to Grid Computing.

## References

Abdul-Rahman A, Hailes S (1997a) A distributed trust model. In: Proceedings of NSPW97, new security paradigms workshop VI, Langdale, Cumbria, 23–26 Sept 1997. ACM Press, New York, pp 48–60

Akerlof G (1970) The market for lemons: quality uncertainty and the market mechanisms. Q J Econ 84:488–500

Alexander R (1987) The biology of moral systems. De Gruyter, New York

Axelrod R (1984) The evolution of cooperation. Basic Books, New York

Bachmann R (1998) Kooperation, vertrauen und macht in systemen verteilter künstlicher intelligenz: eine vorstudie zum verhältnis von soziologischer theorie und technischer modellierung. In: Malsch T (ed) Sozionik. Edition Sigma, Berlin, pp 197–234

Barkow JH (1996) Beneath new culture is old psychology: gossip and social stratification. In: Barkow JH, Cosmides L, Tooby J (eds) The adapted mind: evolutionary psychology and the generation of culture. Oxford University Press, New York, pp 627–638
Bohem C (1999) Hierarchy in the forest: the evolution of egalitarian behavior. Harvard University Press, Cambridge, MA
Bolton G, Katok E, Ockenfels A (2002) How effective are online reputation mechanisms? An experimental investigation (Working paper). Max Planck Institute of Economics, Jena
Cabral L, Hortaçsu A (2006) The dynamics of seller reputation: theory and evidence from eBay (Discussion papers, 4345). Centre for Economic Policy Research, London
Carbo J, Molina JM, Davila J (2002) Comparing predictions of SPORAS vs. a fuzzy reputation agent system. In: Grmela A, Mastorakis N (eds) Proceedings of the 3rd WSEAS international conference on Fuzzy Sets and Fuzzy Systems (FSFS '02), Interlaken, 11–15 Feb 2002. WSEAS, pp 147–153
Castelfranchi C (1998) Modelling social action for AI agents. Artif Intell 103(1–2):157–182
Castelfranchi C, Conte R, Paolucci M (1998) Normative reputation and the costs of compliance. J Artif Soc Soc Simul 1(3). http://jasss.soc.surrey.ac.uk/1/3/3.html
Conte R (1999) Social intelligence among autonomous agents. Comput Math Organ Theory 5: 202–228
Conte R, Castelfranchi C (1995) Cognitive and social action. UCL Press, London
Conte R, Paolucci M (1999) Reproduction of normative agents: a simulation study. Adapt Behav 7 (3/4):301–322
Conte R, Paolucci M (2002) Reputation in artificial societies: social beliefs for social order. Kluwer, Dordrecht
Conte R, Paolucci M (2003) Social cognitive factors of unfair ratings in reputation reporting systems. In: Liu J, Liu C, Klush M, Zhong N, Cercone N (eds) Proceedings of the IEEE/WIC international conference on web intelligence – WI 2003. IEEE Computer Press, Halifax, pp 316–322
Conte R, Castelfranchi C, Dignum F (1998) Autonomous norm-acceptance. In: Müller JP, Singh MP, Rao AS (eds) ATAL '98: proceedings of the 5th international workshop on intelligent agents V, agent theories, architectures, and languages. Springer, London, pp 99–112
Cravens K, Goad Oliver E, Ramamoorti S (2003) The reputation index: measuring and managing corporate reputation. Eur Manage J 21(2):201–212
Dellarocas CN (2003) The digitalization of word-of-mouth: promise and challenges of online feedback mechanisms (MIT Sloan working paper, 4296–03). MIT Sloan School of Management, Cambridge, MA
Di Tosto G, Giardini F, Conte R (2010) Reputation and economic performance in industrial districts: modelling social complexity through multi-agent systems. In: Takadama K, Cioffi-Revilla C, Deffuant G (eds) Simulating interacting agents and social phenomena: the second world congress, vol 7, ABSS. Springer, Tokyo, pp 165–176
Dunbar R (1996) Grooming, gossip, and the evolution of language. Faber and Faber, London
Fehr E, Gachter S (2000) Fairness and retaliation: the economics of reciprocity. J Econ Perspect 14:159–181
Fioretti G (2005) Agent based models of industrial clusters and districts (Working paper, series urban/regional, 0504009). EconWPA http://ideas.repec.org/p/wpa/wuwpur/0504009.html
Frith CD, Frith U (2006) How we predict what other people are going to do. Brain Res 1079 (1):36–46
Giardini F, Cecconi F (2010) Social evaluations, innovation and networks. In: Ahrweiler P (ed) Innovation in complex social systems. Routledge, London, pp 277–289
Giardini F, Di Tosto G, Conte R (2008) A model for simulating reputation dynamics in industrial districts. Simul Model Pract Theory 16:231–241
Gintis H, Smith EA, Bowles S (2001) Costly signaling and cooperation. J Theor Biol 213:103–119
Gluckman M (1963) Gossip and scandal. Curr Anthropol 4:307–316

Gray ER, Balmer JMT (1998) Managing corporate image and corporate reputation. Long Rage Plann 31(5):695–702

Greif A (1993) Contract enforceability and economic institutions in early trade: the maghribi traders' coalition. Am Econ Rev 83(3):525–548

Hales D (2002) Group reputation supports beneficent norms. J Artif Soc Soc Simul 5(4). http://jasss.soc.surrey.ac.uk/5/4/4.html

Huynh D, Jennings NR, Shadbolt NR (2004) Developing an integrated trust and reputation model for open multi-agent systems. In: Falcone R, Barber S, Sabater-Mir J, Singh M (eds) Proceedings of the 7th international workshop on trust in agent societies, New York, 19 July 2004, pp 65–74

IC3 (2005): IC3 2005 Internet crime report http://www.ic3.gov/media/annualreport/2005_IC3Report.pdf

Jøsang A, Ismail R, Boyd C (2007) A survey of trust and reputation systems for online service provision. Decis Support Syst 43(2):618–644

Koenig S, Hudert S, Eymann T, Paolucci M (2008) Towards reputation enhanced electronic negotiations for service oriented computing. In: Falcone R, Barber S, Sabater-Mir J, Singh M (eds) Trust in agent societies: 11th international workshop, TRUST 2008, Estoril, 12–13 May 2008, revised selected and invited papers. Springer, Berlin, pp 273–291

Kreps D, Wilson R (1982) Reputation and imperfect information. J Econ Theory 27:253–279

Leslie AM (1992) Autism and the 'theory of mind' module. Curr Dir Psychol Sci 1:18–21

Marsh S (1992) Trust in distributed artificial intelligence. In: Castelfranchi C, Werner E (eds) Artificial social systems, 4th European workshop on modelling autonomous agents in a multi-agent world, MAAMAW '92, S. Martino al Cimino, 29–31 July 1992, selected papers. Springer, Berlin, pp 94–112

Marsh S (1994a) Formalising trust as a computational concept. PhD Thesis, Department of Computing Science and Mathematics, University of Stirling, Stirling. http://www.cs.stir.ac.uk/research/publications/techreps/pdf/TR133.pdf

Marsh S (1994b) Optimism and pessimism in trust. In: Proceedings of the Ibero-American conference on artificial intelligence (IBERAMIA-94), Caracas, 25–28 Oct 1994

Melnik MI, Alm J (2002) Does a seller's eCommerce reputation matter? Evidence from eBay auctions. J Ind Econ 50(3):337–349

Miceli M, Castelfranchi C (2000) The role of evaluation in cognition and social interaction. In: Dautenhahn K (ed) Human cognition and social agent technology. Benjamins, Amsterdam, pp 225–262, chapter 9

Moukas A, Zacharia G, Maes P (1999) Amalthaea and histos: multiAgent systems for WWW sites and reputation recommendations. In: Klusch M (ed) Intelligent information agents: agent-based information discovery and management on the internet. Springer, Berlin, pp 292–322

Noon M, Delbridge R (1993) News from behind my hand: gossip in organizations. Organ Stud 14:23–36

Nowak MA, Sigmund K (1998a) Evolution of indirect reciprocity by image scoring. Nature 393:573–577

Nowak MA, Sigmund K (1998b) The dynamics of indirect reciprocity. J Theor Biol 194:561–574

Paine R (1967) What is gossip about? An alternative hypothesis. Man 2(2):278–285

Panchanathan K, Boyd R (2004) Indirect reciprocity can stabilize cooperation without the second-order free rider problem. Nature 432:499–502

Paolucci M (2000) False reputation in social control. Adv Complex Syst 3(1–4):39–51

Paolucci M (ed) (2009) Proceedings of the first international conference on reputation: theory and technology – ICORE 09, Gargonza. Institute of Science and Technology of Cognition, National Research Centre (ISTC-CNR), Rome. http://pagesperso-systeme.lip6.fr/Erika.Rosas/pdf/ICORE09.pdf

Paolucci M, Conte R (2009) Reputation: social transmission for partner selection. In: Trajkovski GP, Collins SG (eds) Handbook of research on agent-based societies: social and cultural interactions. IGI, Hershey, pp 243–260

Paolucci M et al (2009) Theory and technology of reputation (Technical report, FP6 Research Project “Social knowledge for e-Governance” (eRep)). Institute of Science and Technology of Cognition, National Research Centre (ISTC-CNR), Rome. http://issuu.com/mario.paolucci/docs/erep_booklet

Pinyol I, Paolucci M, Sabater-Mir J, Conte R (2008) Beyond accuracy: reputation for partner selection with lies and retaliation. In: Antunes L, Paolucci M, Norling E (eds) Multi-agent-based simulation VIII: international workshop, MABS 2007, Honolulu, 15 May 2007, revised and invited papers (Lecture notes in computer science 5003). Springer, Berlin, pp 128–140

Porter M (1998) Clusters and the new economics of competition. Harv Bus Rev 76:77–90

Ramchurn SD, Huynh D, Jennings NR (2004a) Trust in multiagent systems. Knowl Eng Rev 19(1):1–25

Ramchurn SD, Sierra C, Godo L, Jennings NR (2004b) Devising a trust model for multi-agent interactions using confidence and reputation. Int J Appl Artif Intell 18:833–852

Rasmusson L (1996) Socially controlled global agent systems (Working paper). Department of Computer and Systems Science, Royal Institute of Technology, Stockholm

Rasmusson L, Janson S (1996) Simulated social control for secure internet commerce. In: Proceedings of the 1996 workshop on new security Paradigms, Lake Arrowhead. ACM Press, New York, pp 18–25

Regan K, Cohen R (2005) Indirect reputation assessment for adaptive buying agents in electronic markets. In: Proceedings of workshop business agents and the semantic web (BASeWEB’05), 8 May 2005, Victoria, British Columbia

Resnick P, Zeckhauser R (2002) Trust among strangers in internet transactions: empirical analysis of eBay’s reputation system. In: Baye MR (ed) The economics of the internet and e-commerce, vol 11, Advances in applied microeconomics. Elsevier Science, Amsterdam, pp 127–157

Ripperger T (1998) Ökonomik des Vertrauens: Analyse eines Organisationsprinzips. Mohr Siebeck, Tübingen

Rocco E, Warglien M (1995) La comunicazione mediata da computer e l’emergere dell’opportunismo elettronico. Sistemi Intelligenti 7(3):393–420

Rose C, Thomsen S (2004) The impact of corporate reputation on performance: some Danish evidence. Eur Manage J 22(2):201–210

Saam N, Harrer A (1999) Simulating norms, social inequality, and functional change in artificial societies. J Artif Soc Soc Simul 2(1), http://jasss.soc.surrey.ac.uk/2/1/2.html

Sabater J, Paolucci M (2007) On representation and aggregation of social evaluations in computational trust and reputation models. Int J Approx Reason 46(3):458–483

Sabater J, Sierra C (2002) Reputation and social network analysis in multi-agent systems. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems (AAMAS 2002), 15–19 July 2002, Bologna. ACM Press, New York, pp 475–482

Sabater J, Sierra C (2005) Review on computational trust and reputation models. Artif Intell Rev 24(1):33–60

Sabater J, Paolucci M, Conte R (2006) REPAGE: REPutation and ImAGE among limited autonomous partners. J Artif Soc Soc Simul 9(2), http://jasss.soc.surrey.ac.uk/9/2/3.html

Schillo M (1999) Vertrauen: Ein Mechanismus zur sicheren Koalitionsbildung in künstlichen Gesellschaften (Trust: a mechanism for reliable coalition formation in artificial societies). Master’s thesis, Department of Computer Science, Saarland University, Saarbrücken, Germany.

Schillo M, Funk P, Rovatsos M (2000) Using trust for detecting deceitful agents in artificial societies. Appl Artif Intell 14:825–848

Sell J, Wilson R (1991) Levels of information and contributions to public goods. Soc Forces 70: 107–124

Sommerfeld RD, Krambeck H, Milinski M (2008) Multiple gossip statements and their effects on reputation and trustworthiness. Proc R Soc B 275:2529–2536

Squazzoni F, Boero R (2002) Economic performance, inter-firm relations and local institutional engineering in a computational prototype of industrial districts. J Artif Soc Soc Simul 5(1). http://jasss.soc.surrey.ac.uk/5/1/1.html

Utz S, Matzat U, Snijders CCP (2009) On-line reputation systems : the effects of feedback comments and reactions on building and rebuilding trust in on-line auctions. International Journal of Electronic Commerce 13(3):95–118

Wilson DS, Wilczynski C, Wells A, Weiser L (2000) Gossip and other aspects of language as group-level adaptations. In: Heyes C, Huber L (eds) The evolution of cognition. MIT Press, Cambridge, pp 347–366

Yamagishi T, Matsuda M, Yoshikai N, Takahashi H, Usui Y (2009) Solving the lemons problem with reputation. An experimental study of online trading. In: Cook KS, Snijders C, Vincent B, Cheshire C (eds) eTrust: forming relationships in the online world. Russel Sage Foundation, New York, pp 73–108

Yu B, Singh MP (2000) A social mechanism of reputation management in electronic communities. In: Klusch M, Kerschberg L (eds) Cooperative information agents IV, the future of information agents in cyberspace, 4th international workshop, CIA 2000, Boston, 7–9 July 2000, proceedings (Lecture notes in computer science, 1860). Springer, Berlin, pp 154–165

Zacharia G (1999) Trust management through reputation mechanisms. In: Castelfranchi C, Falcone R, Firozabadi BS (eds) Proceedings of the workshop on deception, fraud and trust in agent societies at autonomous agents'99, Seattle. pp 163–167

Zacharia G, Moukas A, Maes P (1999) Collaborative reputation mechanisms in electronic marketplaces. In: Proceedings of the 32nd annual Hawaii international conference on system sciences (HICSS-32), 5–8 Jan 1999, Maui, Track 8: software technology. IEEE computer society, Los Alamitos

# Chapter 16
# Social Networks and Spatial Distribution

**Frédéric Amblard and Walter Quattrociocchi**

**Why Read This Chapter?** To learn about interaction topologies for agents, from social networks to structures representing geographical space, and the main questions and options an agent-based modeller has to face when developing and initialising a model.

**Abstract** In most agent-based social simulation models, the issue of the organisation of the agents' population matters. The topology, in which agents interact, – be it spatially structured or a social network – can have important impacts on the obtained results in social simulation. Unfortunately, the necessary data about the target system is often lacking, therefore you have to use models in order to reproduce realistic spatial distributions of the population and/or realistic social networks among the agents. In this chapter we identify the main issues concerning this point and describe several models of social networks or of spatial distribution that can be integrated in agent-based simulation to go a step forward from the use of a purely random model. In each case we identify several output measures that allow quantifying their impacts.

## 16.1 Introduction

Independent of the methodology followed to build a model, any agent-based modeller has to face not only the design of the agents and their behaviour, but also the design of the topology in which the agents interact. This can be a spatial structure, so that the agents are distributed within a representation of geographical space, or a social structure, linking agents as nodes in a network, or even both.

---

F. Amblard (✉)
IRIT, Université Toulouse 1 Capitole, Toulouse, France
e-mail: frederic.amblard@ut-capitole.fr

W. Quattrociocchi
LABSS, University of Siena, Italy

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_16,

Interaction topologies can be determined explicitly or implicitly. They are *explicit* when they are specified as modelling hypotheses and thus clearly defined within the model. Conversely, topologies are *implicit* when they are inferred from other processes and thus not a definite part of the model. We will come back to the consequences of such a classification in the next section.

There are three major issues to solve when dealing with social and spatial structures:

*Implementation*: How to represent the structure in the model, e.g. continuous vs. discrete representation of space or which data structure to choose for the social network;

*Initialisation*: How to initialise the chosen structure, e.g. which initial shape of the network should be chosen and how should the population of agents be distributed on this network;

*Observation*: How to characterise a given structure and/or its evolution, potentially taking into account agents' states related to their place in the structure. This latter point raises the question of the indicators to observe during the simulation in order to follow changes in either spatial or social structures.

Since the answers to these questions differ quite substantially for spatial and social structures, we will discuss them separately in Sects. 16.3 and 16.4. The following section, while focussing on the topic of explicitly versus implicitly defined interaction topologies, will also discuss the situation where both social and spatial structures have to be taken into account at the same time; leading to either spatially-embedded networks or graph-like representations of spatial structures.

## 16.2 Explicit and Implicit Structures

As mentioned in the introduction, there is a difference between explicit and implicit structures. We define explicit structures as clearly implemented modelling hypotheses, which can therefore be identified in the model. Implicit structures, on the other hand, are not directly defined in the model and are rather determined as a result of other processes and/or hypotheses in the model. We will demonstrate the difference with the help of two examples and then go on to explore the consequences of ex- versus implicitness with regard to the three issues of implementation, initialisation and observation introduced above.

### *16.2.1 Example 1: Schelling's Segregation Model*

To present it briefly, the segregation model of Thomas Schelling (1971) is composed of a set of agents, some red, the others green. Each agent is positioned on an empty square of a chessboard (representing the environment). If the

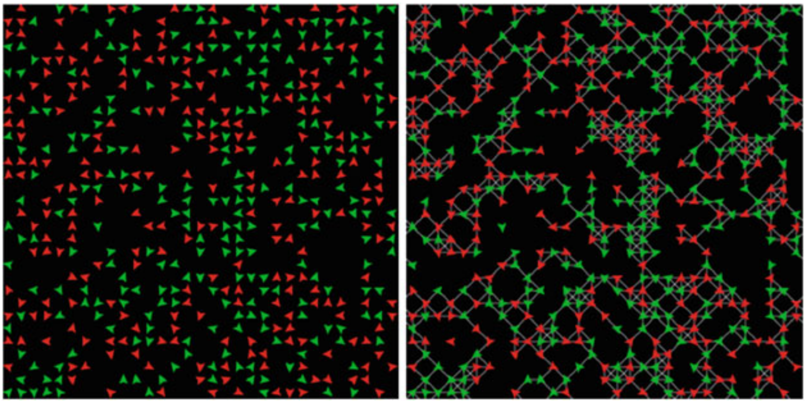


**Fig. 16.1** Schelling's segregation model in NetLogo, the spatial structure is represented on the *left* and the inferred social structure represented as a graph on the *right* figure

proportion of neighbours of the same colour falls below the value of the agent's tolerance variable, the agent moves to a randomly chosen empty square (originally: the nearest empty square at which it would be satisfied with the neighbourhood); otherwise, it stays where it is.

The spatial structure of this model is explicit and is represented explicitly as a grid. As such, it is discrete, regular and static (the distribution of the agents on this structure evolves, not the structure itself). The social structure, on the other hand, is implicit in this model. During the simulation the agents take into account the type (green or red) of their neighbours on the grid, but the corresponding social structure is not defined as such and is inferred from the spatial distribution of the agents (cf. Fig. 16.1). As discussed later, it can be interesting in such a case to characterise the implicit social structure of the model, as this is the one that drives the model, whereas the spatial structure merely acts as a constraint (in particular concerning the maximal number of neighbours) on the evolution of the social structure.

### *16.2.2 Example 2: IMAGES Innovation Dynamics Model*

In the innovation dynamics model developed in the FAIR-IMAGES project (IMAGES 2004) agents, representing farmers, are linked via a social network. This defines the paths, over which information is diffused, and controls which agents influence each other. The social network is determined, at least in part, from the geographical locations of the agents to account for the fact that geographical neighbours tend to know each other. In this example the social structure is explicit as it is built into the model by the hypothesis (Fig. 16.2).


**Fig. 16.2** Reconstruction of the social network among agents incorporating the geographical distance (IMAGES 2004)

### *16.2.3 Questions Linked to the Implicit/Explicit Property of the Structures*

The property of being explicit or implicit enables us to narrow down the range of possible answers to the three questions raised in the introduction. To begin with, the question of implementation can only be asked when dealing with explicit structures; the same is true for the question of initialisation. However, characterising the implicit social structure in spatial models, i.e. identifying at a given time step the whole set of interactions among agents that could or do take place, can give useful hints for understanding the underlying dynamics. Identifying, for instance, separate components in the implicit social network inferred from a spatial model is more informative than solely identifying spatial clusters as it confirms that there is effectively no connection among the different groups.

In the next two sections we will discuss the conditions in which social and spatial structures can be considered as independent features of social analysis and can therefore be presented independently. This is generally the case but, as we will detail in the last section, there are some exceptions.

## 16.3 Social Networks

Social networks have been analysed extensively during the last decade. From the social network of scientists (co-authorship network or co-citation network) (Newman 2001, 2004; Jeong et al. 2002; Meyer et al. 2009) to the social network of dolphins (Lusseau 2003), many empirical studies on large graphs popularised this fascinating subject: 'social' links between social beings. Neither empirical analysis nor theoretical modelling is new in this field. From the formalisation of graphs by Euler in the eighteenth century in order to represent paths in Königsberg, which led to the now well established graph theory, to social network analysis in sociology, originating from the use of the socio-matrix of Moreno, social networks are now quite commonly used in agent-based simulation to explicitly represent the topology of interactions among a population of agents.

In order to clarify the kind of modelling issues you are dealing with, we can divide modelling of social networks into three categories: (a) static networks, (b) dynamic networks with the dynamics independent of the agents' states (for instance random rewiring process), (c) dynamic networks evolving dependent on the agents' states. In this chapter we will concentrate on the first case since it is the most common; although the use of the second case has recently started to grow rapidly, while the third case is still in its incipient stage.

In each case, the same three questions arise with regard to implementation, initialisation and observation:

- Which data structure is best suited to represent the network?
- Which initial (and in the static case, only) network configuration to use?
- How to identify something interesting from my simulation including the network (e.g. a social network effect)?

### 16.3.1 Which Data Structure to Use?

Although it could seem trivial, especially when you use a high-level modelling platform such as NetLogo or Repast, this issue is important concerning the execution efficiency of your model. More important even, depending on your choice, biases are linked to some data structures when using particular classes of networks such as scale-free networks.

Basically using an object-oriented approach, you have two choices: either to embed social links within the agent as pointers to other agents or to externalise the whole set of links as a global collection (called Social Network, for instance). The former is more practical when having $1 - n$ interactions rather than *1–1* interactions, i.e. taking into account all neighbours' states to determine the new state of the agent rather than picking one agent at random in the neighbourhood. The difference between the two solutions is mainly related to the scheduling you will use. You can choose either to first schedule the agents, picking an agent at random from the population and then selecting one (or more) of its social links. Or you can choose to pick a random link from the global collection and then execute the corresponding interaction. While this choice will depend a lot on the kind of model you are implementing, it is crucial when using scale-free networks since both options may produce a bias and you will have to choose the solution that is more relevant for the purpose of your model.

To be able to explain this further, we need to quickly introduce scale-free networks (they will be presented in more detail in Sect. 16.3.2.4). Their main property is that the distribution of the number of links per agent (node) follows a power law, meaning that very few agents – the so-called "hubs" – have a lot of links while the majority of agents have only a few.

If you now choose to schedule the agents first, you effectively apply an egalitarian rule on the individuals, which however results in the links involving the hubs being less frequently scheduled than the links involving agents with few social relations. Take for example a population of 100 agents where one agent has 30 links and another has only one link. Each of these two agents has the same probability to be scheduled (0.01), but if you then proceed to select a random link from the scheduled agent, the links of the hub agent each have a 0.01/30 probability to be chosen while the one link of the other agent still has a 0.01 probability to be picked.

Conversely, if you schedule the collection of links first, i.e. apply an egalitarian rule on the individual links, the influence of the hubs in the global dynamics will be strengthened, as they are involved in more links. Therefore, the initial states of the hub agents are also interesting with respect to the whole dynamic of the model.

## *16.3.2 Which Initial Network Configuration to Use?*

This question mainly arises when choosing which (initial) social structure to implement. There is a large choice of models – each with some advantages and some drawbacks – that can be distinguished into four categories: (a) regular graphs, lattices or grids being the most longstanding structure used in social simulation (inherited from the cellular automata approaches), (b) random graphs, (c) small-world networks, and (d) scale-free networks. Concerning the latter three categories, Newman et al. (2006) regroup an important set of articles that can be useful as advanced material on this point.

### 16.3.2.1 Lattices

The field of social modelling inherited many tools from mathematics and physics and in particular cellular automata (Wolfram 1986). The corresponding underlying interaction structure is then in general a grid and in many cases a torus. The cells of the automata represent the agents and their social neighbourhood is defined from the regular grid with a von Neumann or a Moore neighbourhood. The von Neumann neighbourhood links a cell to its four adjacent cells (North, East, South, West) while a Moore neighbourhood adds four more neighbours (NE, SE, SW, NW; cf. Fig. 16.3).

The main advantage of using regular grids stems from visualisation, regular grids enabling very efficient visualisations of diffusion processes or clustering process (cf. Fig. 16.4).

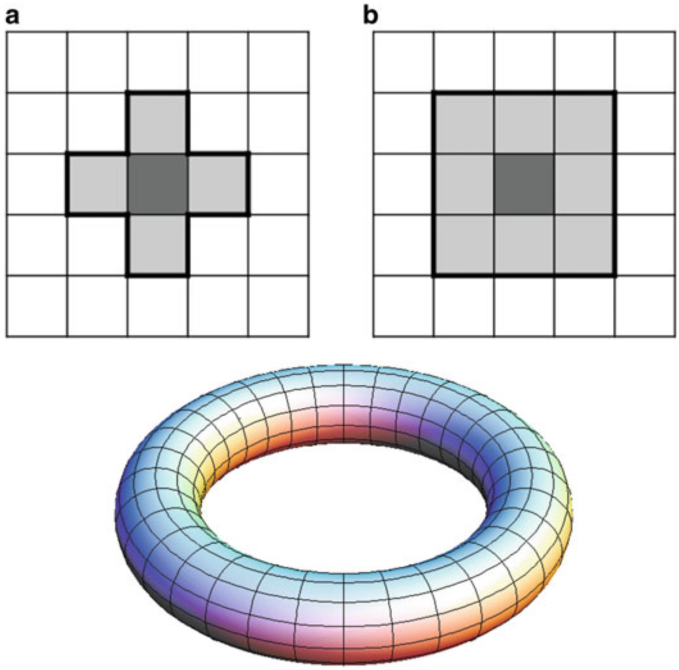



**Fig. 16.3** von Neumann (**a**) and 3 × 3 Moore neighbourhood (**b**) on a regular grid; the illustration on the right shows a torus, i.e. the result of linking the borders (north with south, east with west) of a grid (Flache and Hegselmann 2001)

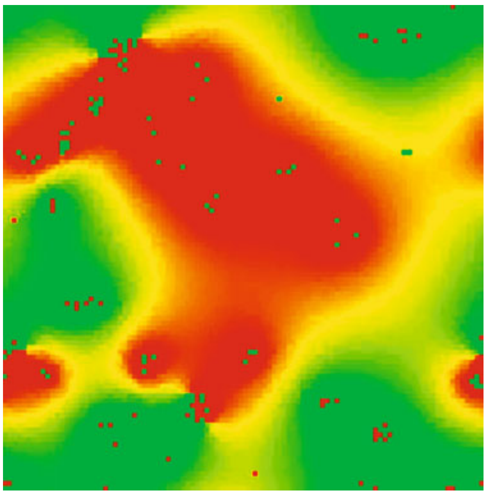


**Fig. 16.4** Opinion dynamics model on a regular grid (Jager and Amblard 2005)

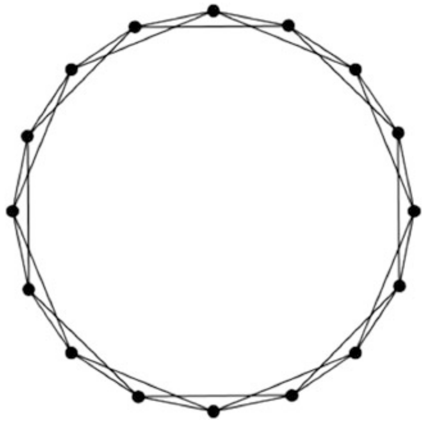


**Fig. 16.5** Regular 1D structure with connectivity = 4

The first important point concerning the regular structures deals with connectivity. In contrast to other kinds of networks (random ones for instance, see next section), using regular networks makes it difficult to change the connectivity of the structure, i.e. the number of links per agent. The exploration of connectivity effects on the model behaviour is limited in this case to specific values (4, 8, 12, 24… in the case of a two-dimensional regular grid).

The second point deals with the dimension of the regular structure. A one-dimensional lattice corresponds to a circle (see Fig. 16.5), two-dimensional structures to grids (chessboard), three-dimensional structures to cubic graphs. However, we have to notice that only 2D regular structures benefit from a visualization advantage, higher dimensions suffering from the classical disadvantage associated with the visualization of dynamics on complex graphs.

The presence or absence of borders is important in regular graphs. Classic example is the 2D grid, which – if not implemented as a torus – is not a regular structure anymore, since agents localised at the borders have fewer connections. Moreover, these agents being linked to each other creates a bias in the simulation (Chopard and Droz 1998). This bias is sometimes needed for any dynamics to happen and could either correspond to a modelling hypothesis or to an unwanted artefact. This point is probably far clearer on one-dimensional graphs, where if you do not close the circle the diameter[1] is approximately the size of the population, whereas if you close it, the diameter is half the population size. This issue corresponds to border effects identified on cellular automata.

[1] The diameter of a graph is defined as the length of the longest shortest-path in the graph.

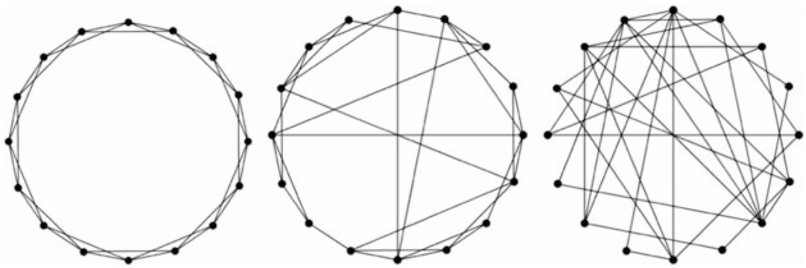


**Fig. 16.6** The β-model of Watts (1999) enables to go from regular graphs (on the *left*) to random graphs (on the *right*) using rewiring of edges

#### 16.3.2.2 Random Graphs

Another kind of model that can be used to generate social structures is the random graph model (Solomonoff and Rapoport 1951; Erdös and Renyi 1960); see Fig. 16.6 on the right for an illustration. As told by Newman et al. (2006), there are two ways to build random graphs containing $n$ vertices: one (denoted as $G_{n,m}$) is to specify the number $m$ of edges between vertex pairs chosen at random, the other (denoted as $G_{n,p}$) is to specify a probability $p$ for an edge to link any two vertices. Both of them correspond to graphs that have on average the same properties when they are big enough. The only difference is that in $G_{n,m}$ the number of edges is fixed, while in $G_{n,p}$ it may fluctuate from one instance to the other; but on average it is also fixed.

The first important property of random graphs is that they show a phase transition when the average degree of a vertex is 1. Below this transition we obtain a number of small components, while above this threshold the model exhibits a giant component with some isolated nodes. The giant component is a subset of the graph vertices, each of which is reachable from any of the others along some path(s). The random graphs that are used most often in agent-based social simulation are slightly above the threshold, in the giant-component phase with some isolated nodes.

Another important point concerns the degree distribution. The properties and behaviour of a network are affected in many ways by its degree distribution (Albert et al. 2000; Cohen et al. 2000; Callaway et al. 2000). $G_{n,p}$ has a binomial degree distribution (Poisson distribution for large $n$), which is sharply peaked and has a tail that decays quicker than any exponential distribution.

It is possible to define random graphs with any desired degree distribution (Bender and Canfield 1978; Luczak 1992; Molloy and Reed 1995). In this case, one considers graphs with a given degree sequence rather than with a degree distribution. A degree sequence is a set of degrees $k_1$, $k_2$, $k_3$... for each of the corresponding vertices *1, 2, 3*.... Molloy and Reed (1995) suggest the following algorithm:

- Create a list in which the label $i$ of each vertex appears exactly $k_i$ times.
- Pair up elements from this list uniformly at random until none remain.
- Add an edge to the graph joining the two vertices of each pair.

According to Molloy and Reed (1995), these graphs possess a phase transition at which a giant component appears, just as in the standard Poisson random graph.

In the context of agent-based social simulation, a great advantage of random graphs over regular graphs is that you can easily change and precisely tune the average connectivity of the graph and – applying Molloy and Reed's algorithm – the distribution of the edges among vertices.

Replacing regular graphs with random graphs, several scientists experienced a "social network" effect, i.e. models having different macroscopic behaviours depending on the chosen interaction structure (Stocker et al. 2001, 2002; Holme and Grönlund 2005; Huet et al. 2007; Deffuant 2006; Gong and Xiao 2007; Kottonau and Pahl-Wostl 2004; Pujol et al. 2005). The fact is that these two classes of networks have very different characteristics. In terms of clustering, regular graphs exhibit more clustering or local redundancy than the random graphs. On the other hand, random graphs lead to a shorter diameter and average path length among the pairs of individuals than a regular graph. Mean path length for a random graph scales logarithmically with graph size. For more details concerning random graphs models, we refer the interested reader to (Bollobas 2001).

#### 16.3.2.3 Small World Networks

The question arising at this stage is: are there classes of graphs between these two extremes (random and regular graphs) that may have other characteristics? Watts and Strogatz (1998) introduced such a model, motivated by the observation that many real world graphs share two main properties:

- *The small-world effect*, i.e. most vertices are connected via a short path in the network.[2]
- *High clustering,* corresponding to the phenomenon that the more neighbours two individuals have in common, the more likely they are to be connected themselves.

Watts and Strogatz defined a network to be a small-world network if it exhibits both of these properties, that is, if the mean vertex-vertex distance $l$ is comparable to that of a random graph and the clustering coefficient is comparable to that of a regular lattice.

To construct such a network, Watts and Strogatz found the following algorithm. Starting from a regular lattice with the desired degree, each link has a probability $p$ to be rewired, i.e. to be disconnected from one of its vertices and reconnected with another vertex chosen uniformly at random. The result is the creation of shortcuts in the regular structure. Watts and Strogatz (1998) imposed additional constraints on the rewiring process: a vertex cannot be linked to itself, and any two vertices cannot

[2] Short path being defined by Watts and Strogatz as comparable to those found in random graphs of the same size and average degree.

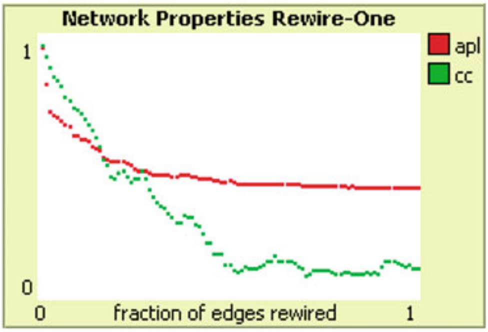



**Fig. 16.7** Average path length (*dark grey*), and clustering coefficient (*light grey*) represented as the probability of rewiring evolves

be linked by more than one edge. Moreover, the rewiring process only rewired one end and not both ends of the link. These added conditions prevent the resulting network from being a random graph even in the limit $p = 1$.

A simplified version of this model was proposed by Newman and Watts (1999). Starting with a lattice and taking each link one by one, add another link between a pair of vertices chosen at random with the probability $p$ without removing the existing one. This corresponds to the addition of $Lkp$ new links on average to the starting lattice, $Lk$ being the initial number of links in the graph (Fig. 16.7).

A number of other models have been proposed to achieve the combination of short path length with high clustering coefficient. The oldest one is the random biased net of Rapoport (1957), in which clustering is added to a random graph by *triadic closure*, i.e. the deliberate completion of connected triples of vertices to form triangles in the network, thereby increasing the clustering coefficient. Another attempt to model clustering was made in the 1980s by Holland and Leinhardt (1981), using the class of network models known as *exponential random graphs*. Another method for generating clustering in networks could be membership of individuals in groups. This has been investigated by Newman et al. (2001; Newman 2003b) using so-called *bipartite graph* models.

#### 16.3.2.4 Scale-Free Networks

Even with the small-world effect, the hypothesis that complex systems such as cells or social systems are based upon components – i.e. molecules or individuals – randomly wired together, has proven to be incomplete.

In fact, several empirical data analysis of real networks found out that for many systems, including citation networks, the World Wide Web, the Internet, and metabolic networks, the degree distribution approximates a power law (Price 1965; Albert et al. 1999; Faloustos et al. 1999; Broder et al. 2000).

This corresponds to a new class of network since neither of the previously discussed networks such as random graphs and small-world models have a power-law degree distribution. Barabási and Albert (1999) called these graphs *scale-free networks* and proposed that power laws could potentially be a generic property of many networks and that the properties of these networks can be explained by having the graph grow dynamically rather than being static. Their paper proposes a specific model of a growing network that generates power-law degree distributions similar to those seen in the World Wide Web and other networks.

Their suggested mechanism has two components: (1) the network is growing, i.e. vertices are added continuously to it, and (2) vertices gain new edges in proportion to the number they already have, a process that Barabási and Albert call *preferential attachment*.[3] Therefore, the network grows by addition of a single new vertex at each time step, with $m$ edges connected to it. The other end of the edge is chosen at random with probability proportional to degree:

$$P(k_i) = \frac{k_i}{\sum_j k_j} \tag{16.1}$$

An extension to this model proposed by the same authors (Albert and Barabási 2000) follows another way of building the graph. In this model one of three events occurs at each time step:

- With probability $p$, $m$ new edges are added to the network. One end of each new edge is connected to a node selected uniformly at random from the network and the other to a node chosen using the preferential attachment process according to the probability given just before.
- With probability $q$, $m$ edges are rewired, meaning that a vertex $i$ is selected at random and one of its edges chosen at random is removed and replaced with a new edge whose other end is connected to a vertex chosen again according to the preferential attachment process.
- With probability $1 - p - q$, a new node is added to the network. The new node has $m$ new edges that are connected to nodes already present in the system via preferential attachment in the normal fashion

This model produces a degree distribution that again has a power-law tail, with an exponent $\gamma$ that depends on the parameters $p$, $q$ and $m$, and can vary anywhere in the range from 2 to $\infty$.

Dorogovtsev et al. (2000) consider a variation on the preceding model applied to directed graphs, in which preferential attachment takes place with respect only to the incoming edges at each vertex. Each vertex has $m$ outgoing edges, which attach to other pre-existing vertices with attachment probability proportional only to those vertices' in-degree.

[3] The preferential attachment mechanism has appeared in several different fields under different names. In information science it is known as cumulative advantage (Price 1976), in sociology as the Matthew effect (Merton 1968), and in economics as the Gibrat principle (Simon 1955).

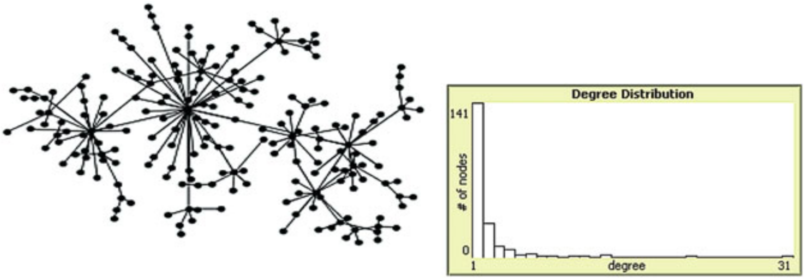



**Fig. 16.8** Scale-free network generated using the preferential attachment model implemented in NetLogo (*left*) and the distribution of number of nodes per degree (*right*), which follows a scale-free distribution

Another extension proposed by Krapivsky et al. (2000) explores what happens when $\Pi(k)$, the probability distribution of connectivity per node, is not linear in $k$. They studied the case in which $\Pi(k)$ takes the power-law form $\Pi(k) \propto k^{\alpha}$

In a subsequent model, Bianconi and Barabási (2001), motivated by the Google effect (emergence of a new hub from an existing scale-free network), introduced the idea that some nodes are intrinsically "better" or faster growing than others. In the model, each vertex in the network has a fixed fitness value $\eta_i$, chosen from some distribution $\rho(\eta)$, that corresponds to an intrinsic ability to compete for edges at the expense of other vertices. Each such vertex connects to $m$ others, the probability of connecting to vertex $i$ being proportional to the product of the degree and the fitness of $i$:

$$\Pi = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \tag{16.2}$$

The use of models like preferential attachment in an agent-based social simulation context follows two motivations that are relatively distinct. On the one hand, such models are used as initial configurations for the simulation: therefore the construction of the scale-free network, even if it could be considered as a model of a growing network, should rather be envisaged as an algorithm to build the initial state of the simulation. On the other hand, it can also be seen as a subject of research, the focus of a model, for instance in the field of modelling social network dynamics. In this case, preferential attachment mechanisms could be included in the model of social network evolution (if such a mechanism is relevant, of course) (Fig. 16.8).

As the structure and the evolution of a scale-free network cannot be considered as separate concepts, the same holds true for the topology and the overlying interactions among components. For instance, Pastor-Satorras and Vespignani (2001) discovered that the epidemic threshold on top of scale-free networks converges towards 0, meaning that in scale-free topologies even weak viruses diffuse fast within the network. In particular the dynamic aspects of real phenomena occurring on top of scale-free networks would allow investigating several properties of real social networks, for instance how changes in the network structure trigger emergent phenomena by local interaction.

### *16.3.3 How to Distribute Agents on the Network?*

This question deals with the efficiency of the model, because each link between nodes (agents) represents an interaction, the shape of the topology, i.e. the interaction space that we are interested in, characterising its emergent properties. To exemplify, let us take the example of an opinion dynamics model including extremist agents that have a significant impact on the global dynamics. What should the initial state of the population be? Are the extremists homogeneously distributed in the population, i.e. over the social network, or are they organised in dense communities, or any case in between? This question could have an important impact on the emerging phenomenon and is the consequence of two choices: a modelling choice, i.e. what is the initial state of your model, and a more technical choice, i.e. the possibility to generate the distribution wanted. With regard to the latter, you can always generate an ad-hoc situation either using real data or building them artificially; however, it is often the case that you wish to use a generic way to distribute agents among a network. Usually, the starting point would be to use a random distribution. You first generate you network and you distribute the different agents uniformly over this network assuming that each agent's state is independent from its location in the network. However, you do not have to believe strongly in any social determinism to question this independence and to wish to explore other kinds of distributions. Two cases occur then:

You have modelling hypotheses concerning the relation between the existence of a link between two agents and their actual state. In this case you would probably proceed using a stochastic algorithm that enables some flexibility applying this relation.

You are only able to describe the global state you want to attain. In this case, you probably have to operate iterative permutations among the agents. For each permutation, you have to compute the result. If it is better (i.e. closer to the state wanted) then you keep the permutation if not you reject this permutation and keep the previous state.

Note that the latter solution can become very expensive with regard to the evaluation of a given state and is not guaranteed to obtain the wanted final state in a finite time. Therefore the most prudent approach would be to take the agents' characteristics into account when deciding on the presence of a link between two agents. Such a solution has been proposed by Thiriot and Kant (2008) explicitly drawing on agents' attributes and link types for the generation of interaction networks. In this perspective, the generated social network is a part of the modelling hypotheses and it has been shown that the corresponding graphs differ a lot from the ones obtained with abstract models like the ones described in this section.

### 16.3.4 Which Measures to Use?

Once you have your model running, especially when it deals with agents changing their state over a (possibly evolving, irregular) social network, the first image is usually a messy network with overlapping links and nodes changing colours over time. Without specific efforts dedicated to the network layout (which is not a trivial task) the global picture is not understandable at first glance. Therefore you would want to add some indicators enabling you to better understand what happens in the model. There are two main categories: the first one – which is currently under development – concerns indicators linking the states of nodes with their position in the network, while the more classical indicators of the second category help to characterise the structure of the network only.

Some important properties associated with graphs influence the characteristics of the dynamics taking place on the graph. This is mainly the case for the diameter, especially dealing with diffusion processes, and the clustering coefficient, dealing for instance with influence processes. In the particular case of a regular network, where all nodes are equal, it can be determined depending on the connectivity and the dimension of the graph. The average path length gives information that is equivalent. The clustering coefficient is defined locally as, for a given node, the rate of existing links among its neighbours compared to the number of possible links. Local redundancy of links is therefore important as an inertial process that can reinforce or go against an influence dynamics. In the following paragraphs we will briefly describe the main indicators you can use to gain some insight about the phenomena emerging in the model.

A first question you could ask is, looking at a particular dimension of the agents' state vector, do they have a tendency to regroup or cluster according to this state or not. Or phrased in a different way: do connected agents tend to become similar? A useful indicator for this would be an averaged similarity measure over the network, calculating the mean distance among all connected agents of the population.

In order to characterise the two main features of small-world networks (i.e. small world effect and high clustering) several indicators are used. The small-world effect is measured simply by averaging the distance among any pair of vertices in the graph. This results in the average path length index.

Concerning the clustering, a number of indicators have been proposed. Watts and Strogatz (1998) suggested the *clustering coefficient*, used classically in social network analysis and consisting in averaging a local clustering coefficient over all vertices of the network. The local clustering coefficient $C_i$ of the vertex $i$ is defined as the number of existing links among the neighbours of the vertex $i$, divided by the number of possible links among these neighbours. This quantity is 1 for a fully connected graph but tends towards 0 for a random graph as the graph becomes large.

The problem is that this indicator is heavily biased in favour of vertices with a low degree due to the small number of possible links (denominator) they have. When averaging the local clustering coefficient over all nodes without additional weighting, this could make a huge difference in the value of $C$. A better way to calculate the average probability of a pair of neighbours being connected is to count the total number of pairs of vertices on the entire graph that have a common neighbour and the total number of such pairs that are themselves connected, and divide the one by the other. Newman et al. (2001) have expressed this, i.e. the fraction of transitive triples, as:

$$C = 3^* \text{ (number of triangles on the graph)}/\text{(numberof connected triples of vertices)}$$

Another definition for clustering coefficient proposed by Newman (2003a) consists in calculating the probability that the friend of your friend is also your friend, resulting in:

$$C = 6^* \text{ (number of triangles on the graph)}/\text{(number of paths of length 2)}$$

Dealing with small-world networks, Watts and Strogatz (1998) defined a network to have high clustering if $C >> C_{rg}$ (this latter being the clustering coefficient for random graphs).

Watts and Strogatz defined a network to be a small-world network if it shows both of those properties, that is, if the mean vertex-vertex distance $l$ is comparable with that on a random graph ($l_{rg}$) and the clustering coefficient is much greater than that for a random graph. Walsh (1999) used this idea to define the *proximity ratio:*

$$\mu = \frac{C/C_{rg}}{l/l_{rg}} \tag{16.3}$$

which is of order 1 on a random graph, but much greater than 1 on a network obeying the Watts-Strogatz definition of a small-world network.

One of the most important properties of social networks is the so-called notion of *power*. As a shared definition of power is still object of debate, the design of metrics able to characterise its causes and consequences is a pressing challenge. In particular the social network approach emphasises the concept of power as inherently relational, i.e., determined by the network topology. Hence, the focus must be put on the relative positions of nodes. In order to characterise such a property the concept of *centrality* has emerged. The simplest centrality metric, namely the degree centrality, measures the number of edges that connect a node to other nodes in a network. Over the years many more complex centrality metrics have been proposed and studied, including status score (Katz 1953), α-centrality (Bonacich and Lloyd 2001), betweenness centrality (Freeman 1979), and several others based on the random walk, the most famous of which is the eigenvector centrality used by Google's PageRank algorithm (Page et al. 1999). The temporal declination of these concepts is meaningful, and Kossinets et al. (2008) have shown that nodes that are topologically more central are not necessarily central from a temporal point of view, hence the concept of temporal centrality. The temporalization of network metrics is currently a pressing scientific challenge (Casteigts et al. 2010; Santoro et al. 2011).

### 16.3.5 What Kind of Network Effect Can You Anticipate?

As mentioned by Newman et al. (2006), the behaviour of dynamical systems on networks is one of the most challenging areas of research. At the end of their paper, Watts and Strogatz (1998) present one of the most tractable problems in the field that is the spread of disease using the SIR (susceptible/infective/removed) epidemic model. They measure the position of the epidemic threshold – the critical value of the probability $r$ of infection of a susceptible individual by an infective one at which the disease first becomes an epidemic, spreading through the population rather than dying out. They found a clear decline in the critical $r$ with increasing $p$, the probability of links rewiring, indicating that the small-world topology makes it easier for the disease to spread. They also found that the typical time for the disease to spread through the population decreases while increasing $p$.

One of the simplest models of the spread of non-conserved information on a network is one in which an idea or a disease starts at a single vertex and spreads first to all its neighbouring vertices. From these it spreads to all of their neighbours, and so forth, until there are no accessible vertices left that have not yet been infected. This process is known in computer science as breadth-first search. Depending on the type of network model chosen, it could happen (this is mostly the case on real networks) that you obtain a large number of small components plus, optionally, a single giant component. If we simulate the spread of a rumour or disease on such a network using breadth-first search, either we will see a small outbreak corresponding to one of the small components, or, if there is a giant component and we happen to start a breadth-first search within it, we will see a giant outbreak that fills a large portion of the system. The latter is precisely what we call an epidemic, when we are talking about disease, and the phase transition at which the giant component appears in a graph is also a phase transition between a regime in which epidemics are not possible and a regime in which they are. In fact, if the disease starts its breadth-first search at a randomly chosen vertex in the network, then the probability of seeing an epidemic is precisely equal to the fraction of the graph occupied by the giant component. By studying, either analytically or numerically, the behaviour of various epidemiological models on networks, the authors hope to get a better idea of how real diseases will behave, and in some cases they have found entirely new behaviours that had not previously been observed in epidemiological models.

Ball et al. (1997) consider networks with two levels of mixing, meaning that each vertex in the network belongs both to the network as a whole and to one of a specified set of subgroups (e.g. family) with different properties of spread within the network. Disease spreading is again modelled using the SIR model. In the model, people can be in one of three states: susceptible (S), meaning they can catch the disease but haven't yet; infective (I), meaning they have caught the disease and can pass it on to others; and removed (R), meaning they have recovered from the disease and can neither pass it on nor catch it again, or they have died. Ball et al. (1997) found that the rapid spread of a disease within groups such as families can lead to epidemic outbreaks in the population as a whole, even when the probability of inter-family communication of the disease is low enough that epidemic outbreaks normally would not be possible. The reason for this is the following. If transmission between family members takes place readily enough that most members of a family will contract the disease once one of them does, then we can regard the disease as spreading on a super network in which vertices are the families, not individuals. Roughly speaking, spread of the disease between families will take place with $n^2$ times the normal person-to-person probability, where $n$ is the number of people in a family.

An alternative approach to calculating the effect of clustering on SIR epidemics has been presented by Keeling (1999). What Keeling's method does is to include, in approximate form, the effect of the short-scale structure of the network – the clustering – but treat everything else using a standard fully mixed approximation. Thus things like the effect of degree distributions are absent from his calculations. But the effect of clustering is made very clear. Keeling finds that a lower fraction of the population need to be vaccinated against a disease to prevent an epidemic if the clustering coefficient is high.

In another paper, Pastor-Satorras and Vespignani (2001) address the behaviour of an endemic disease model on networks with scale-free degree distributions. Their motivation for the work was an interest in the dynamics of computer virus infections, which is why they look at scale-free networks; computer viruses spread over the Internet and the Internet has a scale-free form, as demonstrated by Faloutsos et al. (1999). They used a derivation of the SIR model, the SIS (susceptible/infected/susceptible) model, which simply considers that individuals recover with no immunity to the disease and are thus immediately susceptible once they have recovered. In their work, Pastor-Satorras and Vespignani grow networks according to the scale-free model of Barabási and Albert (1999) and then simulate the SIS model on this network starting with some fixed initial number of infective computers. Pastor-Satorras and Vespignani do not find oscillations in the number of infected individuals for any value of the independent parameter of the model. No matter what value the parameter takes, the system is always in the epidemic regime; there is no epidemic threshold in this system. No matter how short a time computers spend in the infective state or how little they pass on the virus, the virus remains endemic. Moreover, the average fraction of the population infected at any one time decreases exponentially with the infectiousness of the disease.

Watts (2002) has looked at the behaviour of cascading processes. Unlike disease, the spread of some kinds of information, such as rumours, fashions, or opinion, depends not only on susceptible individuals having contacts with infective ones, but on their having contact with such individuals in sufficient numbers to persuade them to change their position or beliefs on an issue. People have a threshold for adoption of trends. Each individual has a threshold $t$ for adoption of the trend being modelled, which is chosen at random from a specified distribution. When the proportion of a person's contacts that have already adopted the trend rises above this threshold, the person will also adopt it. This model is similar to the rioting model of Granovetter (1978). Watts gives an exact solution for his model on random graphs for the case where initially a low density of vertices has adopted the trend. The solution depends crucially on the presence of individuals who have very low thresholds $t$. In particular, there must exist a sufficient density of individuals in the network whose thresholds are so low that they will adopt the trend if only a single one of their neighbours does. As Watts argues, the trend will only propagate and cause a cascade if the density of these individuals is high enough to form a percolating sub-graph in the network. The fundamental result of this analysis is that, as a function of the average degree $z$ of a vertex in the graph, there are two different thresholds for cascading spread of information. Below $z = 1$, no cascades happen because the network itself has no giant component. Cascades also cease occurring when $z$ is quite large, the exact value depending on the chosen distribution of $t$. The reason for this upper threshold is that as $z$ becomes large, the value of $t$ for vertices that adopt the trend when only a single neighbour does so becomes small, and hence there are fewer such vertices. For large enough $z$ these vertices fail to percolate and so cascades stop happening.

Another field of application deals with the robustness of networks. If we have information or disease propagating through a network, how robust is the propagation to failure or removal of vertices? The Internet, for example, is a highly robust network because there are many different paths by which information can get from any vertex to any other. The question can also be rephrased in terms of disease. If a certain fraction of all the people in a network are removed in some way from a network – by immunization against disease, for instance – what effect will this have on the spread of the disease?

Albert et al. (2000) discuss network resilience for two specific types of model networks, random graphs and scale-free networks. The principal conclusion of the paper is that scale-free networks are substantially more robust to the random deletion of vertices than Erdös-Rényi random graphs, but substantially less robust to deletion specifically targeting the vertices with the highest degrees. The mean vertex-vertex distance in the scale-free network increases as vertices are deleted, but it does so much more slowly than in the random graph. Similarly, the size of the largest component goes down substantially more slowly as vertices are deleted in the scale-free network than in the random graph. By contrast, the random graph's behaviour is almost identical whether one deletes the highest-degree vertices or vertices chosen at random. Thus scale-free networks are highly robust to random failure, but highly fragile to targeted attacks. More recent work has shown that there are networks with even higher resilience than scale-free networks (Costa 2004; Rozenfeld and ben-Avraham 2004; Tanizawa et al. 2005).

As a conclusion on networks effects, the existence of isolated components and the average degree of the graph are the first important factors that play a role in the dynamics occurring in agent-based social simulation. After that, depending on the kind of phenomenon studied, social influence or diffusion dynamics for instance, clustering coefficient and average path length should be considered.

## 16.4 Spatial Structure

### *16.4.1 Which Spatial Structure to Use?*

Despite the increased use of social networks in agent-based simulation, in many cases one has to take into account the spatial dimension. In this section, we will first present some issues dealing with this point, then some models that allow the distribution of a population of agents in a geographical space, as well as some measures to try and qualify such distributions and their effects on the simulation.

#### 16.4.1.1 Torus or Not?

There may also be arbitrary effects introduced by the spatial bounds or limits placed on the phenomenon or study area. This occurs since spatial phenomena may be unbounded or have ambiguous transition zones in reality. In the model, ignoring spatial dependency or interaction outside the study area may create edge effects. For example, in Schelling's segregation model there is a higher probability to stabilise when you have fewer neighbours (e.g. three in the corners or five on a border) for a particular density.

The choice of spatial bounds also imposes artificial shapes on the study area that can affect apparent spatial patterns such as the degree of clustering. A possible solution is similar to the sensitivity analysis strategy for the modifiable areal unit problem, or MAUP: change the limits of the study area and compare the results of the analysis under each realization. Another possible solution is to over-bound the study area, i.e. to deliberately model an area that encompasses the actual study area. It is also feasible to eliminate edge effects in spatial modelling and simulation by mapping the region to a boundless object such as a torus or sphere.

### *16.4.2 How to Distribute Agents on the Space?*

To begin with the simplest case, let us present the Poisson process in a 2D continuous space. Such a distribution can be used in particular to define the null hypothesis of a totally random structure for spatial distribution that will enable us to compare other kinds of spatial structures. In order to simulate a Poisson process of intensity $\lambda$ on a domain of surface $D$, we first define the number of points $N$ to be distributed, picking it at random from a Poisson law of parameter $\lambda D$. For each point $A_i$ the coordinates $x_i$ and $y_i$ are therefore taken at random from a uniform law. For a rectangular domain, it is sufficient to bound the values of $x$ and $y$ depending on the studied domain. For more complex zones, the method can be the same, deleting the points in a larger rectangle that are not part of the specific zone. Studying a population where the number of agents $N$ is known, we can use the same method

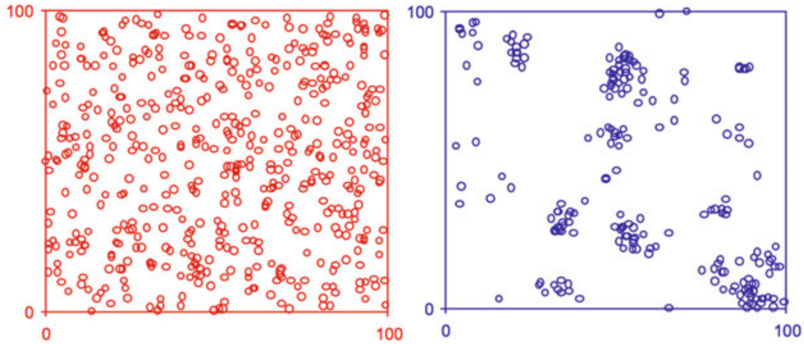



**Fig. 16.9** Two different spatial distributions, random (*left*) and aggregated (*right*) (From Goreaud 2000)

without picking the number $N$ at random and using the known value instead. The corresponding process is called a binomial process (Tomppo 1986). Moreover, note that because in a Poisson process random points can be infinitely close, the space occupied by the represented agents is not taken into account. Some processes have been developed to deal with this (Ripley 1977; Cressie 1993), corresponding to a random repartition of the points under the constraint that two points have to be at least $2R$ apart, where $R$ is the specified radius of an agent. It can also be interpreted as the random repartition of non-overlapping disks of radius $R$.

Practically, spatial structures are rarely totally random and it is quite frequent to have more aggregated structures. The Neyman-Scott cluster process can be used to simulate such more elaborated structures (Ripley 1977; Cressie 1993). The Neyman-Scott process is built from a master Poisson process whose $N_{ag}$ (number of aggregates) points are used as centres for the aggregates. Each aggregate is then composed of a random number of points whose positions are independent and distributed within a radius $R$ around the centre of the aggregate (Fig. 16.9).

In order to simulate complex interaction, the Gibbs process can be useful. The general class of Gibbs processes allows obtaining various complex structures (Tomppo 1986). The main idea of the Gibbs process is to distribute the points of the pattern according to attraction/repulsion relations with a range. Such a process can be defined using a cost function $f(r)$ which represents the cost associated to the presence of two points in the pattern separated by a distance $r$. For a given point pattern we can calculate a total cost, equal to the sum of the costs associated to each couple of points (see Eq. 16.4 below). By definition, the lower the total cost obtained on a pattern, the stronger is the probability for this pattern to be the result of the Gibbs process considered. For a given $r$, if the cost function $f(r)$ is positive, the probability is low to have two points at a distance $r$ (we could say that there is repulsion at distance $r$). Conversely, if the cost function $f(r)$ is negative, it is highly probable to find a couple of points separated by a distance $r$ (we could say there is attraction at a distance $r$).

$$totalCost = \sum_{i,j} f(d(A_i, A_j)) \tag{16.4}$$

Where $d$ is the distance between points $A_i$ and $A_j$.

To obtain the most probable realization of the process, we aim to simulate patterns whose total costs are low enough, i.e. get closer to a minimum of the cost function. Such patterns can be obtained using an algorithm of depletions and replacements: starting for instance from a Poisson pattern, we iterate a big number of times a modification of this pattern, which consists in moving randomly an arbitrarily chosen point. If this move increases the total cost, the previous configuration is kept unchanged, else the new position is kept. Then, the attractions and repulsions among points lead to a progressive reorganization of the original pattern toward a pattern that is more aggregated or more regular. Following Tomppo (1986), this algorithm converges quickly enough towards a realization of the process.

In the preceding paragraphs, we were discussing the spatial distribution of homogeneous points. What if the points are heterogeneous? In the case of heterogeneous patterns that do not correspond to the superposition of independent distributions of different homogeneous groups, the preceding methodology can be used independently for each class of points.

In order to consider a more difficult case, let us consider that now each point can have a radius as well as a colour. More generally speaking, the variables associated to the points are called "marks" and the following approach can be applied either to qualitative (the colour) or quantitative (the radius) marks. We can then define stochastic processes whose realizations are marked point patterns. These are marked punctual processes, some mathematical objects that can generate an infinite number of marked point patterns, all different, but sharing some common properties, in particular the spatial structure of points and marks (Stoyan 1987; Goulard et al. 1995).

For a marked pattern, one can consider the position of points (the pattern as such) and the attribution of marks separately. The positioning of points can be done using tools and methods from the preceding part. Rather than to consider probabilities of having points separated by a distance $r$, we now have to reason about the joint probability $P$ (see Eq. 16.5) that two points at a distance $r$ have marks $m$ and $m'$. This function expresses neighbourhood relationships among values of a mark. The function $g_M(r,m,m')$ can moreover be linked to the number of neighbours having a mark $m'$ at a distance $r$ of a point marked $m$.

$$\mathrm{P}(\mathrm{M}(\mathrm{A}_1) = \mathrm{m} \text{ and } \mathrm{M}(\mathrm{A}_2) = \mathrm{m}') = \lambda_\mathrm{M}(\mathrm{m})\ \lambda_\mathrm{M}(\mathrm{m}')\mathrm{g}_\mathrm{M}(\mathrm{r}, \mathrm{m}, \mathrm{m}') \tag{16.5}$$

The hypothesis of a totally random distribution of marks can be used as null hypothesis for marked stochastic processes. Then if, for a given case, the points of a subpopulation $SP_2$ are less often near the points of a subpopulation $SP_1$ than in the null hypothesis, we can speak about repulsion between the two subpopulations considered. Conversely, if the points from $SP_2$ are more numerous in the vicinity of $SP_1$ than in the null hypothesis we can talk about attraction of $SP_2$ by $SP_1$.

One can easily use a "depletions and replacements" algorithm with an associated cost function in this particular case to generate marked point patterns of the wanted properties with regard to inter- and intra-attraction/repulsion among the elements of the different groups.

This situation can be seen as two different cases that could be treated using nearly the same methods: the case where mark values are drawn at random from a known statistical law and the case where mark values are fixed from an existing distribution (histogram for instance) and where we have to distribute these mark values spatially. This corresponds to a random distribution of marks and to a list of given marks, respectively.

### 16.4.3 Which Measures to Use?

To analyse a spatial structure, there are several established methods from spatial statistics to characterise the spatial structure of point patterns (Ripley 1981; Diggle 1983; Cressie 1993).

One can distinguish between methods based on quadrants, the required data for which are the number of individuals in the quadrants, having variable position and size (Chessel 1978), and methods based on distance, for which the distances among points or individuals or positions are used as input. Indicators à la "Clark and Evans" (Clark and Evans 1954) are classical examples of methods based on distance. They calculate for each point an index of spatial structure considering the $n$ (fixed) closest neighbours. The average value of this index can then be compared to the theoretical value calculated for a null hypothesis (for instance a Poisson process) in order to characterise the spatial structure with regard to aggregates or regularity, attraction or repulsion, etc. (Pretzsch 1993; Füldner 1995).

For a homogeneous punctual process having a density $\lambda$, Ripley (1976, 1977) showed that it is possible to characterise the second order propriety using a function $K(r)$ such that $\lambda K(r)$ is the expected value of the number of neighbours at a distance $r$ of a point chosen at random from the pattern. This function is linked directly to the function of density of a couple of points $g(r)$ defined previously.

$$\lambda K(r) = E(number\ of\ \ neighbours\ at\ distance <= r\ from\ A_i)$$

$$K(r) = \int_{s=0}^{r} g(s) 2\pi \ sds \qquad (16.6)$$

For the Poisson process, which serves as null hypothesis, the expected value of the number of neighbours at distance $r$ of a given point from the pattern is $\lambda\pi r^2$, and then $K(r) = \pi r^2$. For an aggregated process, the points have on average more

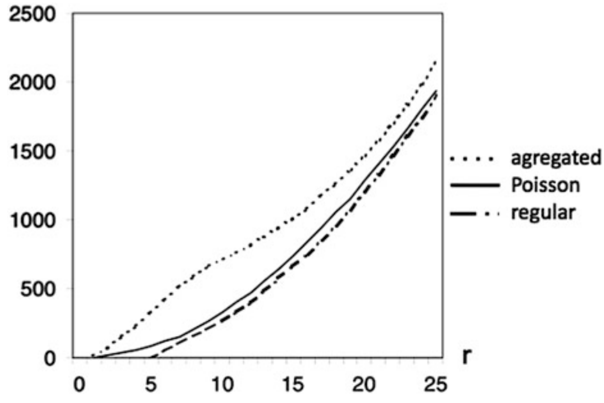



**Fig. 16.10** Estimation of the K(r) function for different repartitions (aggregated, poisson and regular) (From Goreaud 2000)

neighbours than in the null hypothesis and thus $K(r) > \pi r^2$. Conversely, for a regular process, the points have on average fewer neighbours than in the null hypothesis and $K(r) < \pi r^2$ (Fig. 16.10).

In the most frequent case where we do not know the process, the function $K(r)$ has to be estimated with the unique known realization: the real point pattern. We then approach the expected value of the number of neighbours around a given point using its average on the whole set of individuals belonging to the pattern. We then obtain a first approximated indicator of $K(r)$, noted as $\hat{K}(r)$ defined as follows:

$$\hat{K}(r) = \frac{1}{\lambda}\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i} k_{ij} \tag{16.7}$$

$N$ is the number of points in the studied domain of surface D; $\lambda = N/D$ is the estimator of the process density and $k_{ij}$ takes the value 1 if the distance between points $i$ and $j$ is below $r$ and 0 otherwise.

This first estimator is unfortunately biased as it underestimates the values of $K(r)$ because of the points situated close to the border of the studied domain having fewer neighbours than a given point of the process. This general problem, known as border or edge effect and discussed in Sect. 16.3.2.1, is met in a lot of methods for the analysis of spatial structure.

To correct the border effect, two methods are known from the literature:

- Ohser and Stoyan (Stoyan et al. 1995) propose a global correction method, which is based on the geometrical properties of the studied domain. This method has the advantage of being quite simple, so quick enough to calculate.
- Ripley (1977) proposes a local correction method that is based on the contribution of each point situated nearby the border. This local correction has the advantage of being equally usable to calculate the individual indices relative to each point of the pattern, just as those proposed by Getis and Franklin (1987). Some works seem to show that this class of estimators is more robust than the estimators of the Ohser and Stoyan type (Kiêu and Mora 1999).

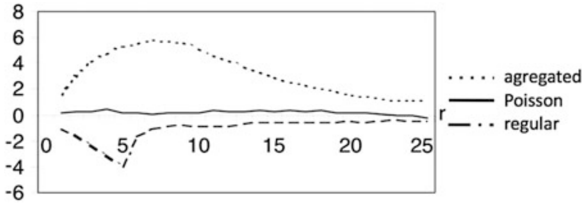



**Fig. 16.11** Estimation of the L(r) function for different repartitions (aggregated, poisson and regular) (From Goreaud 2000)

This latter solution consists in replacing the coefficient $k_{ij}$ with the inverse of the proportion of the perimeter of the circle $C_{ij}$ (centred on $A_i$ and passing through $A_j$) for the points situated near the border of the domain studied. It corresponds to an estimation of the number of points situated at the same distance that would be outside of this domain. Ripley (1977) shows this estimator is not biased.

$$k_{ij} = \frac{totalPerimeter}{PerimeterInsideTheZone} = \frac{2\pi \mathrm{r}}{C_{inside}} \geq 1 \tag{16.8}$$

An alternative to the function *K(r)* is the function *L(r)* proposed by Besag (1977) which is easier to interpret. For a Poisson process, for any distance *r*, *L(r)* = 0. Aggregated processes and regular ones are situated under and above the x-axis, respectively.

$$L(r) = \sqrt{\frac{K(r)}{\pi}} - r \text{ estimated by } \hat{L}(r) = \sqrt{\frac{\hat{K}(r)}{\pi}} - r \tag{16.9}$$

However appropriate such methods of punctual processes are to deal with agent-based simulations and allow initialising and measuring and therefore characterising such systems, agent-based simulations have an important aspect that is not addressed by this literature: the dynamics. In many cases, the agents in such systems will move in the environment and the spatial properties of the system is represented more accurately considering a set of trajectories of the agents rather than a succession of static spatial repartition. Even though punctual processes enable to capture spatial clustering effects, when considering for example the Boids model (Reynolds 1987) the dynamics of the flock will be overlooked. Therefore there is a need for statistical tools that enable to characterise the interrelations between sets of trajectories to analyze some models properly (Fig. 16.11).

### *16.4.4 What Effects Can You Anticipate?*

In spatial agent-based simulation, and without aiming at being exhaustive, the main phenomena you can look at and therefore characterise depend greatly on the population of agents in the model. Dealing with a homogeneous population of agents, you may observe spatial clustering, i.e. the formation of groups in space, that could be characterised by the methods presented beforehand. Introducing heterogeneity into the population, the main question deals with the spatial mixing of subpopulations, as present for instance in Schelling's segregation model. Dealing with a population of agents which may change their states, methods to characterise diffusion processes could be used. However, one of the most interesting points in this case does not really deal with the efficiency of the diffusion (evolution of the number of infected agents in an epidemiological model for instance), but rather with the characterization of the shape of such a diffusion. In this case, considering a diffusion process that takes place, you aim at characterising the shape of the interface rather than the global phenomenon that takes place. To this aim, fractal analysis gives interesting hints but is not appropriate for many phenomena, especially dynamical ones.

Finally, an important phenomenon that can occur at the macro-level of any spatial agent-based simulation concerns the density-dependence of the results. For an identified phenomenon (if we take for instance the segregation model of Schelling), there will be an impact of the density of agents (either locally or globally) on the appearance of the phenomenon. This is exactly the case of segregation in Schelling's model, where a very low number of empty places (thus, a high density of agents) can freeze the system.

## 16.5 Conclusion

This chapter aims at presenting ways to deal with the distribution of agents in social simulation. The two kinds of distributions considered are distribution over a graph or social network on the one hand or spatial distribution on the other hand. While these two cases are very common in social simulation too little effort is spent on either the characterization or the investigation of the impact of the distribution on the final results. The methods presented in this chapter are certainly not exhaustive and not even pertinent to all cases but they do present a first step towards pointing the curiosity of agent-based modellers at techniques that would definitely be useful for social simulation.

## Further Reading

The literature on dynamic aspects of social networks is rapidly developing. For social network models and their analysis we currently recommend (Newman et al. 2006). For spatial aspects we recommend (Diggle 1983). For more details concerning random graphs models, we refer the interested reader to (Bollobas 2001).

## References


Albert R, Barabási A-L (2000) Topology of evolving networks: local events and universality. Phys Rev Lett 85:5234–5237

Albert R, Jeong H, Barabási A-L (1999) Diameter of the World-Wide Web. Nature 401:130–131

Albert R, Jeong H, Barabási A-L (2000) Error and attack tolerance of complex networks. Nature 406:379–381

Ball F, Mollisson D, Scalia-Tomba G (1997) Epidemics with two levels of mixing. Ann Appl Probab 7:46–89

Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512

Bender EA, Canfield ER (1978) The asymptotic number of labeled graphs with given degree sequences. J Comb Theory A 24:296–307

Besag J (1977) Contribution to the discussion of Dr Ripley's paper. J Roy Stat Soc B 39:193–195

Bianconi G, Barabási A-L (2001) Competition and multiscaling in evolving networks. Europhys Lett 90:436–442

Bollobas B (2001) Random graphs, 2nd edn. Academic, New York

Bonacich P, Lloyd P (2001) Eigenvector-like measures of centrality for asymmetric relations. Soc Networks 23(3):191–201

Broder A et al (2000) Graph structure in the web. Comput Networks 33:309–320

Callaway DS, Newman MEJ, Strogatz SH, Watts DJ (2000) Network robustness and fragility: percolation on random graphs. Phys Rev Lett 85:5468–5471

Casteigts A, Flocchini P, Quattrociocchi W, Santoro N (2010) Time-varying graphs and dynamic networks. arXiv: 1012.0009v3

Chessel D (1978) Description non paramétrique de la dispersion spatiale des individus d'une espèce. In: Legay JM, Tomassone R (eds) Biométrie et écologie. Société française de biométrie, pp 45–133

Chopard B, Droz M (1998) Cellular automata modeling of physical systems. Cambridge University Press, Cambridge

Clark PJ, Evans FC (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 35(4):445–453

Cohen R, Erez K, Ben-Avraham D, Havlin S (2000) Resilience of the internet to random breakdowns. Phys Rev Lett 85:4626–4628

Costa LD (2004) Reinforcing the resilience of complex networks. Phys Rev E 69, 066127

Cressie NAC (1993) Statistics for spatial data. Wiley, New York

de Price DJS (1965) Networks of scientific papers. Science 149:510–515

de Price DJS (1976) A general theory of bibliometric and other cumulative advantage processes. J Am Soc Inform Sci 27:292–306

Deffuant G (2006) Comparing extremism propagation patterns in continuous opinion models. J Artif Soc Soc Simul 9(3). http://jasss.soc.surrey.ac.uk/9/3/8.html

Diggle PJ (1983) Statistical analysis of spatial point patterns. Academic, New York

Dorogovtsev SN, Mendes JFF, Samukhin AN (2000) Structure of growing networks with preferential linking. Phys Rev Lett 85:4633–4636

Erdös P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5:17–61

Faloustos M, Faloustos O, Faloustos C (1999) On power-law relationships of the internet topology. Comput Commun Rev 29:251–262

Flache A, Hegselmann R (2001) Do irregular grids make a difference? Relaxing the spatial regularity assumption in cellular models of social dynamics. J Artif Soc Soc Simul 4(4). http://www.soc.surrey.ac.uk/JASSS/4/4/6.html

Freeman LC (1979) Centrality in social networks: conceptual clarification. Soc Networks 1:215–239

Fuldner K (1995) Strukturbeschreibung von buchen-edellaubholz-Mischwäldern. Cuvillier Verlag, Göttingen

Getis A, Franklin J (1987) Second order neighborhood analysis of mapped point patterns. Ecology 68(3):473–477

Gong X, Xiao R (2007) Research on multi-agent simulation of epidemic news spread characteristics. J Artif Soc Soc Simul 10(3). http://jasss.soc.surrey.ac.uk/10/3/1.html

Goreaud F (2000) Apports de l'analyse de la structure spatiale en forêt tempérée à l'étude et la modélisation des peuplements complexes (Diploma Thesis). ENGREF (École nationale du génie rural des eaux et forêts), Nancy

Goulard M, Pages L, Cabanettes A (1995) Marked point process: using correlation functions to explore a spatial data set. Biom J 37(7):837–853

Granovetter M (1978) Threshold models of collective behaviour. Am J Sociol 83(6):1420–1443

Holland PW, Leinhardt S (1981) An exponential family of probability distributions for directed graphs. J Am Stat Assoc 76:33–65

Holme P, Grönlund A (2005) Modelling the dynamics of youth subcultures. J Artif Soc Soc Simul 8(3). http://jasss.soc.surrey.ac.uk/8/3/3.html

Huet S, Edwards M, Deffuant G (2007) Taking into account the variations of neighbourhood sizes in the mean-field approximation of the threshold model on a random network. J Artif Soc Soc Simul 10(1). http://jasss.soc.surrey.ac.uk/10/1/10.html

IMAGES (2004) Improving agri-environmental policies: a simulation approach to the role of the cognitive properties of farmers and institutions (Final report of the FAIR3 CT 2092 – IMAGES project), http://wwwlisc.clermont.cemagref.fr/ImagesProject/FinalReport/final%20report%20V2last.pdf

Jager W, Amblard F (2005) Uniformity, bipolarisation and pluriformity captured as generic stylised behaviour with an agent-based simulation model of attitude change. Comput Math Organ Theory 10:295–303

Jeong H et al (2002) Evolution of the social network of scientific collaborations. Physica A 311: 590–614

Katz L (1953) A new status derived from sociometric analysis. Psychometrika 18:39–43

Keeling MJ (1999) The effects of local spatial structure on epidemiological invasion. Proc R Soc London B 266:859–867

Kieu K, Mora M (1999) Estimating the reduced moments of a random measure. Adv Appl Probab Sect Stoch Geom Stat Appl 31(1):48–62

Kossinets G, Kleinberg J, Watts D (2008) The structure of information pathways in a social communication network. In: Li Y, Liu B, Sarawagi S (eds) proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'08). ACM, New York, pp 435–443

Kottonau J, Pahl-Wostl C (2004) Simulating political attitudes and voting behaviour. J Artif Soc Soc Simul 7(4). http://jasss.soc.surrey.ac.uk/7/4/6.html

Krapivsky PL, Redner S, Leyvraz F (2000) Connectivity of growing random networks. Phys Rev Lett 85:4629–4632

Luczak T (1992) Sparse random graphs with a given degree sequence. In: Frieze AM, Luczak T (eds) Proceedings of the symposium on random graphs, Poznan 1989. Wiley, New York, pp 165–182

Lusseau D (2003) The emergent properties of a dolphin social network. Proc R Soc London B 270(2): S186–S188

Merton RK (1968) The Matthew effect in science. Science 159:56–63

Meyer M, Lorscheid I, Troitzsch K (2009) The development of social simulation as reflected in the first ten years of JASSS: a citation and co-citation analysis. J Artif Soc Soc Simul 12(4). http://jasss.soc.surrey.ac.uk/12/4/12.html

Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. Random Struct Algor 6:161–179

Newman MEJ (2001) The structure of scientific collaboration networks. PNAS 98(2):404–409

Newman MEJ (2003a) Ego-centered networks and the ripple effect. Soc Networks 25:83–95
Newman MEJ (2003b) Properties of highly clustered networks. Phys Rev E 68, 016128
Newman MEJ (2004) Co-authorship networks and patterns of scientific collaboration. PNAS 101:5200–5205
Newman MEJ, Watts DJ (1999) Scaling and percolation in the small-world network model. Phys Rev E 60:7332–7342
Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. Phys Rev E 64, 026118
Newman MEJ, Barabási A-L, Watts DJ (2006) The structure and dynamics of networks. Princeton University Press, Princeton
Page L, Brin S, Motwani R, Winograd T (1999) The pageRank citation ranking: bringing order to the web (Technical report, 1999–66). Stanford InfoLab, Stanford University, http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf
Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Phys Rev Lett 86:3200–3203
Pretzsch H (1993) Analyse und reproduktion räumlicher bestandesstrukturen: versuche mit dem strukturgenerator STRUGEN Writings from the forestry faculty of the University of Gottingen 954 and NFV, 114
Pujol JM, Flache A, Delgado J, Sanguesa R (2005) How can social networks ever become complex? Modelling the emergence of complex networks from local social exchanges. J Artif Soc Soc Simul 8(4). http://jasss.soc.surrey.ac.uk/8/4/12.html
Rapoport A (1957) Contribution to the theory of random and biased nets. Bull Math Biophys 19:257–277
Reynolds CW (1987) Flocks, herds, and schools: a distributed behavioral model. Comput Graph 21(4):25–34
Ripley BD (1976) The second order analysis of stationary point process. J Appl Probab 13:255–266
Ripley BD (1977) Modelling spatial patterns. J R Stat Soc B 39:172–212
Ripley BD (1981) Spatial statistics. Wiley, New York
Rozenfeld HD, ben-Avraham D (2004) Designer nets from local strategies. Phys Rev E 70, 056107
Santoro N, Quattrociocchi W, Flocchini P, Casteigts A, Amblard F (2011) Time-varying graphs and social network analysis: temporal indicators and metrics. In: Proceedings of SNAMAS 2011, pp 33–38
Schelling T (1971) Dynamic models of segregation. J Math Sociol 1(1):143–186
Simon HA (1955) On a class of skew distribution functions. Biometrika 42:425–440
Solomonoff R, Rapoport A (1951) Connectivity of random nets. Bull Math Biophys 13:107–117
Stocker R, Green DG, Newth D (2001) Consensus and cohesion in simulated social networks. J Artif Soc Soc Simul 4(4). http://jasss.soc.surrey.ac.uk/4/4/5.html
Stocker R, Cornforth D, Bossomaier TRJ (2002) Network structures and agreement in social network simulations. J Artif Soc Soc Simul 5(4). http://jasss.soc.surrey.ac.uk/5/4/3.html
Stoyan D (1987) Statistical analysis of spatial point process: a soft-core model and cross correlations of marks. Biom J 29:971–980
Stoyan D, Kendall WS, Mecke J (1995) Stochastic geometry and its applications. Wiley, New York
Tanizawa T, Paul G, Cohen R, Havlin S, Stanley HE (2005) Optimization of network robustness to waves of targeted and random attacks. Phys Rev E 71, 047101
Thiriot S, Kant J-D (2008) Generate country-scale networks of interaction from scattered statistics. In: Squazzioni F (ed) Proceedings of the 5th conference of the European social simulation association, ESSA '08, Brescia
Tomppo E (1986) Models and methods for analysing spatial patterns of trees, vol 138, Communicationes instituti forestalis fenniae. The Finnish Forest Research Institute, Helsinki

Walsh T (1999) Search in a small world. In: Dean T (ed) Proceedings of the 16th international joint conference on artificial intelligence, IJCAI '99. Morgan Kauffman, San Francisco, pp 1172–1177
Watts DJ (1999) Small worlds: the dynamics of networks between order and randomness. Princeton University Press, Princeton
Watts DJ (2002) A simple model of global cascades on random networks. PNAS 99:5766–5771
Watts DJ, Strogatz SH (1998) Collective dynamics of small world networks. Nature 393:440–442
Wolfram S (1986) Theory and applications of cellular automata. World Scientific, Singapore

# Chapter 17
# Learning

**Michael W. Macy, Stephen Benard, and Andreas Flache**

**Why Read This Chapter?** To understand the properties of various individual or collective learning algorithms and be able to implement them within an agent (where evolution is a particular kind of collective learning).

**Abstract** Learning and evolution are adaptive or "backwards-looking" models of social and biological systems. Learning changes the probability distribution of traits within an individual through direct and vicarious reinforcement, while evolution changes the probability distribution of traits within a population through reproduction and selection. Compared to forward-looking models of rational calculation that identify equilibrium outcomes, adaptive models pose fewer cognitive requirements and reveal both equilibrium and out-of-equilibrium dynamics. However, they are also less general than analytical models and require relatively stable environments. In this chapter, we review the conceptual and practical foundations of several approaches to models of learning that offer powerful tools for modeling social processes. These include the Bush-Mosteller stochastic learning model, the Roth-Erev matching model, feed-forward and attractor neural networks, and belief learning. Evolutionary approaches include replicator dynamics and genetic algorithms. A unifying theme is showing how complex patterns can arise from relatively simple adaptive rules.

M.W. Macy (✉)
Department of Information Science, Cornell University, Ithaca NY 14853, USA
e-mail: m.macy@cornell.edu

S. Benard
Department of Sociology, Indiana University, 744 Ballantine Hall, Bloomington IN 47405

A. Flache
Department of Sociology, ICS, University of Groningen, Faculty of Behavioral and Social Sciences, NL-9712 TG Groningen, The Netherlands

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_17,

## 17.1 Introduction

Evolution and learning are basic explanatory mechanisms for consequentialist theories of adaptive self-organization in complex systems.[1] These theories are consequentialist in that behavioral traits are selected by their outcomes. Positive outcomes increase the probability that the associated trait will be repeated (in learning theory) or reproduced (in evolutionary theory), while negative outcomes reduce it. Explanatory outcomes might be rewards and punishments (in learning theory), survival and reproduction (in evolutionary models), systemic requisites (in functionalism), equilibrium payoffs (in game theory), or the interests of a dominant class (in conflict theory).

An obvious problem in consequentialist models is that the explanatory logic runs in the opposite direction from the temporal ordering of events. Behavioral traits are the explanandum and their outcomes the explanans. This explanatory strategy collapses into teleology unless mechanisms can be identified that bridge the temporal gap. While expected utility theory and game theory posit a forward-looking and analytic causal mechanism, learning and evolution provide a backward-looking and experiential link. In everyday life, decisions are often highly routine, with little conscious deliberation. These routines can take the form of social norms, protocols, habits, traditions, and rituals. Learning and evolution explain how these routines emerge, proliferate, and change in the course of consequential social interaction, based on *experience* instead of calculation. In these models, *repetition*, not prediction, brings the future to bear on the present, by recycling the lessons of the past. Through repeated exposure to a recurrent problem, the consequences of alternative courses of action can be iteratively explored, by the individual actor (learning) or by a population (evolution).

Backward-looking rationality is based on rules rather than choices (Vanberg 1994). A choice is an instrumental, case-specific comparison of alternative courses of action, while rules are behavioral routines that provide standard solutions to recurrent problems. Rules can take the form of strategies, norms, customs, habits, morals, conventions, traditions, rituals, or heuristics. Rule-based decision-making is backward-looking in that the link between outcomes and the actions that produce them runs backward in time. The outcomes that explain the actions are not those the action will produce in the future, they are the outcomes that were previously experienced when the rule was followed in the past.

Learning alters the probability distribution of behavioral traits within a given individual, through processes of direct and vicarious reinforcement. Evolution alters the frequency distribution of traits within a population, through processes of reproduction and selection. Whether selection operates at the individual or population level, the units of adaptation need not be limited to human actors but may include larger entities such as firms or organizations that adapt their behavior

[1] Much of the material in this chapter has been previously published in (Macy 1996, 1997, 1998, 2004; Macy and Flache 2002; Flache and Macy 2002).

in response to environmental feedback. Nor is evolutionary adaptation limited to genetic propagation. In cultural evolution, norms, customs, conventions, and rituals propagate via role modeling, occupational training, social influence, imitation, and persuasion. For example, a firm's problem-solving strategies improve over time through exposure to recurrent choices, under the relentless selection pressure of market competition. Sub-optimal routines are removed from the repertoires of actors by learning and imitation, and any residuals are removed from the population by bankruptcy and takeover. The outcomes may not be optimal, but we are often left with well-crafted routines that make their bearers look much more calculating than they really are (or need to be), like a veteran outfielder who catches a fly ball as if she had computed its trajectory.

## 17.2 Evolution

Selection pressures influence the probability that particular traits will be replicated, in the course of competition for scarce resources (ecological selection) or competition for a mate (sexual selection). Although evolution is often equated with ecological selection, sexual selection is at least as important. By building on partial solutions rather than discarding them, genetic crossover in sexual reproduction can exponentially increase the rate at which a species can explore an adaptive landscape, compared to reliance on trial and error. Paradoxically, sexual selection can sometimes inhibit ecological adaptation, especially among males. Gender differences in parental investment cause females to be choosier about mates and thus sexual selection to be more pronounced in males. An example is the peacock's large and cumbersome tail, which attracts the attention of peahens (who are relatively drab) as well as predators. Sexually selected traits tend to become exaggerated as males trap one another in an arms race to see who can have the largest antlers or to be bravest in battle.

Selection pressures can operate at multiple levels in a nested hierarchy, from groups of individuals with similar traits, down to individual carriers of those traits, down to the traits themselves. Evolution through group selection was advanced by Wynne-Edwards (1962, 1986) as a solution to one of evolution's persistent puzzles – the viability of altruism in the face of egoistic ecological counter pressures. Pro-social in-group behavior confers a collective advantage over rival groups of rugged individualists. However, the theory was later dismissed by Williams in *Adaptation and Natural Selection* (1966), which showed that between-group variation gets swamped by within-group variation as group size increases. Moreover, group selection can depend on differential rates of extinction, with no plausible mechanism for the whole-cloth replication of successful groups.

Sexual selection suggests a more plausible explanation for the persistence of altruistic behaviors that reduce the chances of ecological selection. Contrary to Herbert Spencer's infamous view of evolution as "survival of the fittest," generosity can flourish even when these traits are ecologically disadvantageous, by attracting females who have evolved a preference for "romantic" males who are ready to sacrifice for their partner. Traits that reduce the ecological fitness of an individual carrier can also flourish if the trait increases the selection chances of other individuals with that trait. Hamilton (1964) introduced this gene-centric theory of kin altruism, later popularized by Dawkins' in the *Selfish Gene* (1976).

Allison (1992) extended the theory to benevolence based on cultural relatedness, such as geographical proximity or a shared cultural marker. This may explain why gene-culture co-evolution seems to favor a tendency to associate with those who are similar, to differentiate from "outsiders", and to defend the in-group against social trespass with the emotional ferocity of parents defending their offspring.

This model also shows how evolutionary principles initially developed to explain biological adaptation can be extended to explain social and cultural change. Prominent examples include the evolution of languages, religions, laws, organizations, and institutions. This approach has a long and checkered history. Social Darwinism is a discredited nineteenth century theory that used biological principles as analogs for social processes such as market competition and colonial domination. Many sociologists still reject all theories of social or cultural evolution, along with biological explanations of human behavior, which they associate with racist and elitist theories of "survival of the fittest". Others, like the socio-biologist E. O. Wilson (1988, p. 167), believe "genes hold culture on a leash", leaving little room for cultural evolution to modify the products of natural selection. Similarly, evolutionary psychologists like Cosmides and Tooby search for the historical origins of human behavior as the product of ancestral natural selection rather than on-going social or cultural evolution.

In contrast, a growing number of sociologists and economists are exploring the possibility that human behaviors and institutions may be heavily influenced by processes of social and cultural selection that are independent of biological imperatives. These include DiMaggio and Powell (the new institutional sociology), Nelson and Winter (evolutionary economics), and Hannan and Freeman (organizational ecology).

One particularly compelling application is the explanation of cultural diversity. In biological evolution, speciation occurs when geographic separation allows populations to evolve in different directions to the point that individuals from each group can no longer mate. Speciation implies that all life has evolved from a very small number of common ancestors, perhaps only one. The theory has been applied to the evolution of myriad Indo-European languages that are mutually incomprehensible despite having a common ancestor. In socio-cultural models, speciation operates through homophily (attraction to those who are similar), xenophobia (aversion to those who are different), and influence (the tendency to become more similar to those to whom we are attracted and to differentiate from those we despise).

Critics counter that socio-cultural evolutionists have failed to identify any underlying replicative device equivalent to the gene. Dawkins has proposed the "meme" as the unit of cultural evolution but there is as yet no evidence that these exist. Yet

Charles Darwin developed the theory of natural selection without knowing that phenotypes are coded genetically in DNA. Perhaps the secrets of cultural evolution are waiting to be unlocked by impending breakthroughs in cognitive psychology.

The boundary between learning and evolution becomes muddied by a hybrid mechanism, often characterized as "cultural evolution." In cultural evolution, norms, customs, conventions, and rituals propagate via role modeling, occupational training, social influence, and imitation. Cultural evolution resembles learning in that the rules are soft-wired and can therefore be changed without replacing the carrier. Cultural evolution also resembles biological evolution in that rules used by successful carriers are more likely to propagate to other members of the population. For example, norms can jump from one organism to another by imitation (Dawkins 1976; Durham 1992; Boyd and Richerson 1985; Lopreato 1990). A successful norm is one that can cause its carrier to act in ways that increase the chances that the norm will be adopted by others. Cultural evolution can also be driven by social learning (Bandura 1977) in which individuals respond to the effects of vicarious reinforcement. Social learning and role-modeling can provide an efficient shortcut past the hard lessons of direct experience.

Imitation of successful role models is the principal rationale for modeling cultural evolution as an analog of natural selection (Boyd and Richerson 1985; Dawkins 1976). However, social influence differs decisively from sociobiological adaptation. Softwired rules can spread without replacement of their carriers, which means that reproductive fitness loses its privileged position as the criteria for replication. While "imitation of the fittest" is a reasonable specification of cultural selection pressures, it is clearly not the only possibility. Replication of hardwired rules may be a misleading model for cultural evolution, and researchers need to be cautious in using Darwinian analogs as templates for modeling the diffusion of cultural rules. In cultural models of "imitation of the fittest," actors must not only know which actor is most successful, they must also know the underlying strategy that is responsible for that success. Yet successful actors may not be willing to share this information. For very simple strategies, it may be sufficient to observe successful behaviors. However, conditional strategies based on "if-then" rules cannot always be deduced from systematic observation. Researchers should therefore exercise caution in using biological models based on Darwinian principles to model cultural evolution, which is a hybrid of the ideal types of evolution and learning.

## 17.3 Learning

The most elementary principle of learning is simple reinforcement. Thorndike (1898) first formulated the theory of reinforcement as the "Law of Effect," based on the principle that "pleasure stamps in, pain stamps out." If a behavioral response has a favorable outcome, the neural pathways that triggered the behavior are strengthened, which "loads the dice in favor of those of its performances which make for the most permanent interests of the brain's owner" (James 1981,

p. 143). This connectionist theory anticipates the error back-propagation used in contemporary neural networks (Rumelhart and McClelland 1988). These models show how highly complex behavioral responses can be acquired through repeated exposure to a problem.

Reinforcement theory relaxes three key behavioral assumptions in models of forward-looking rationality:

1. Propinquity replaces causality as the link between choices and payoffs.
2. Reward and punishment replace utility as the motivation for choice.
3. Melioration replaces optimization as the basis for the distribution of choices over time.

We consider each of these in turn.

1. *Propinquity, not causality.* Compared to forward-looking calculation, the Law of Effect imposes a lighter cognitive load on decision makers by assuming experiential induction rather than logical deduction. Players explore the likely consequences of alternative choices and develop preferences for those associated with better outcomes, even though the association may be coincident, "superstitious," or causally spurious. The outcomes that matter are those that have already occurred, not those that an analytical actor might predict in the future. Anticipated outcomes are but the consciously projected distillations of prior exposure to a recurring problem. Research using fMRI supports the view that purposive assessment of means and ends can take place *after* decisions are made, suggesting that "rational choice" may be not so much a theory of decision but a theory of how decisions are rationalized to self and others.

   Reinforcement learning applies to both intended and unintended consequences of action. Because repetition, not foresight, links payoffs back to the choices that produce them, learning models need not assume that the payoffs are the intended consequences of action. Thus, the models can be applied to expressive behaviors that lack a deliberate or instrumental motive. Frank's (1988) evolutionary model of trust and commitment formalizes the backward-looking rationality of emotions like vengeance and sympathy. An angry or frightened actor may not be capable of deliberate and sober optimization of self-interest, yet the response to the stimulus has consequences for the individual, and these in turn can modify the probability that the associated behavior will be repeated.
2. *Reward and punishment, not utility.* Learning theory differs from expected utility theory in positing two distinct cognitive mechanisms that guide decisions toward better outcomes, *approach* (driven by reward) and *avoidance* (driven by punishment). The distinction means that aspiration levels are very important for learning theory. The effect of an outcome depends decisively on whether it is coded as gain or loss, satisfactory or unsatisfactory, pleasant or aversive.
3. *Melioration, not optimization.* Melioration refers to suboptimal gradient climbing when confronted with what Herrnstein and Drazin (1991) call "distributed choice" across recurrent decisions. A good example of distributed choice is the decision whether to cooperate in an iterated Prisoner's Dilemma game. Suppose

each side is satisfied when the partner cooperates and dissatisfied when the partner defects. Melioration implies a tendency to repeat choices with satisfactory outcomes even if other choices have higher utility, a behavioral tendency March and Simon (1958) call “satisficing.” In contrast, unsatisfactory outcomes induce searching for alternative outcomes, including a tendency to revisit alternative choices whose outcomes are even worse, a pattern we call “dissatisficing.” While satisficing is suboptimal when judged by conventional game-theoretic criteria, it may be more effective in leading actors out of a suboptimal equilibrium than if they were to use more sophisticated decision rules, such as “testing the waters” to see if they could occasionally get away with cheating. Gradient search is highly path dependent and not very good at backing out of evolutionary *cul de sacs*. Course correction can sometimes steer adaptive individuals to globally optimal solutions, making simple gradient climbers look much smarter than they need to be. Often, however, adaptive actors get stuck in local optima. Both reinforcement and reproduction are biased toward *better* strategies, but they carry no guarantee of finding the highest peak on the adaptive landscape, however relentless the search. Thus, learning theory can be usefully applied to the equilibrium selection problem in game theory. In repeated games (such as an on-going Prisoner’s Dilemma), there is often an indefinitely large number of analytic equilibria. However, not all these equilibria are learnable, either by individuals (via reinforcement) or by populations (via evolution). Learning theory has also been used to identify a fundamental solution concept for these games – stochastic collusion – based on a random walk from a self-limiting non-cooperative equilibrium into a self-reinforcing cooperative equilibrium (Macy and Flache 2002).

## 17.4 Modeling Evolution

Replicator dynamics are the most widely used model of evolutionary selection (Taylor and Jonker 1978). In these models, the frequency of a strategy changes from generation to generation as a monotonic function of its “payoff advantage,” defined in terms of the difference between the average payoff of that strategy and the average payoff in the population as a whole. The more successful a strategy is on average, the more frequent it tends to be in the next generation.

Replicator dynamics typically assume that in every generation every population member encounters every other member exactly once, and replication is based on the outcome of this interaction relative to the payoff earned by all other members of the population. However, in natural settings, actors are unlikely to interact with or have information about the relative success of every member of a large population. The mechanism can also be implemented based on local interaction (limited to network “neighbors”) and local replication (neighbors compete only with one another for offspring).

The outcomes of replicator dynamics depend on the initial distribution of strategies, since the performance of any given strategy will depend on its effectiveness in interaction with other strategies. For example, aggressive strategies perform much better in a population that is accommodating than one that is equally aggressive. It is also not possible for replicator dynamics to invent new strategies that were not present at the outset.

These limitations are minimized by using genetic algorithms. The genetic algorithm was proposed by Holland (1975) as a problem-solving device, modeled after the recursive system in natural ecologies. The algorithm provides a simple but elegant way to write a computer program that can improve through experience. The program consists of a string of symbols that carry machine instructions. The symbols are often binary digits called "bits" with values of 0 and 1. The string is analogous to a chromosome containing multiple genes. The analog of the gene is a bit or combination of bits that comprises a specific instruction. The values of the bits and bit-combinations are analogous to the alleles of the gene. A one-bit gene has two alleles (0 and 1), a two-bit gene has four alleles (00, 01, 10, and 11), and so on. The number of bits in a gene depends on the instruction. An instruction to go left or right requires only a single bit. However, an instruction to go left, right, up, or down requires two bits. When the gene's instructions are followed, there is some performance evaluation that measures the program's reproductive fitness relative to other programs in a computational ecology. Relative fitness determines the probability that each strategy will propagate. Propagation occurs when two mated programs recombine through processes like "crossover" and "inversion." In crossover, the mated programs (or strings) are randomly split and the "left" half of one string is combined with the "right" half of the other, and vice versa, creating two new strings. If two different protocols are each effective, but in different ways, crossover allows them to create an entirely new strategy that may combine the best abilities of each parent, making it superior to either. If so, then the new rule may go on to eventually displace both parent rules in the population of strategies. In addition, the new strings contain random copying errors. These mutations continually refresh the heterogeneity of the population, in the face of selection pressures that tend to reduce it.

To illustrate, consider the eight-bit string **10011010** mated with *11000101*. (The typefaces might represent gender, although the algorithm does not require sexual reproduction.) Each bit could be a specific gene, such as whether to trust a partner under eight different conditions (Macy and Skvoretz 1998). In mating, the two parent strings are randomly broken, say after the third gene. The two offspring would then be **100***00101* and *110***11010**. However, a chance copying error on the last gene might make the second child a mutant, with *110***11011**.

At the end of each generation, each individual's probability of mating is a monotonic (often linear) function of relative performance during that generation, based on stochastic sampling (Goldberg 1989).

$$P_{ij} = \frac{F_i}{\sum_{n=1}^{N} F_n} \text{ for } j = 1 \text{ to } N, \; j \neq i \tag{17.1}$$

where $P_{ij}$ is the probability that $j$ is mated with $i$, $F_i$ is $i$'s "fitness" (or cumulative payoff over all previous rounds in that generation), and $N$ is the size of the population. If the best strategy had only a small performance edge over the worst, it had only a small edge in the race to reproduce. With stochastic sampling, each individual, even the least fit, selects a mate from the fitness-weighted pool of eligibles. In each pairing, the two parents combined their chromosomes to create a single offspring that replaces the less-fit parent. The two chromosomes are combined through crossover.

## 17.5 Learning Models

The need for a cognitive alternative to evolutionary models is reflected in a growing number of formal learning-theoretic models of behavior (Macy 1991; Roth and Erev 1995; Fudenberg and Levine 1998; Young 1998; Cohen et. al. 2001). In general form, learning models consist of a probabilistic decision rule and a learning algorithm in which outcomes are evaluated relative to an aspiration level, and the corresponding decision rules are updated accordingly.

All stochastic learning models share two important principles, the law of effect and probabilistic decision-making (Macy 1989, 1991; Börgers and Sarin 1996; 1997; Roth and Erev 1995; Erev and Roth 1998; Erev et al. 1999; for more references cf. Erev et al. 1999). The law of effect implies that the propensity of an action increases if it is associated with a positively evaluated outcome and it declines if the outcome is negatively evaluated. Probabilistic choice means that actors hold a propensity $qX$ for every action $X$. The probability $pX$ to choose action $X$ then increases in the magnitude of the propensity for $X$ relative to the propensities for the other actions.

Whether an outcome is evaluated as positive or negative depends on the evaluation function. An outcome is positive if it exceeds the actor's aspirations. There are basically three substantively different approaches for modeling the aspiration level, fixed interior aspiration, fixed zero aspiration and moving average aspiration. Fixed interior aspiration assumes that some payoffs are below the aspiration level and are evaluated negatively, while other payoffs are above the aspiration level and are evaluated positively (e.g. Macy 1989, 1991; Fudenberg and Levine 1998). The fixed zero aspiration approach also fixes the aspiration level, but it does so at the minimum possible payoff (Roth and Erev 1995; Börgers and Sarin 1997; Erev and Roth 1998). In other words: in the fixed zero aspiration model every payoff is deemed "good enough" to increase or at least not reduce the corresponding propensity, but higher payoffs increase propensities more than lower payoffs do. Finally, moving average aspiration models assume that the aspiration level approaches the average of the payoffs experienced recently, so that players get used to whatever outcome they may experience often enough (Macy and Flache 2002; Börgers and Sarin 1997; Erev and Rapoport 1998; Erev et al. 1999). Clearly, these assumptions have profound effects on model dynamics.

### 17.5.1 Bush-Mosteller Stochastic Learning Model

One of the simplest models of reinforcement learning is the Bush-Mosteller model (Bush and Mosteller 1950). The Bush-Mosteller stochastic learning algorithm updates probabilities following an action $a$ as follows:

$$p_{a,t+1} = \begin{cases} p_{a,t} + (1 - p_{a,t})\, l\, s_{a,t}, & s_{a,t} \geq 0 \\ p_{a,t} + p_{a,t}\, l\, s_{a,t}, & s_{a,t} < 0 \end{cases}, \tag{17.2}$$

In Eq. 17.2, $p_{a,t}$ is the probability of action $a$ at time $t$ and $s_{a,t}$ is a positive or negative stimulus ($0 \leq |s| \leq 1$). The change in the probability for the action not taken, $b$, obtains from the constraint that probabilities always sum to one, i.e. $p_{b,t+1} = 1 - p_{a,t+1}$. The parameter $l$ is a constant ($0 < l < 1$) that scales the learning rate. With $l \approx 0$, learning is very slow, and with $l \approx 1$, the model approximates a "win-stay, lose-shift" strategy (Catania 1992).

For any value of $l$, Eq. 17.2 implies a decreasing effect of reward as the associated propensity approaches unity, but an increasing effect of punishment. Similarly, as the propensity approaches zero, there is a decreasing effect of punishment and a growing effect of reward. This constrains probabilities to approach asymptotically their natural limits.

### 17.5.2 The Roth-Erev Matching Model

Roth and Erev (Roth and Erev 1995; Erev and Roth 1998; Erev et al. 1999) have proposed a learning-theoretic alternative to the Bush-Mosteller formulation. Their model draws on the "matching law" which holds that adaptive actors will choose between alternatives in a ratio that matches the ratio of reward. Like the Bush-Mosteller model, the Roth-Erev payoff matching model implements the three basic principles that distinguish learning from utility theory – experiential induction (vs. logical deduction), reward and punishment (vs. utility), and melioration (vs. optimization). The similarity in substantive assumptions makes it tempting to assume that the two models are mathematically equivalent, or if not, that they nevertheless give equivalent solutions.

Flache and Macy (2002) show that Bush-Mosteller and Roth-Erev are special cases of a more general learning model and identify important differences between the two specifications. Each specification implements reinforcement learning in different ways, and with different results. Roth and Erev (1995; Erev and Roth 1998) propose a baseline model of reinforcement learning with a fixed zero reference point. The law of effect is implemented such that the propensity for action $X$ is simply the sum of all payoffs a player ever experienced when playing $X$. The probability to choose action $X$ at time $t$ is then the propensity for $X$ divided by the sum of all action propensities at time $t$. The sum of the propensities increases over time, such that payoffs have decreasing effects on choice probabilities. However, this also undermines the law of effect. Suppose, after some time a new action is carried out and yields a higher payoff than every other action experienced before. The probability of repetition of this action will nevertheless be negligible, because its recent payoff is small in comparison with the accumulated payoffs stored in the propensities for the other actions. As a consequence, the baseline model of Roth and Erev (1995) fails to identify particular results, because it has the tendency to lock the learning dynamics into any outcome that occurs sufficiently often early on. Roth and Erev amend this problem by introducing a "forgetting parameter" that keeps propensities low relative to recent payoffs. With this, they increase the sensitivity of the model to recent reinforcement.

Roth and Erev used a variant of this baseline model to estimate globally applicable parameters from data collected across a variety of human subject experiments. They concluded that "low rationality" models of reinforcement learning may often provide a more accurate prediction than forward looking models. Like the Bush-Mosteller, the Roth-Erev model is stochastic, but the probabilities are not equivalent to propensities. The propensity $q$ for action $a$ at time $T$ is the sum of all stimuli $s_a$ a player has ever received when playing $a$:

$$q_{a,T} = \sum_{t=1}^{T} s_{a,t}, \tag{17.3}$$

Roth and Erev then use a "probabilistic choice rule" to translate propensities into probabilities. The probability $p_a$ of action $a$ at time $t+1$ is the propensity for $a$ divided by the sum of the propensities at time $t$:

$$p_{a,t+1} = \frac{q_{a,t}}{q_{a,t} + q_{b,t}}, \qquad a \neq b \tag{17.4}$$

where $a$ and $b$ represent binary choices. Following action $a$, the associated propensity $q_a$ increases if the payoff is positive relative to aspirations (by increasing the numerator in Eq. 17.4) and decreases if negative. The propensity for $b$ remains constant, but the probability of $b$ declines (by increasing the denominator in the equivalent expression for $p_{b,t+1}$).

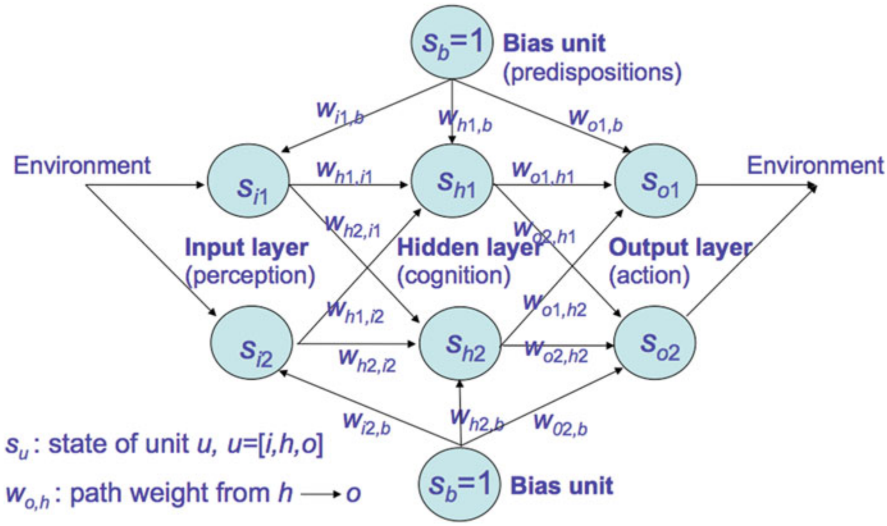



**Fig. 17.1** An example feed-forward network

## *17.5.3 Artificial Neural Networks*

Bush-Mosteller and Roth-Erev are very simple learning models that allow an actor to identify strategies that generally have more satisfactory outcomes. However, the actor cannot learn the conditions in which a strategy is more or less effective. Artificial neural nets add perception to reinforcement, so that actors can learn conditional strategies.

An artificial neural network is a simple type of self-programmable learning device based on parallel distributed processing (Rumelhart and McClelland 1988). Like genetic algorithms, neural nets have a biological analog, in this case, the nerve systems of living organisms. In elementary form, the device consists of a web of neuron-like nodes (or neurodes) that fire when triggered by impulses of sufficient strength, and in turn stimulate other nodes when fired. The magnitude of an impulse depends on the strength of the connection (or "synapses") between the two neurodes. The network learns by modifying these path coefficients, usually in response to environmental feedback about its performance.

There are two broad classes of neural networks that are most relevant to social scientists, feed-forward networks and attractor networks. Feed-forward networks consist of four types of nodes, usually arranged in layers, as illustrated in Fig. 17.1. The most straightforward are input (sensory) and output (response) nodes. Input nodes are triggered by stimuli from the environment. Output nodes, in turn, trigger action by the organism on the environment.

The other two types of nodes are less intuitive. Intermediate (or "hidden") nodes link sensory and response nodes so as to increase the combinations of multiple stimuli that can be differentiated. Figure 17.1 shows a network with a single layer containing two hidden nodes, H1 and H2. The number of hidden layers and the number of hidden nodes in each layer vary with the complexity of the stimulus patterns the network must learn to recognize and the complexity of the responses the network must learn to perform. Unlike sensory and response nodes, hidden nodes have no direct contact with the environment, hence their name.

Bias nodes are a type of hidden node that has no inputs. Instead, a bias node continuously fires, creating a predisposition toward excitation or inhibition in the nodes it stimulates, depending on the size and sign of the weights on the pathway from the bias node to the nodes it influences. If the path weight is positive, the bias is toward excitation, and if the path is negative, the bias is toward inhibition. The weighted paths from the bias node to other hidden and output nodes correspond to the activation thresholds for these nodes.

A feed-forward network involves two processes – action (firing of the output nodes) and learning (error correction). The action phase consists of the forward propagation of influence (either excitation or inhibition) from the input nodes to the hidden nodes to the output nodes. The influence of a node depends on the state of the node and the weight of the neural pathway to a node further forward in the network. The learning phase consists of the backward propagation of error from the output nodes to the hidden nodes to the input nodes, followed by the adjustment of the weights so as to reduce the error in the output nodes.

The action phase begins with the input nodes. Each node has a "state" which can be binary (e.g. 0 or 1 to indicate whether the node "fires") or continuous (e.g. .9 to indicate that the node did not fire as strongly as one whose state is 1.0). The states of the input nodes are controlled entirely by the environment and correspond to a pattern that the network perceives. The input nodes can influence the output nodes directly as well as via hidden nodes that are connected to the input nodes.

To illustrate, consider a neural network in which the input nodes are numbered from $i = 1$ to $I$. The $i$th input is selected and the input register to all nodes $j$ influenced by $i$ are then updated by multiplying the state of $i$ times the weight $w_{ij}$ on the $ij$ path. This updating is repeated for each input node. The pathways that link the nodes are weighted with values that determine the strength of the signals moving along the path. A low absolute value means that an input has little influence on the output. A large positive weight makes the input operate as an excitor. When it fires, the input excites an otherwise inhibited output. A large negative path weight makes the input operate as an inhibitor. When it fires, the input inhibits an otherwise excited output.

Next, the nodes whose input registers have been fully updated (e.g. the nodes in the first layer of hidden nodes) must update their states. The state of a node is updated by aggregating the values in the node's input register, including the input from its bias node (which always fires, e.g. $B_i = 1$).

Updating states is based on the activation function. Three activation functions are commonly used. Hard-limit functions fire the node if the aggregate input exceeds zero. Sigmoid stochastic functions fire the node with a probability given by the aggregate input. Sigmoid deterministic functions fire the node with a magnitude given by the aggregate input.[2] For example, if node $k$ is influenced by two input nodes, $i$ and $j$ and a bias node $b$, then $k$ sums the influence $ik = i{*}w_{ij} + k{*}w_{ik} + b{*}w_{ib}$. Positive weights cause $i$, $j$, and $b$ to activate $k$ and negative weights inhibit. If $j$ is hard-limited, then if $ik > 0$, $k = 1$, else $k = 0$. If the activation function is stochastic, $k$ is activated with a probability $p = 1/1 + e^{-ik}$. If the sigmoid function is deterministic, $k$ is activated with magnitude $p$.

Once the states have been activated for all nodes whose input registers have been fully updated (e.g., the first layer of hidden nodes and/or one or more output nodes that are directly connected to an input node), these nodes in turn update the input registers of nodes they influence going forward (e.g. the second layer of hidden nodes and/or some or all of the output nodes). Once this updating is complete, all nodes whose input registers have been fully updated aggregate across their inputs and update their states, and so on until the states of the output nodes have been updated. This completes the action phase.

The network learns by modifying the path weights linking the neurodes. Learning only occurs when the response to a given sensory input pattern is unsatisfactory. The paths are then adjusted so as to reduce the probability of repeating the mistake the next time this pattern is encountered. In many applications, neural nets are trained to recognize certain patterns or combinations of inputs. For example, suppose we want to train a neural net to predict stock prices from a set of market indicators. We first train the net to correctly predict known prices. The net begins with random path coefficients. These generate a prediction. The error is then used to adjust the weights in an iterative process that improves the predictions. These weights are analogous to those in a linear regression, and like a regression, the weights can then be applied to new data to predict the unknown.

#### 17.5.3.1 Back Propagation of Error

A feed-forward neural network is trained by adjusting the weights on the paths between the nodes. These weights correspond to the influence that a node will have in causing nodes further forward in the network to fire. When those nodes fire incorrectly, the weights responsible for the error must be adjusted accordingly. Just as the influence process propagates forward, from the input nodes to the hidden layer to the output nodes, the attribution of responsibility for errors propagates backward, from the output nodes to the hidden layers.

The back propagation of error begins with the output nodes and works back through the network to the input nodes, in the opposite direction from the influence process that "feeds forward" from input nodes to hidden nodes to output nodes. First, the output error is calculated for each output node. This is simply the

[2] A multilayer neural net requires a non-linear activation function (such as a sigmoid). If the functions are linear, the multilayer net reduces to a single-layer I-O network.

difference between the expected state of the $i$th output node ($\hat{S}_i$) and the state that was observed ($S_i$). For an output node, error refers to the difference between the expected output for a given pattern of inputs and the observed output.

$$e_o = s_o(1 - s_o)(\hat{s}_o - s_o) \tag{17.5}$$

where the term $s_o(1 - s_o)$ limits the error. If the initial weights were randomly assigned, it is unlikely that the output will be correct. For example, if we observe an output value of 0.37 and we expected 1, then the error is $-0.53$.

Once the error has been updated for each output node, these errors are used to update the error for the nodes that influenced the output nodes. These nodes are usually located in the last layer of hidden nodes but they can be anywhere and can even include input nodes that are wired directly to an output.[3] Then the errors of the nodes in the last hidden layer are used to update error back further still to the hidden nodes that influenced the last layer of hidden nodes, and so on, back to the first layer of hidden nodes, until the errors for all hidden nodes have been updated. Input nodes cannot have error, since they simply represent an exogenous pattern that the network is asked to learn.

Back propagation means that the error observed in an output node $o$ $(\hat{s}_o - s_o)$ is allocated not only to $o$ but to all the hidden nodes that influenced $o$, based on the strength of their influence. The total error of a hidden node $h$ is then simply the summation over all $n_h$ allocated errors from the $n$ nodes $i$ that $h$ influenced, including output nodes as well as other hidden nodes:

$$e_h = s_h(1 - s_h)\sum_{i=1}^{n} w_{hi}e_i \tag{17.6}$$

Once the errors for all hidden, bias, and output nodes have been back propagated, the weight on the path from $i$ to $j$ is updated:

$$w'_{ij} = w_{ij} + \lambda s_i e_j \tag{17.7}$$

where λ is a fractional learning rate. The learning algorithm means that the influence of $i$ on $j$ increases if $j$'s error was positive (i.e. the expected output exceeded the observed) and decreases if $j$'s influence was negative.

Note that the Bush-Mosteller model is equivalent to a neural net with only a single bias unit and an output, but with no sensory inputs or hidden units. Such a device is capable of learning *how often* to act, but not *when* to act, that is, it is incapable of learning conditional strategies. In contrast, a feed-forward network can

[3] However, if an input node is wired to hidden nodes as well as output nodes, the error for this node cannot be updated until the errors for all hidden nodes that it influenced have been updated.

learn to differentiate environmental cues and respond using more sophisticated protocols for contingent strategies.

#### 17.5.3.2 Attractor Neural Network

Feed-forward networks are the most widely used but not the only type of artificial neural network. An alternative design is the attractor neural network (Churchland and Sejnowski 1994; Quinlan 1991), originally developed and investigated by Hopfield (1982; Hopfield and Tank 1985). In a recent article, Nowak and Vallacher (1998) note the potential of these computational networks for modeling group dynamics. This approach promises to provide a fertile expansion to social network analysis, which has often assumed that social ties are binary and static. A neural network provides a way to dynamically model a social network in which learning occurs at both the individual and structural levels, as relations evolve in response to the behaviors they constrain.

Unlike feed-forward neural networks (Rumelhart and McClelland 1988), which are organized into hierarchies of input, hidden, and output nodes, attractor (or “feed-lateral”) networks are internally undifferentiated. Nodes differ only in their states and in their relational alignments, but they are functionally identical. Without input units to receive directed feedback from the environment, these models are “unsupervised” and thus have no centralized mechanism to coordinate learning of efficient solutions. In the absence of formal training, each node operates using a set of behavioral rules or functions that compute changes of state (“decisions”) in light of available information. Zeggelink (1994) calls these “object-oriented models”, where each agent receives input from other agents and may transform these inputs into a change of state, which in turn serves as input for other agents.

An important agent-level rule that characterizes attractor networks is that individual nodes seek to minimize “energy” (also “stress” or “dissonance”) across all relations with other nodes. As with feed-forward networks, this adaptation occurs in two discrete stages. In the action phase, nodes change their states to maximize similarity with nodes to which they are strongly connected. In the learning phase, they update their weights to strengthen ties to similar nodes. Thus, beginning with some (perhaps random) configuration, the network proceeds to search over an optimization landscape as nodes repeatedly cycle through these changes of weights and states.

In addition to variation in path strength, artificial neural networks typically have paths that inhibit as well as excite. That is, nodes may be connected with negative as well as positive weights. In a social network application, agents connected by negative ties might correspond to “negative referents” (Schwartz and Ames 1977), who provoke differentiation rather than imitation.

Ultimately, these systems are able to locate stable configurations (called “attractors”), for which any change of state or weight would result in a net increase in stress for the affected nodes. Hopfield (1982) compares these attractors to memories, and shows that these systems of undifferentiated nodes can learn to implement higher-order cognitive functions. However, although the system may converge at a stable equilibrium that allows all nodes to be locally satisfied (i.e., a "local optimum"), this does not guarantee that the converged pattern will minimize overall dissonance (a "global optimum").

This class of models generally uses complete networks, with each node characterized by one or more binary or continuous states and linked to other nodes through endogenous weights. Like other neural networks, attractor networks learn stable configurations by iteratively adjusting the weights between individual nodes, without any global coordination. In this case, the weights change over time through a Hebbian learning rule: the weight $w_{ij}$ is a function of the correlation of states for nodes $i$ and $j$ over time. Specifically, Hebbian learning implies the following rules:

- To the extent that nodes $i$ and $j$ adopt the same states at the same time, the weight of their common tie will increase until it approaches some upper limit (e.g., 1.0).
- To the extent that nodes $i$ and $j$ simultaneously adopt different states, the weight of their common tie will decrease until it approaches some lower limit (e.g., 0.0).

Although Hebbian learning was developed to study memory in cognitive systems, it corresponds to the homophily principle in social psychology (Homans 1951) and social network theory (McPherson and Smith-Lovin 1987), which holds that agents tend to be attracted to those whom they more closely resemble. This hypothesis is also consistent with structural balance theory (Cartwright and Harary 1956; Heider 1958) and has been widely supported in studies of interpersonal attraction and interaction, where it has been called "the Law of Attraction" (Byrne 1971; Byrne and Griffit 1966).

An important property of attractor networks is that individual nodes seek to minimize "energy" (or dissonance) across all relations with other nodes – a process that parallels but differs from the pursuit of balanced relations in structural balance theory. These networks also posit self-reinforcing dynamics of attraction and influence as well as repulsion and differentiation.

Following Nowak and Vallacher (1998), Macy et al. (2003) apply the Hopfield model of dynamic attraction to the study of polarization in social networks. In this application, observed similarity/difference between states determines the strength and valence of the tie to a given referent. This attraction and repulsion may be described anecdotally in terms of liking, respect, or credibility and their opposites.

In their application of the Hopfield model, each node has $N - 1$ undirected ties to other nodes. These ties include weights, which determine the strength and valence of influence between agents. Formally, social pressure on agent $i$ to adopt a binary state $s$ (where $s = \pm 1$) is the sum of the states of all other agents $j$, where influence from each agent is conditioned by the weight ($w_{ij}$) of the dyadic tie between $i$ and $j$ ($-1.0 < w_{ij} < 1.0$):

$$P_{is} = \frac{\sum_{j=1}^{N} w_{ij} s_j}{N - 1}, \; j \neq i \tag{17.8}$$

Thus, social pressure ($-1 < P_{is} < 1$) to adopt $s$ becomes increasingly positive as $i$'s "friends" adopt $s$ ($s = 1$) and $i$'s "enemies" reject $s$ ($s = -1$). The pressure can also become negative in the opposite circumstances. The model extends to multiple states in a straightforward way, where Eq. 17.8 independently determines the pressure on agent $i$ for each binary state $s$.

Strong positive or negative social pressure does not guarantee that an agent will accommodate, however. It is effective only if $i$ is willing and able to respond to peer influence. If $i$ is closed-minded or if a given trait is not under $i$'s control (e.g., ethnicity or gender), then no change to $s$ will occur. The probability $\pi$ that agent $i$ will change state $s$ is a cumulative logistic function of social pressure:

$$\pi_{is} = \frac{1}{1 + e^{-10P_{is}}} \tag{17.9}$$

Agent $i$ adopts $s$ if $\pi > C \pm \varepsilon$, where $C$ is the inflection point of the sigmoid, and $\varepsilon$ is an exogenous error parameter ($0 \leq \varepsilon \leq 1$). At one extreme, $\varepsilon = 0$ produces highly deterministic behavior, such that any social pressure above the trigger value always leads to conformity and pressures below the trigger value entail differentiation. Following Harsanyi (1973), $\varepsilon > 0$ allows for a "smoothed best reply" in which pressure levels near the trigger point leave the agent relatively indifferent and thus likely to explore behaviors on either side of the threshold.

In the Hopfield model, the path weight $w_{ij}$ changes as a function of similarity in the states of node $i$ and $j$. Weights begin with uniformly distributed random values, subject to the constraints that weights are symmetric ($w_{ij} = w_{ji}$). Across a vector of $K$ distinct states $s_{ik}$ (or the position of agent $i$ on issue $k$), agent $i$ compares its own states to the observed states of another agent $j$ and adjusts the weight upward or downward corresponding to their aggregated level of agreement or disagreement. Based on the correspondence of states for agents $i$ and $j$, their weight will change at each discrete time point $t$ in proportion to a parameter $\lambda$, which defines the rate of structural learning ($0 < \lambda < 1$):

$$w_{ij,t+1} = w_{ijt}(1 - \lambda) + \frac{\lambda}{K} \sum_{k=1}^{K} s_{jkt} s_{ikt}, j \neq i \tag{17.10}$$

As correspondence of states can be positive (agreement) or negative (disagreement), ties can grow positive or negative over time, with weights between any two agents always symmetric.

Note one significant departure from structural balance theory. Although the agents in this model are clearly designed to maintain balance in their behaviors with both positive and negative referents, this assumption is not "wired in" to the relations themselves. That is, two agents $i$ and $j$ feel no direct need for consistency in their relations with a third agent $h$. Indeed, $i$ has no knowledge of the $jh$ relationship and thus no ability to adjust the $ij$ relation so as to balance the triad.

Given an initially random configuration of states and weights, these agents will search for a profile that minimizes dissonance across their relations. Structural balance theory predicts that system-level stability can only occur when the group either has become uniform or has polarized into two (Cartwright and Harary 1956) or more (Davis 1967) internally cohesive and mutually antipathetic cliques. However, there is no guarantee in this model that they will achieve a globally optimal state in structural balance.

### *17.5.4 Belief Learning*

Actors learn not only what is useful for obtaining rewards and avoiding punishments, they also update their beliefs about what is true and what is false. There are two main models of belief learning in the literature, Bayesian belief learning and fictitious play[4] (cf. Offerman 1997; Fudenberg and Levine 1998). These models differ in their assumptions about how players learn from observations. Both models assume that players believe that something is true with some fixed unknown probability $p$. In Bayesian learning, players then use Bayes' learning rule to rationally update over time their beliefs about $p$. In a nutshell, Bayes' learning rule implies that actors' assessment of the true value of $p$ converges in the long run on the relative frequency of events that they observe. However, in the short and medium term, Bayesian learners remain suspicious in the sense that they take into account that observed events are an imperfect indication of $p$ (Offerman 1997).

Fudenberg and Levine note that fictitious play is a special case of Bayesian learning. Fictitious play is a form of Bayesian learning that always puts full weight on the belief that the true value of $p$ corresponds to the relative frequency observed in past events. Fudenberg and Levine (1998) note that it is an implausible property of fictitious play that a slight change in beliefs may radically alter behavior. The reason is that the best reply function always is a step function. As a remedy, Fudenberg and Levine introduce a smooth reply curve. The reply curve assigns a probability distribution that corresponds to the relative frequency of events. With strict best reply, the reply curve is a step function. Instead, a smooth reply curve assigns some probability to the action that is not strict best reply. This probability decreases in the difference between the expected payoffs. Specifically, when expected payoffs are equal, actors choose with equal probability, whereas their choice probabilities converge on pure strategies when the difference in expected payoffs approaches the maximum value.

[4] The Cournot rule may be considered as a third degenerate model of belief learning. According to the Cournot rule, players assume that the behavior of the opponent in the previous round will always occur again in the present round.

The strict best reply function corresponds to the rule, "play X if the expected payoff for X is better than the expected payoff for Y, given your belief $p$. Otherwise play Y". Smooth best reply is then introduced with the modification to play the strict best reply strategy only with a probability of $1 - \eta$, whereas the alternative is played with probability $\eta$. The probability $\eta$, in turn, decreases in the absolute difference between expected payoffs where $\eta = 0.5$ if players are indifferent.

Belief learning generally converges with the predictions of evolutionary selection models. The approaches are also similar in the predicted effects of initial conditions on end results. Broadly, the initial distribution of strategies in independent populations in evolutionary selection corresponds to the initial beliefs players' hold about their opponent. For example, when two pure strategy Nash equilibria are feasible, then the one tends to be selected towards which the initial strategy distribution in evolutionary selection or initial beliefs in belief learning are biased.

## 17.6 Conclusion

Evolution and learning are examples of backward-looking consequentialist models, in which outcomes of agents' past actions influence their future choices, either through selection (in the case of evolution) or reinforcement (in the case of learning). Backward-looking models make weaker assumptions about agents' cognitive capacities than forward looking models, and thus may be appropriate for settings in which agents lack the ability, resources, information, and motivation to engage in intensive cognitive processing, as in most everyday instances of collective action. Backward-looking models may also be useful for understanding behavior driven by affect, rather than calculation. Forward looking models may be more appropriate in applications such as investment decisions, international diplomacy, or military strategy, where the stakes are high enough to warrant collection of all relevant information and the actors are highly skilled strategists. However, even where the cognitive assumptions of the models are plausible, forward-looking models are generally limited to the identification of static equilibria but not necessarily whether and how agents will reach those equilibria.

When implemented computationally, backward-looking models can show how likely agents are to reach particular equilibria, as well as the paths by which those equilibria may be reached. However, computational models are also less general than analytical models. Furthermore, backward-looking models will be of little help if change in the environment outpaces the rate of adaptation. These limitations underscore the importance of robustness testing over a range of parameter values.

Evolutionary models are most appropriate for theoretical questions in which adaptation takes place at the population level, through processes of selection. Biological evolution is the most obvious analog, but social and cultural evolution are likely to be more important for social scientists. However, as we note above, researchers must be cautious about drawing analogies from the biological to social/ cultural dynamics.

Learning models based on reinforcement and Bayesian updating are useful in applications that do not require conditional strategies based on pattern recognition. When agents must learn more complex conditional strategies, feed-forward neural networks may be employed. Furthermore, attractor neural networks are useful for modeling structural influence, such as conformity pressures from peers.

Models of evolution and learning are powerful tools for modeling social processes. Both show how complex patterns can arise when agents rely on relatively simple, experience-driven decision rules. This chapter seeks to provide researchers with an overview of promising research in this area, and the tools necessary to further develop this research.

## Further Reading

We refer readers interested in particular learning models and their application in agent-based simulation to (Macy and Flache 2002), which gives a brief introduction into principles of reinforcement learning and discusses by means of simulation models how reinforcement learning affects behavior in social dilemma situations, whereas (Macy 1996) compares two different approaches of modeling learning behavior by means of computer simulations. (Fudenberg and Levine 1998) gives a very good overview on how various learning rules relate to game theoretic rationality and equilibrium concepts.

For some wider background reading, we recommend (Macy 2004), which introduces the basic principles of learning theory applied to social behavior, (Holland et al. 1986), which presents a framework in terms of rule-based mental models for understanding inductive reasoning and learning, and Sun (2008), which is a handbook of computational cognitive modeling.

## References

Allison P (1992) The cultural evolution of beneficent norms. Soc Forces 71:279–301

Bandura A (1977) Social learning theory. Prentice Hall, Englewood Cliffs

Börgers T, Sarin R (1996) Naive reinforcement learning with endogenous aspirations. International Economic Review 4:921–950

Börgers T, Sarin R (1997) Learning through reinforcement and replicator dynamics. J Econ Theory 77:1–14

Boyd R, Richerson PJ (1985) Culture and the evolutionary process. University of Chicago Press, Chicago

Bush RR, Mosteller F (1950) Stochastic models for learning. Wiley, New York

Byrne DE (1971) The attraction paradigm. Academic, New York

Byrne DE, Griffith D (1966) A development investigation of the law of attraction. J Pers Soc Psychol 4:699–702

Cartwright D, Harary F (1956) Structural balance: a generalization of Heider's theory. Psychol Rev 63:277–293
Catania AC (1992) Learning, 3rd edn. Prentice Hall, Englewood Cliffs
Churchland PS, Sejnowski TJ (1994) The computational brain. MIT Press, Cambridge, MA
Cohen MD, Riolo R, Axelrod R (2001) The role of social structure in the maintenance of cooperative regimes. Ration Soc 13(1):5–32
Davis JA (1967) Clustering and structural balance in graphs. Hum Relat 20:181–187
Dawkins R (1976) The selfish gene. Oxford University Press, Oxford
Durham WH (1992) Coevolution: genes, culture and human diversity. Stanford University Press, Stanford
Erev I, Rapoport A (1998) Coordination, "magic", and reinforcement learning in a market entry game. Game Econ Behav 23:146–175
Erev I, Roth AE (1998) Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am Econ Rev 88(4):848–879
Erev I, Bereby-Meyer Y, Roth AE (1999) The effect of adding a constant to all payoffs: experimental investigation, and implications for reinforcement learning models. J Econ Behav Organ 39(1):111–128
Flache A, Macy MW (2002) Stochastic collusion and the power Law of learning: a general reinforcement learning model of cooperation. J Confl Resolut 46(5):629–653
Frank R (1988) Passions within reason: the strategic role of the emotions. Norton, New York
Fudenberg D, Levine D (1998) The theory of learning in games. MIT Press, Boston
Goldberg D (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, New York
Hamilton W (1964) The genetic evolution of social behaviour. J Theor Biol 17:1–54
Harsanyi J (1973) Games with randomly disturbed payoffs. Int J Game Theory 2:1–23
Heider F (1958) The psychology of interpersonal relations. Wiley, New York
Herrnstein RJ, Drazin P (1991) Meliorization: a theory of distributed choice. J Econ Perspect 5 (3):137–156
Holland J (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor
Holland JH, Holyoak KJ, Nisbett RE, Thagard PR (1986) Induction: processes of inference, learning, and discovery. MIT Press, Cambridge, MA
Homans GC (1951) The human group. Harcourt Brace, New York
Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Am Acad Sci 79:2554–2558
Hopfield JJ, Tank DW (1985) "Neural" computation of decisions in optimization problems. Bio Cybernetics 52:141–152
James W (1981) Principles of psychology. Harvard University Press, Cambridge, MA
Lopreato J (1990) From social evolutionism to biocultural evolutionism. Sociol Forum 5:187–212
Macy M (1989) Walking Out of social traps: a stochastic learning model for the Prisoner's dilemma. Ration Soc 2:197–219
Macy M (1991) Learning to cooperate: stochastic and tacit collusion in social exchange. Am J Sociol 97:808–843
Macy M (1996) Natural selection and social learning in prisoner's dilemma: co-adaptation with genetic algorithms and artificial neural networks. Sociol Method Res 25:103–137
Macy M (1997) Identity, interest, and emergent rationality. Ration Soc 9:427–448
Macy M (1998) Social order in an artificial world. J Artif Soc Soc Simul 1(1). http://jasss.soc.surrey.ac.uk/1/1/4.html
Macy M (2004) Learning theory. In: Ritzer G (ed) Encyclopedia of social theory. Sage, Thousand Oaks
Macy M, Flache A (2002) Learning dynamics in social dilemmas. Proc Natl Acad Sci 99:7229–7236

Macy M, Skvoretz J (1998) The evolution of trust and cooperation between strangers: a computational model. Am Sociol Rev 63:638–660

Macy M, Kitts J, Flache A, Benard S (2003) Polarization in dynamic networks: a Hopfield model of emergent structure. In: Breiger R, Carley K (eds) Dynamic social network modeling and analysis: workshop summary and papers. National Academy Press, Washington, DC, pp 162–173

March JG, Simon HA (1958) Organizations. Wiley, New York

McPherson JM, Smith-Lovin L (1987) Homophily in voluntary organizations: status distance and the composition of face to face groups. Am Sociol Rev 52:370–379

Nowak A, Vallacher RR (1998) Toward computational social psychology: cellular automata and neural network models of interpersonal dynamics. In: Read SJ, Miller LC (eds) Connectionist models of social reasoning and social behavior. Lawrence Erlbaum, Mahwah, pp 277–311

Offerman T (1997) Beliefs and decision rules in public good games. Kluwer, Dordrecht

Quinlan PT (1991) Connectionism and psychology: a psychological perspective on New connectionist research. University of Chicago Press, Chicago

Roth AE, Erev I (1995) Learning in extensive-form games: experimental data and simple dynamic models in intermediate term. Game Econ Behav 8:164–212

Rumelhart DE, McClelland JL (1988) Parallel distributed processing: explorations in the microstructure of cognition. MIT Press, Cambridge, MA

Schwartz SH, Ames RE (1977) Positive and negative referent others as sources of influence: a case of helping. Sociometry 40:12–21

Sun R (ed) (2008) The Cambridge handbook of computational psychology. Cambridge University Press, Cambridge

Taylor PD, Jonker L (1978) Evolutionary stable strategies and game dynamics. Math Biosci 40:145–156

Thorndike EL (1898) Animal intelligence: an experimental study of the associative processes in animals, vol 8, Psychological review, monograph supplements. Macmillan, New York

Vanberg V (1994) Rules and choice in economics. Routledge, London

Williams GC (1966) Adaptation and natural selection. Princeton University Press, Princeton

Wilson EO (1988) On human nature. Harvard University Press, Cambridge, MA

Wynne-Edwards VC (1962) Animal dispersion in relation to social behaviour. Oliver & Boyd, Edinburgh

Wynne-Edwards VC (1986) Evolution through group selection. Blackwell, Oxford

Young HP (1998) Individual strategy and social structure. An evolutionary theory of institutions. Princeton University Press, Princeton

Zeggelink E (1994) Dynamics of structure: an individual oriented approach. Soc Network 16:295–333

# Chapter 18
# Evolutionary Mechanisms

**Edmund Chattoe-Brown and Bruce Edmonds**

**Why Read This Chapter?** To learn about techniques that may be useful in designing simulations of adaptive systems including Genetic Algorithms (GA), Classifier Systems (CS) and Genetic Programming (GP). The chapter will also tell you about simulations that have a fundamentally evolutionary structure – those with variation, selection and replications of entities – showing how this might be made relevant to social science problems.

**Abstract** After an introduction, the abstract idea of evolution is analysed into four processes which are illustrated with respect to a simple evolutionary game. A brief history of evolutionary ideas in the social sciences is given, illustrating the different ways in which the idea of evolution has been used. The technique of GA is then described and discussed including: the representation of the problem and the composition of the initial population, the fitness function, the reproduction process, the genetic operators, issues of convergence, and some generalisations of the approach including endogenising the evolution. GP and CS are also briefly introduced as potential developments of GA. Four detailed examples of social science applications of evolutionary techniques are then presented: the use of GA in the Arifovic "cobweb" model, using CS in a model of price setting developed by Moss, the role of GP in understanding decision making processes in a stock market model and relating evolutionary ideas to social science in a model of survival for "strict" churches. The chapter concludes with a discussion of the prospects and difficulties of using the idea of biological evolution in the social sciences.

E. Chattoe-Brown (✉)
Department of Sociology, University of Leicester, Leicester, UK
e-mail: ecb18@le.ac.uk

B. Edmonds
Centre for Policy Modelling, Manchester Metropolitan University, Manchester, UK
e-mail: bruce@edmonds.name

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_18,

## 18.1 Introduction

There are now many simulations of complex social phenomena that have structures or component processes analogous to biological evolution (see Arifovic 1994; Chattoe 2006a; Dosi et al. 1999; Lomborg 1996; Nelson and Winter 1982; Oliphant 1996; Windrum and Birchenhall 1998 to get a flavour of the diversity of approach and applications). Clearly the process of biological evolution is complex and has resulted in the development of complex (and in several cases social) systems. However, biological evolution follows very specific mechanisms and is clearly not strictly isomorphic with social processes. For a start, biological evolution occurs over larger time spans than most social processes. Further, it is unlikely, as sociobiology (Wilson 1975) and evolutionary psychology (Buss 1998) are sometimes supposed to imply, that the domain of social behaviour will actually prove reducible to genetics. Thus it is not immediately apparent *why* evolutionary ideas have had such an influence upon the modelling of social processes. Nevertheless, simulations of social phenomena have been strongly influenced by our understanding of biological evolution and this has occurred via two main routes: through *analogies* with biological evolution and through computer science approaches.

In the first case, conceptions of evolution have been used as a way of understanding social processes and then simulations have been made using these conceptions. For example Nelson and Winter (1982) modelled growth and change in firms using the idea of random variation (new products or production processes) and selective retention (whether these novelties in fact sustain profitability – the survival requirement for firms – in an environment defined by what other firms are currently doing).

In the second case, computer science has taken up the ideas of evolution and applied it to engineering problems. Most importantly in Machine Learning, ideas from biological evolution have inspired whole families of techniques in what has become known as "Evolutionary Computation". The most famous of these techniques are Genetic Algorithms (Holland 1975; Mitchell 1996) and Genetic Programming (Koza 1992a, 1994) discussed below. These algorithms have then been applied in social simulations with different degrees of modification; from using them unchanged as "off the shelf" plug-ins (for example to model learning processes) to specifying simulation processes that use the core evolutionary idea but are completely re-engineered for a particular modelling purpose or domain. There is no a priori reason to suppose that a particular technique from computer science will be the most appropriate algorithm in a social simulation (including those with a biological inspiration) as we shall see below, but it certainly presents a wealth of evolutionary ideas and results that are potentially applicable in some form. Like any theory, the trick is to use good judgement and a clear specification in applying an algorithm to a particular social domain (Chattoe 1998, 2006b).

What is certainly the case is that biological evolution offers an example of how complex and self-organised phenomena can emerge using randomness, so it is natural to look to this as a possible conceptual framework with which to understand social phenomena with similar properties. (In particular, while it may be reasonable to assume deliberation and rationality in some social contexts, it is extremely unlikely to apply to all social structures and phenomena. As such, some kind of blind variation and retention – resulting in evolution – is probably the only well defined theoretical alternative.) The extent to which evolution-like processes are generally applicable to social phenomena is unknown (largely because this foundational issue has not received much attention to date), but these processes certainly are a rich source of ideas and it may be that there are some aspects of social complexity that will prove to be explicable by models thus inspired. It is already the case that many social simulation models have taken this path and thus have the potential to play a part in helping us to understand social complexity (even if they only serve as horrible examples).

This chapter looks at some of the most widely used approaches to this kind of modelling, discusses others, gives examples and critically discusses the field along with areas of potential development.

## 18.2 An Abstract Description of Biological Evolution

We will not provide full details of biological evolution as currently understood in the Neo-Darwinian synthesis.[1] Rather we will take from this a generalised model of evolution that will potentially cover a variety of social processes. This description will then be used to discuss an example from Evolutionary Game Theory (Vega-Redondo 1996). This will unpack the implications of the abstract description and demonstrate its generality. This generalisation is a preliminary to discussing evolutionary simulations of social phenomena based on the abstract description as a framework.

### *18.2.1 The Four Process Description*

The basic components in the biological theory are the genotype (the set of instructions or genome) and the phenotype (the "body" which the genotype specifies) in which these instructions are embedded. The phenotype is constructed using "instructions" encoded in the genotype. The phenotype has various capabilities including reproduction. Maintenance of the phenotype (and the embedded genotype) requires a number of potentially scarce inputs (food, water). The phenotypic capabilities include management of inputs and outputs to the organism. Poor adaptation of these capabilities with respect to either external or internal environment will

[1] For details about this see any good textbook on biology (e.g. Dobzhansky et al. 1977).

result in malfunction and consequent death. The death of a particular phenotype also ends its reproductive activity and removes the corresponding genotype from the population. Variation occurs by mutation during reproduction, giving rise to novel genotypes (and hence subsequent phenotypes) in the resulting offspring. Genotypic variations are not selected directly by the environment but according to the overall capabilities of their phenotypes. In biology, phenotype alterations cannot be transmitted to the genotype for physiological reasons but in social systems this "Lamarckian" adjunct to evolution (which is not, however, adequate to explain change in its own right) is both possible and plausible. In particular it allows for combinations of evolutionary learning at the social level and deliberate action at the individual level (Chattoe 2006a).

A full specification of an evolutionary model requires descriptions of the following processes:

1. *Generation of phenotypes*: A specification of the genotypes and the phenotypes these correspond to. This may not specify a 1–1 mapping between genotypes and phenotypes but describe the process by which phenotypes are actually constructed from genotypes. This is necessary when genotypes cannot be enumerated.
2. *Capabilities of the phenotypes*: A specification of ways in which phenotypes may use their capabilities to affect the internal and external environment, including the behaviour and numbers of other phenotypes. Lamarckian systems include the capability to modify the genotype using environmental feedback during the lifetime of the phenotype.
3. *Mechanisms of reproduction and variation*: A specification of the process by which phenotypes reproduce including possible differences between ancestor and successor genotypes resulting from reproduction. Reproduction may involve a single ancestor genotype (parthenogenesis) or a pair (sexual reproduction). In principle, multiple parents could be modelled if appropriate for particular social domains (like policies decided by committees) though this approach has not been used so far.
4. *Mechanism of selection*: A specification of all the processes impinging on the phenotype and their effects. This is the converse of the second mechanism, the capabilities of one phenotype form part of the selection process for the others. Some processes, such as fighting to the death, can be seen as directly selective. However, even indirect processes like global warming may interact with phenotypic capabilities in ways that affect fitness.

In these process specifications it may be convenient to distinguish (and model separately) the "environment" as the subset of objects impinging on phenotypes which display no processes of the first three types. Whether a separate representation of the environment is useful depends on the process being modelled. At one extreme, a person in a desert is almost exclusively dealing with the environment. At the other, rats in an otherwise empty cage interact almost entirely with each other.

Obviously, some of these specifications could be extremely complex depending on the system being modelled. The division into system components is necessarily imprecise but not arbitrary. It is based on the considerable observed integrity of organisms relative to their environment. (This integrity is also observed in social "organisms" like firms which have clearly – and often legally – defined boundaries.) The first and third specifications involve processes internal to the organism while the second and fourth represent the organism's effect on the external world and the converse.

Of course, social processes, even "evolutionary social processes" are not limited by the above specification. For example what most closely corresponds to the genotype might not be separable from what corresponds to the phenotype. Nevertheless, however, for a very broad class of evolutionary simulations it will be necessary to implement something very similar to the above four categories.

### *18.2.2 Illustrative Example: A Simple Evolutionary Game*

Despite the potential complexity of specifying complete models for biological systems, this description can also be used to clarify and analyse relatively simple evolutionary systems. In this section we shall provide a description for an evolutionary game. The purpose is not to comment on Evolutionary Game Theory per se but to show how the description raises issues relevant to our understanding of evolutionary models.

For each agent, the genotype is one of a set of finite state automata producing a single action in each period, for example the complete set of one and two-state automata leading to the actions "Co-operate" (C) and "Defect" (D) in a Prisoner's Dilemma (see, for example, Lomborg 1996). The action is compared with the action of a co-player (another agent) and the result is an adjustment to the "energy level" for each agent depending on the game payoffs and chosen strategies. If agents reach a certain energy level, they produce an exact copy. (This model dispenses with variation and involves asexual reproduction.) If the energy level of any agent reaches zero, it dies and is removed from the environment. Reproduction reduces the energy level considerably. Merely existing also does so but at a much lower rate.

With some qualifications, this is an example of a complete description discussed in the last section. It reveals some interesting things about the process of constructing such descriptions.

Firstly, this model involves a very attenuated environment compared to real social systems. Agents have a single external capability involving one of two actions and thus affecting the energy levels of their co-players. The effect of these actions is also the only environmental process that impinges on agents. The model of the environment just consists of mechanisms for deciding when and which agents will play, administering actions and energy changes, producing copies and removing dead agents. In real social systems, exogenous events (both social and environmental) are likely to be very important.

Secondly, the discussion of energy levels still sounds biological but this is simply to make the interpretation of the example more straightforward in the light of the recent discussion. As we shall show subsequently, the survival criterion can just as easily be profitability or organisation membership levels.

Thirdly (and perhaps most importantly) there has been sleight of hand in the description of the model. We have already described Lamarckism (genotype modification by the phenotype during the organism's lifetime) and the construction of the phenotype by the genotype during gestation is a fundamental part of the evolutionary process. But in this example the genotype is effectively "reconstructing" the phenotype every time the finite state automaton generates an action. There is nothing observable about a particular agent, given the description above, except the sequence of actions they choose. There is no way for an agent to establish that another is actually the same when it plays D on one occasion and C on another, or that two plays of D in successive periods actually come from two different agents. In fact, this is the point at which models of social evolution develop intuitively from the simple description of biological evolution used so far. The capabilities of social agents (such as consumers, families, churches and firms) include the "senses" that give them the ability to record actions and reactions in memory. Furthermore, they have mental capabilities that permit the processing of sense data in various ways, some subset of which we might call rationality. The simple automata described above are reactive, in that their actions depend in systematic ways on external stimuli, but they can hardly be said to be rational or reflective in that their "decision process" involves no choice points, internal representations of the world or "deliberation". Such distinctions shelve into the deep waters of defining intelligence, but the important point is that we can make useful distinctions between different kinds of adaptability based on the specifications of process we use in our models without compromising the evolutionary framework we have set up. It is in this way that the complex relationship between selection and reasoned action may begin to be addressed.

Thus, even in this simple example, one can see not only the general evolutionary structure of the simulation but also that the conception differs in significant ways from the corresponding biological process.

## 18.3 Evolutionary Ideas in the Social Sciences

From early on, since the publication of "The Origin of Species" (Darwin 1859), Darwin's ideas of evolution have influenced those who have studied social phenomena. For example, Tarde (1884) published a paper discussing "natural" and "social" Darwinism. This marked a shift from looking at the social organisation of individuals to the patterns of social products (fashions, ideas, tunes, laws and so on). Tarde (1903 p. 74) put it like this:

> but self-propagation and not self-organisation is the prime demand of the social as well as of the vital thing. Organisation is but the means of which propagation, of which *generative* or *imitative* imitation, is the end.

However, it was from the latter half of the twentieth Century that the full force of the analogy with biological evolution (as understood in the Neo-Darwinian Synthesis) was felt in the social sciences. There were those who sought to understand the continuous change in cultural behaviours over long time-scales in this way, e.g. (Boyd and Richerson 1985; Campbell 1965; Cavalli-Sforza and Feldman 1973; Cloak 1975; Csányi 1989). Richard Dawkins coined the term 'meme' as a discrete and identifiable unit of cultural inheritance corresponding to the biological gene (Dawkins 1976, 1982), an idea which has influenced a multitude of thinkers including (Costall 1991; Lynch 1996; Dennett 1990; Heyes and Plotkin 1989; Hull 1982, 1988; Westoby 1994). Another stream of influence has been the Philosophy of Science via the idea that truth might result from the evolution of competing hypotheses (Popper 1979), a position known as Evolutionary Epistemology since (Campbell 1974). The ultimate reflection of the shift described by Tarde above is that the human mind is "merely" the niche where memes survive (Blackmore 1999) or which they exploit (as "viruses of the mind", Dawkins 1993) – the human brain is programmed by the memes, rather than using them (Dennett 1990). This fits in with the idea of the Social Intelligence Hypothesis (Kummer et al. 1997) that the biological reason the brain evolved is because it allows specific cultures to develop in groups giving specific survival value with respect to the ecological niches they inhabit (Reader 1970). All of these ideas hinge on the importance of imitation (Dautenhahn and Nehaniv 2002), since without this process individual memes, ideas or cultural patterns would be quickly lost.

Evolutionary theories are applied in a wide variety of disciplines. As mentioned above, evolutionary theories are applied to culture and anthropology, as in the work of Boyd and Richerson, Cavalli-Sforza and Feldman, and Csányi. The evolution of language can be seen as an analogy to biological evolution, as described by Hoenigswald and Wiener (1987). In computer science, Genetic Programming and Genetic Algorithms (as well as the more rarely used Classifier Systems) are descendants of the evolutionary view as well, for example in the work of several individuals at the Santa Fe Institute (Holland 1975; Kauffman 1993). Learning theories of humans, applied to individuals, groups and society can be tied to evolutionary theory, as shown in the work of Campbell (1965, 1974). The work of several philosophers of science also shows an evolutionary perspective on knowledge, as in the work of Popper (1979) and Kuhn (1970). Such theories have been used to account for brain development by Gerald Edelman (1992), and extended to the ms-to-minutes time scale of thought and action by William Calvin (1996a, b). Evolutionary theory (and in some cases, explicit modelling) is present in economics, often tied to the development of technology, as in the work of Nelson and Winter (1982) or to the evolution of institutions and practices as in the work of Dosi et al. (1999), Hodgson (1993) and North (1990). Sociology too has used evolutionary ideas and simulations to understand the evolution of social order (Lomborg 1996; Macy 1996), changing populations of organisations (Hannan and Freeman 1993) and the survival of so-called "strict" churches (Chattoe 2006a).

Interestingly, however, against these creative approaches must be set forces, in particular social sciences that have slowed or marginalised their adoption. In sociology, the conversion of functionalism (potentially a form of social evolution) into a virtual religion was followed by a huge backlash against untestable grand theory which made these ideas virtually beyond the pale for 20 years or so (Chattoe 2002; Runciman 1998). It is quite likely that confused associations with Social Darwinism, eugenics and even Nazism have not helped the use of biological analogies in social science from the 1940s until quite recently. In economics, the focus on deliberate rationality and well defined equilibria has meant that evolutionary approaches are judged ad hoc unless they can be reinterpreted to support the core assumptions of economics. (This can be observed, for example, in evolutionary approaches to equilibrium selection where the object is not to understand the dynamics of the system but to support the claim that particular equilibria are robust.) In psychology, while there appears to be no overt objection to evolutionary approaches, it seems to be the case (perhaps for historical reasons) that the main interest in these ideas is to explain behaviour using genetic accounts of cognitive structure rather than using evolutionary analogies.

In subsequent sections, having shown that interest in evolutionary ideas is widespread, we turn to technical details of various kinds of evolutionary algorithm, their strengths, weaknesses and social applicability so the reader is able to evaluate their use and consider applications in their own areas of research interest. We start with the Genetic Algorithm, which is easiest to describe, then move to Genetic Programming and the (more rarely used but in some sense more satisfactory as an analogy) Classifier Systems. The final example doesn't rely directly on the use of an evolutionary algorithm but clearly attempts to model a social process using a biological analogy.

## 18.4 The Basic Genetic Algorithm

This section describes the basic operation and limitations of the Genetic Algorithm. This leads to a description of ways in which the Genetic Algorithm can be generalised and a detailed discussion of one specific way of generalising it (Genetic Programming) in the subsequent section.

### *18.4.1 What Is a Genetic Algorithm?*

The Genetic Algorithm is actually a family of programmes developed by John Holland (1975) and his co-workers at the University of Michigan. The following algorithm describes the structure of a typical Genetic Algorithm. It is the different ways in which various parts of the algorithm can be implemented which produces the wide variety of Genetic Algorithms available. Each part of the algorithm will be discussed in more detail in a subsequent section. For the purposes of illustration, consider an attempt to solve the notorious Travelling Salesman Problem that involves producing the shortest tour of a set of cities at known distances visiting each once only (Grefenstette et al. 1985).

1. Represent potential solutions to the problem as data structures.
2. Generate a number of these solutions/structures and store them in a composite data structure called the Solution Pool.
3. Evaluate the "fitness" of each solution in the Solution Pool using a Fitness Function.
4. Make copies of each solution in the Solution Pool, the number of copies depending positively on its fitness according to a Reproduction Function. These copies are stored in a second (temporary) composite data structure called the Breeding Pool.
5. Apply Genetic Operators to copies in the Breeding Pool chosen as "parents" and return one or more of the resulting "offspring" to the Solution Pool, randomly overwriting solutions which are already there. Repeat this step until some proportion of the Solution Pool has been replaced.
6. Repeat steps 3, 4 and 5 until the population of the Solution Pool satisfies a Stopping Condition. One such condition is that the Solution Pool should be within a certain distance of homogeneity.

There is an obvious parallel between this algorithm and the process of biological evolution that inspired it. The string representing a solution to a problem corresponds to the genotype and each element to a gene. The Fitness Function represents the environment that selects whole genotypes on the basis of their relative performance. The Genetic Operators correspond to the processes causing genetic variation in biology that allow better genes to propagate while poorer ones are selected out. This class of Genetic Algorithms has a number of interesting properties (for further discussion see Goldberg 1989).

1. It is evolutionary. Genetic Operators combine and modify solutions directly to generate new ones. Non-evolutionary search algorithms typically generate solutions "from scratch" even if the location of these solutions is determined by the current location of the search process. The common Genetic Operators are based on biological processes of variation. Genetic Operators permit short subsections of parent solutions to be propagated unchanged in their offspring. These subsections (called schemata) are selected through their effect on the overall fitness of solutions. Schemata that produce high fitness for the solutions in which they occur continue to be propagated while those producing lower fitness tend to die out. (Note that while it is not possible to assign a meaningful fitness to single genes, it is possible to talk about the relative fitness of whole genotypes differing by one or more genes. By extension, this permits talk about successful "combinations" of genes.) The Genetic Operators also mix "genetic material" (different solutions in the Breeding Pool) and thus help to ensure that all the promising areas of the Problem Space are explored continuously. These ideas clearly resonate with the social production of knowledge, in science for example.

2. It is non-local. Each solution is potentially exploring a different area of the Problem Space although solutions can "cluster" in promising areas to explore them more thoroughly. This allows for societies to be "smarter" than their members.
3. It is probabilistic. The Fitness Function ensures that fitter solutions participate in Genetic Operators more often because they have more copies in the Breeding Pool and are thus more likely to propagate their useful schemata. However, it sometimes happens that a solution of low overall fitness contains useful schemata. The probabilistic replacement of only a proportion of the Solution Pool with new solutions means that a small number of poor solutions will survive for sufficient generations that these schemata have a chance of being incorporated into fitter solutions. This probabilistic approach to survival (when coupled with non-locality and the use of Genetic Operators) means that the Genetic Algorithm have an ability to avoid getting stuck on non-optimal peaks in the Problem Space. Consider a problem space with two peaks, one higher than the other. A simple hill climbing algorithm, if it happens to start "near" the lower peak, will climb up it and then be stuck at a non optimal position. By contrast, there is nothing to prevent the Genetic Operators from producing a new solution somewhere on the higher peak. Once this happens, there is a possibility of solutions fitter than those at the top of the lower peak and these will come to dominate the population. The search process can thus "jump" from one peak to another which most variants of hill climbing don't do.
4. It is implicitly parallel. In contrast with the behaviour of serial search algorithms that operate on a single best solution and improve it further, the Genetic Algorithm uses a population of solutions and simultaneously explores the area each occupies in the Problem Space. The results of these explorations are repeatedly used to modify the direction taken by each solution. The parallelism arises because the "side effects" of exploring the area surrounding each solution affect all the other solutions through the functioning of Genetic Operators. The whole is thus greater than the sum of its parts.
5. It is highly general. The Genetic Algorithm makes relatively few assumptions about the Problem Space in advance. Instead, it tries to extract the maximum amount of information from the process of traversing it. For example, non-evolutionary heuristic search algorithms use features like the gradient (first differential) which may not be calculable for highly irregular Problem Spaces. By contrast, in the Genetic Algorithm all operations take place directly on a representation of the potential solution. The Fitness Function also evaluates fitness directly from solutions rather than using derived measures. Although no search technique escapes the fact that all such techniques exploit *some* properties of the problem space they are applied upon, in practice Genetic Algorithms are good at finding acceptable solutions to hard problems (which, in some cases, defeat other methods), albeit not always the best solution. Ironically, social evolutionary learning may be *better* at finding the solutions to difficult problems than rationality which struggles without high levels of knowledge about environmental structure.

### 18.4.2 The Problem Representation and Initial Population

The most important step in developing a Genetic Algorithm also requires the most human ingenuity. A good representation for solutions to the problem is vital to efficient convergence. Some solutions have more obvious representations than others do. In the Travelling Salesman Problem, for example, the obvious representation is an ordered list of numbers representing cities. For example, the solution (1 4 3 2) involves starting at city 1, then going to city 4 and so on. Once a representation has been developed, a number of solutions are generated and form the initial population in the Solution Pool. These solutions can be generated randomly or they may make use of some other (perhaps "quick and dirty") algorithm producing better than random fitness. The optimum size of the initial population depends on the size of the Problem Space. A population of almost any size will ultimately converge. But the efficiency of the Genetic Algorithm relies on the availability of useful genetic material that can be propagated and developed by the Genetic Operators. The larger the initial population, the greater the likelihood that it will already contain schemata of an arbitrary quality. This must be set against the increased computational cost of manipulating the larger Solution Pool. The initial population must also be sufficiently large that it covers the Problem Space adequately. One natural criterion is that any given point in the Problem Space should not be more than a certain "distance" from some initial solution. A final requirement for a good solution representation is that all "genes" should be similarly important to overall fitness, rather than some "genes" being much more important than others. Equivalent variations at different positions should have a broadly similar effect on overall fitness. In the Travelling Salesman Problem all the positions in the list are equivalent. They all represent cities. The efficiency of the Genetic Algorithms relies on the exponential propagation of successful schemata and this efficiency is impaired if schemata differ too much in importance as the system then becomes "bottlenecked" on certain genes.

### 18.4.3 The Fitness Function

The Fitness Function is almost as important as the solution representation for the efficiency of the Genetic Algorithm. It assigns fitness to each solution by reference to the problem that solution is designed to solve. The main requirement for the Fitness Function is that it must generate a fitness for any syntactically correct solution. (These are commonly referred to as "legal" solutions.) In the Travelling Salesman Problem, an obvious Fitness Function satisfying this requirement would be the reciprocal of the tour length. The reciprocal is used because the definition of the problem involves finding the shortest tour. Given this goal we should regard shorter tours as fitter. More complicated problems like constrained optimisation can also be handled using the Fitness Function. One approach is simply to reject all solutions that do not satisfy the constraints. This involves assigning them a fitness of 0. However, where solutions satisfying the constraints are sparse, a more efficient method is to add terms to the Fitness Function reflecting the extent of constraint satisfaction. These "penalty terms" lower the fitness of solutions that fail to satisfy the constraints but do not necessarily reduce it to zero.

### *18.4.4 The Process of Reproduction*

Reproduction is sometimes classified as a Genetic Operator in that it takes a number of solutions (the Solution Pool) and produces a new set (the Breeding Pool). However, it is a Genetic Operator of a special type in that it uses additional information (the fitness of solutions and the Reproduction Function) in generating that population. The Reproduction Function links the fitness of individual solutions and the number of copies they produce. This process mimics the reproductive success of fitter organisms in biological systems. The number of copies depends on the type of Genetic Algorithm. Typically the fittest solutions in the Solution Pool may produce two or three copies while the worst may produce none. In order that potentially useful "genetic material" be retained, it is important that fitter solutions do not proliferate too rapidly, nor less fit solutions die out too fast. Despite their low fitness, poor solutions may contain useful schemata that need to be incorporated into better solutions. Ensuring "adequate" survival for instrumental efficiency is a matter of trial-and-error and depends on the problem and the type of Genetic Algorithm being used. There are two main types of reproduction strategies.

In the first, the Holland-Type algorithm (Holland 1975), the copies of each solution make up the Breeding Pool as described above. The Breeding Pool thus contains more copies of fitter solutions. There are two main kinds of Reproduction Function. The first is proportional fitness: here the number of copies produced for each solution is equal to the size of the Solution Pool normalised by some function according to the "share of fitness" accruing to each particular solution. Fitter solutions, responsible for a larger share of total fitness, produce more copies. This system is similar to that used in Replicator Dynamics (Vega-Redondo 1996): it is performance relative to the average that determines the number of offspring. The second possibility is rank-based fitness. In this case, the number of copies depends on fitness rank. For example, the fittest five solutions may receive two copies each, the least fit receive no copies and all others receive one. Both types of function have probabilistic equivalents. Instead of determining the actual number of copies, the function can determine the probability of drawing each type. The reproduction operator is then applied repeatedly, drawing from the probability distribution until the Breeding Pool is full. Clearly, this will still result in a greater proportion of fitter solutions in the Breeding Pool. The Reproduction Function can be linear or arbitrarily complex. In practice, the "shape" of the Reproduction Function is chosen on the basis of experience to optimise the performance of the Genetic Algorithm.

The second reproduction strategy, the GENITOR algorithm (Whitley 1989) does not involve a Breeding Pool. Instead of copying solutions into the Breeding Pool and then copying the results of Genetic Operators back again, the GENITOR takes parent solutions sequentially from the Solution Pool, applies Genetic Operators and returns the offspring immediately to the Solution Pool. The Solution Pool is kept sorted by rank and new solutions are appropriately placed according to fitness. A new solution either overwrites the solution with fitness nearest to its own or it is inserted into the Solution Pool so that all solutions with lower fitness move down one place and the solution with the lowest fitness is removed altogether. The GENITOR algorithm ensures that fitter solutions are more likely to become parents by using a skewed distribution to select them.

The differences between these strategies are instructive. The GENITOR algorithm is more similar to the interaction of biological organisms. The parents produce offspring that are introduced into a population that probably still contains at least one parent. Fitness affects which parents will mate, rather than generating offspring from all individuals in the Solution Pool. Even the "pecking order" interpretation of the introduction of offspring seems relatively intelligible. By contrast, the Breeding Pool in the Holland-Type algorithm seems to be an abstraction with little descriptive plausibility. The Holland-Type algorithm effectively splits the process of reproduction into two parts: the proliferation of fitter individuals and the subsequent generation of variation in their offspring. In biological systems, both processes result from the "same" act of reproduction. Furthermore, the differential production of offspring emerges from the relative fitness of parents. It is not explicitly designed into the system. In functional terms, both types of algorithm promote the survival of the fittest through variation and selective retention. In instrumental terms, one is sometimes more suitable than the other for a particular Problem Space. In descriptive terms, the GENITOR algorithm seems more appropriate to biological systems. (It can also be given a more plausible behavioural interpretation in social contexts.)

### *18.4.5 The Genetic Operators*

There are two main types of Genetic Operator that correspond to the biological phenomena of recombination and mutation. These are the original Genetic Operators developed by Holland (1975). Recombination Genetic Operators involve more than one solution and the exchange of genetic material to produce offspring. The commonest example is the Crossover Operator. Two solutions are broken at the same randomly selected point (n) and the "head" of each solution is joined to the "tail" of the other to produce two new solutions as shown in Fig. 18.1. Here $a_i$ identifies an ordered set of k genes from one parent and $b_i$ identifies those from another.

One of the two new solutions is then chosen with equal probability as the offspring to be placed in the Solution Pool.

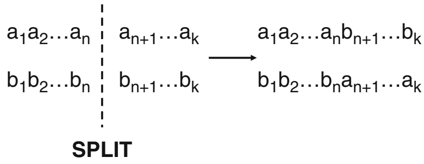



**Fig. 18.1** The Crossover Operator

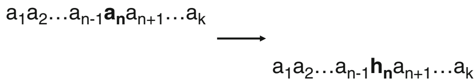



**Fig. 18.2** The (Point) Mutation Operator

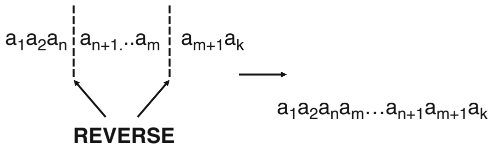



**Fig. 18.3** The Inversion Operator

Mutation Genetic Operators involve a single solution and introduce new genetic possibilities. The two main kinds of mutation Genetic Operators used in Genetic Algorithms correspond to so-called “large scale” and “point” mutation. In the Mutation Operator (realising point mutation) one gene is altered to another value from the legal range (selected with equal probability) and shown in Fig. 18.2 as $h_n$.

One commonly used Genetic Operator (corresponding to large scale chromosomal mutation) is the Inversion Operator (see Fig. 18.3) which involves reversing the order of a set of genes between two randomly selected points (n and m) in the genotype.

The Inversion Operator provides an opportunity to discuss positional effects in solution representations although these can arise in all Genetic Operators except point mutation. It is not problematic to invert (reverse) the order in which a section of a city tour takes place in the Travelling Salesman Problem. However, there may be problem representations for which we have no reason to expect that the Inversion Operator will generate solutions that are even syntactically correct let alone fit. There are two solutions to the production of illegal solutions by Genetic Operators. One is to use the penalty method. The other is simply to avoid unsuitable Genetic Operators by design. Positional effects pose particular problems if genes have different meanings, some representing one sort of object and some another. In this case, inverted solutions are almost certain not to be legal. This difficulty will be addressed further in the section on developing the Genetic Algorithm.

In biological systems, recombination ensures that the genes of sexually reproduced offspring are different from those of both parents. Various forms of mutation guarantee that entirely new genetic possibilities are also being introduced continuously into the gene pool. In the Genetic Algorithm, the Genetic Operators perform the same function, but the probability with which each is applied has to be tuned to ensure that useful genetic material can be properly incorporated before it is lost. Typically the Crossover Operator is applied with a high probability to each solution and the Mutation Operator with a low probability to each gene leading to a moderate probability of some mutation occurring in each solution. Other Genetic Operators are applied with intermediate probability. These probabilities are intended to reflect very approximately the relative importance of each process in biological systems.

For instrumental uses of the Genetic Algorithm, the setting of probabilities is a matter of experience. If the probabilities of application are too low, especially for the Mutation Operator, there is a danger of premature convergence on a local optimum followed by inefficient "mutation only" search. (In such cases, the advantages of parallel search are lost and the Genetic Algorithm effectively reverts to undirected serial search.) By contrast, if the probabilities of application are too high, excessive mixing destroys useful schemata before they can be combined into fit solutions.

There is a wide variety of other Genetic Operators discussed in the literature (Goldberg 1989; Mitchell 1996), some developed descriptively from biological systems and others designed instrumentally to work on particular problems. The descriptive use of Genetic Operators in the example provided here means that although it is important to bear the instrumental examples in mind, they should not be regarded as definitive. The processes of variation that affect the social analogues of genotypes should be established empirically just as they were for biological genes.

### *18.4.6 Convergence*

Because the Genetic Algorithm is a powerful technique, many of the problems it is used to solve are very hard to tackle by other means. Although it is possible to test the Genetic Algorithm by comparison with other techniques for simple problems, there is a danger that conclusions about performance will not scale to more complex cases. One consequence of this is the difficulty of defining satisfactory conditions for convergence. Provided the Problem Space is suitable, a non-evolutionary algorithm will find the best solution within a certain time. In the same time, the Genetic Algorithm is only statistically likely to converge (though, in practice, it will actually do so for a far larger class of problems). As a result, unlike some iterative procedures, the Genetic Algorithm cannot simply be stopped after a fixed number of generations. Instead, the properties of the Solution Pool must be analysed to determine when the programme should stop. The simplest method involves stopping when the fittest solution is "good enough". Clearly, this involves a value judgement external to the definition of the problem. Another possibility is to stop the programme when the rate of change in best solution fitness drops below a specified level. Unfortunately the behaviour of the Genetic Algorithm means that improvements in fitness are often "stepped" as the Genetic

Operators give rise to whole ranges of new possibilities to be explored. For this reason more sophisticated approaches analyse the Solution Pool continuously and measure fitness in the whole population. Another advantage of this technique is that it allows for the fact that convergence is never total because of the Mutation Operator. There is always a certain amount of "mutation noise" in the Solution Pool even when it has converged.

### *18.4.7 Developing the Genetic Algorithm*

The previous section was intended to provide a summary of the main aspects of design and a feel for the operation of a typical instrumental Genetic Algorithm (one that is supposed to solve a pre-defined problem as efficiently as possible). In the next two subsections, we describe a variety of generalisations that move the Genetic Algorithm away from the instrumental interpretation and towards the possibility of realistic *description* of certain social processes. This involves enriching the syntax for solution representations, developing formal techniques for analysing the behaviour of evolutionary models and making various aspects of the evolutionary process endogenous. The fact that these generalisations develop naturally from previous discussions suggests that a suitably sophisticated Genetic Algorithm might serve as a framework for evolutionary models of (carefully chosen) social phenomena. We shall try to show that Genetic Programming (as an extension of Genetic Algorithms) is particularly suitable for this purpose.

#### 18.4.7.1 Generalising the Solution Representation

In the simplest Genetic Algorithm, the solution representation is just a list of numbers with a fixed length. Each gene (number) in the genotype (list) represents an object like a city in the Travelling Salesman Problem. But there is no reason why the Genetic Algorithm should be limited to solving problems using such a restricted representation. The enrichment of the syntax for solution representations has proceeded in three overlapping domains: the computational improvement of programmes implementing Genetic Algorithms, the incorporation of useful insights from biology and the study of theoretical requirements for the use of different solution representations.

Developments of the first sort are those which broaden the capabilities of the Genetic Algorithm itself. Instead of solutions of fixed length "hard coded" by the programmer, Goldberg et al. (1990) have developed a "messy" Genetic Algorithm. This evolves an encoding of optimal length by varying the lengths of potential solutions as well as their encoding interpretations. Schraudolph and Belew (1992) have also addressed this problem, developing a technique called Dynamic Parameter Encoding that changes the solution encoding in response to an analysis of the current Solution Pool. (This technique avoids the loss of efficiency that results from premature convergence and the consequent failure of parallel search.) Finally, Harvey (1993) has stressed the importance of variable length genotypes in systems that are to display genuine increases in behavioural complexity.

Developments of the second sort have arisen from the study of biological systems. Smith et al. (1992) have developed a Genetic Algorithm that produces a diverse coexistent population of solutions in "equilibrium" rather than one dominated by a single "optimal" solution. In this way, the coexistent population is capable of generalisation. This approach also forms the basis of the Classifier Systems discussed in Forrest (1991). Here groups of "IF [condition] THEN [action]" rules form coexistent data structures that can jointly perform computational tasks. Belew (1989, 1990) has developed this notion further by considering models in which the solutions themselves take in information from the environment and carry out a simple form of learning. Koza (1992b, c) considers the possibility of co-evolution. This is a process in which the fitness of a solution population is not defined relative to a fixed environment or Fitness Function but rather in terms of another population. He applies this technique to game strategy learning by Genetic Programmes. Clearly this development is important to models of social systems where we can seldom define, let alone agree, a clear objective ranking of alternative social arrangements. In a sense, it is the existence of a Fitness Function that identifies instrumental (rather than descriptive) applications of Evolutionary Algorithms. The exception might be a model in which different solutions to a problem were created "subconsciously" in the style of a Genetic Algorithm but were then evaluated "rationally" by an agent. For an example see Chattoe and Gilbert (1997).

Developments of the third sort involve the adaptation of formal systems such as grammars to serve as solution representations. Antoinisse has developed a representation and set of Genetic Operators that can be used for any problem in which legal solutions can be expressed as statements in a formal grammar (Antoinisse 1991). Koza (1992a, 1994) has developed a similar though far more general representation involving the syntax of computer languages. This approach (called Genetic Programming) will receive detailed discussion in its own section shortly.

#### 18.4.7.2 Making the Process of Evolution Endogenous

So far, most of the Genetic Algorithm generalisations discussed have been instrumental in their motivation and use. The abstractions and limitations in the simple Genetic Algorithm have not been viewed as unrealistic but merely unhelpful (since they are engineering solutions rather than attempts to describe and understand complex social behaviour). The interesting question from the perspective of this chapter is how it is possible to develop simulations based on Evolutionary Algorithms which are not just instrumentally effective (allowing firms to survive by learning about their market situation for example) but actually provide a convincing ("descriptive") insight into their decision processes and the complexity of the resulting system. At the same time, the powerful self-organising capabilities of Evolutionary Algorithms may serve to provide an alternative explanation of observed stability (and instability) in social systems which do not (or cannot) involve a high level of individual rationality. Despite the instrumental nature of most current developments in Genetic Algorithms, the trend of these developments suggests an important issue for the design of descriptive models.

Most of the developments discussed above can be characterised as making various aspects of the process of evolution endogenous. Instead of exogenous system level parameters that are externally "tuned" by the programmer for instrumental purposes, various parts of the evolutionary process become internalised attributes of the individual solutions. They need not be represented in the solution explicitly as numerical parameters. They are parameters in the more general sense that they alter the process of evolution and may be adjusted by the programmer. For example, the level of mutation may emerge from some other process (such as endogenous copying of information through imitation) rather than being "applied" to the solutions. Co-evolution provides a good example of this approach. In the instrumental Genetic Algorithm, the Fitness Function is specified by the programmer and applied equally to all solutions, producing an answer to some question of interest. To follow an old Darwinian example, this is equivalent to the deliberate breeding of particular dog breeds. In co-evolving Genetic Algorithms, as in biological evolution, there is no fixed Fitness Function. Fitness can only be measured relative to the behaviour of other agents that constitute an important part of the environment. This is equivalent to the production of the dog species by biological evolution. Another example is provided by the Classifier Systems briefly discussed above. The simple Genetic Algorithm assumes that the fitness of an individual solution is independent of the fitness of other solutions. In practice, the fitness of one solution may depend on the existence and behaviour of other solutions. In biology, this is acknowledged in the treatment of altruism (Becker 1976; Boorman and Levitt 1980) and of group selection (Hughes 1988).

The use of solutions that are syntactically identical also abstracts from another important feature of evolution. Because the solutions only differ semantically there is no sense in measuring the relative "cost" of each. By contrast, when solutions differ syntactically, selection pressure may operate to produce shorter solutions as well as better ones. In descriptive models, "fitness" no longer measures an abstract quantity but describes the efficient scheduling of all scarce resources used including time. The less time is spent making decisions (provided they are sensible) the more time can be spent on other things. To put this point in its most general terms, organisms (and firms) are dynamic solutions to a dynamic environment while the simple Genetic Algorithm is a static solution to a static environment. Since social environments are dynamic, one way in which social agents can evolve or adapt is by evolving or adapting their models of that environment. Thus, an important way in which descriptive models can make the evolutionary process endogenous is by simulating agents that develop and test their own interpretations of the world in an evolutionary manner rather than being "gifted" with a fixed set of interpretations or decision processes by the modeller (Dosi et al. 1999).

The value of making parts of the process specification endogenous can only be assessed in specific cases using descriptive plausibility as the main criterion. For example, if the rate of mutation can realistically be treated as fixed over the lifetime of a given evolutionary process it makes little practical difference whether it is represented as an extra global parameter or as part of the representation for each solution. In such cases, instrumental considerations such as computational efficiency may as well decide the matter. By contrast, making fitness endogenous will probably have a major effect on the behaviour of the system. In particular, there will be a tension between the descriptive plausibility of this change and the "instrumental" desirability of convergence to a unique optimum facilitated by an external Fitness Function.

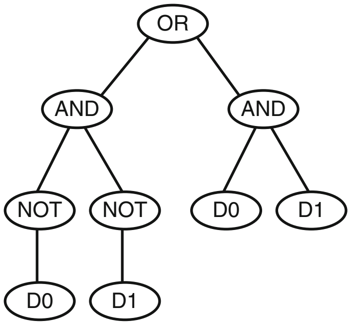
This aspect of Genetic Algorithm design provides a new insight into the distinction between instrumental and descriptive models. Instrumental models are those that allow the programmer to achieve her goals whatever they are. By contrast, the only goal that is permitted to shape a descriptive model is that of effective description as determined by empirical evidence. What determines the extent to which mutation should be modelled as a process inhering in agents is the extent to which the mutation process inheres in agents. Only once it has been shown that the mutation rate does not vary significantly across agents should it be represented as an environmental variable.

To sum up then, Genetic Algorithms constitute a broad class of powerful evolutionary search mechanisms with an active research agenda. Some (but not all) of the subsequent developments to the basic Genetic Algorithm are valuable to the descriptive modelling of social systems. (In addition some developments may have value in the characterisation of models. In the long term it may be possible to prove formal convergence results for descriptively realistic systems.) We now turn to a discussion of Genetic Programming, a significant variant of the Genetic Algorithm based on the idea of "evolving" computer programmes which can both solve instrumental problems and represent sets of practices agents use to address the problems their environment creates for them.

## 18.5 Genetic Programming

The fundamental insight of Genetic Programming (Koza 1992a, 1994) is that Evolutionary Algorithms do not need to be limited to static representations or adaptation in a static environment. The approach originated in an instrumental concern, the possibility of evolving efficient computer programmes rather than having to design them explicitly (Koza 1991). However, it rapidly became clear that the power of the technique could be extended to any process which could be represented as an algorithm provided the fitness of different solutions could be measured (Koza 1992d). The possibility of developing descriptive models of agents was also considered early on (Koza 1992c). In most models of this kind, however, the fitness of the programme representing an agent is assessed by its ability to fulfil exogenous goals. Agents typically "compete" against the environment on an equal footing rather than constituting that environment.

**Fig. 18.4** An S-expression




The potential of such an approach is tremendous. It involves the possibility of an evolutionary process that operates on the richest representation language we can envisage: the set of computable functions. These functions can model the capability to collect, abstract, store and process data from the environment, transfer it between agents and use it to determine action. Furthermore, we know that (in principle at least) languages within the class of computable functions can also represent important features of human consciousness like self-awareness and self-modification of complex mental representations (Kampis 1991; Metcalfe 1994; Fagin et al. 1995).

A simple example illustrates the most common solution representation used in Genetic Programming. This can be visualised as a tree structure and translates exactly into the set of "S-expressions" available in the LISP programming language (Friedman and Felleisen 1987). This is convenient for programming purposes because LISP comes already equipped to perform operations on S-expressions and can therefore easily and efficiently implement suitable Genetic Operators. The tree structure in Fig. 18.4 is equivalent to the S-expression (OR (AND (NOT D0) (NOT D1)) (AND D0 D1)). This is the definition of the XOR (exclusive or) function. For obvious reasons, D0 and D1 are referred to as terminals and the set {AND, OR, NOT} are referred to as functions. The choice of a suitable set of functions and terminals (the equivalent of the solution representation in Genetic Algorithms) is a key part of Genetic Programming. Functions are by no means limited to the logical operators. They can also include mathematical operators and programming language instructions. Similarly, terminals can represent numerical (or physical) constants, a variety of "sensor" inputs from the environment (including the observable actions of other agents) and "symbolic" variables like "true" and "false".

The instrumental measurement of fitness involves providing the S-expressions with different "inputs" (in this case truth values for D0 and D1) and assessing the extent to which the desired "output" results. For example, in Koza (1991) a programme to generate random numbers was tested by measuring the statistical properties of the number sequences it generated and rewarding such features as uncorrelated residuals. If S-expressions represent agents that are capable of action in an environment, success can be measured by the ability to modify the relationship between the agent and the environment in a certain way, for example by following a trail successfully. (The further along the trail an agent gets the fitter its programme.) It should be noted that the instrumental measurement of fitness requires a fairly precisely defined problem and solution grammar. On the other hand, the descriptive modelling of interaction need not. In order to do "well enough" in the market, a firm only needs to make some profit in every period sufficient to cover its costs. It may or may not have an internal goal to do better than this or even to make as much profit as it possibly can but this goal is not required for its survival (and may, in some cases, actually be counter-productive).

This discussion raises several potential difficulties with the descriptive use of Genetic Programming. However, these appear to recede on further consideration of the corresponding solutions to these problems in instrumental applications. The first difficulty is designing Genetic Operators that are guaranteed to produce meaningful offspring. In the S-expression representation, it is clear that a cut can be made at any point on the tree and the crossing of two such fragmented parents will always result in two legal offspring. However, the price to be paid for this advantage is that solutions must have a hierarchical form. More complicated function sets, mixing numerical and logical functions for example, must restrict crossover to prevent such outcomes as (+ 4 TRUE) or (NOT 6).[2]

However, given the descriptive interpretation of Genetic Operators, it is plausible that agents should know the syntactic rules of combination for the set of terminals and operators they possess. As such the relevant "descriptive" Genetic Operators may execute rather more slowly than the simple instrumental ones but it is not unreasonable to suppose that only syntactically correct trees will result. However, this raises another interesting possibility for using Genetic Operators. A good illustration is provided by a second difficulty with Genetic Programming, that of "bloating". This occurs because Genetic Programmes sometimes grow very large and contain substantial amounts of syntactically redundant material. (If a tree is trying to converge on a specific numerical value, for example, any sub trees evaluating to 0 are syntactically redundant.) Bloating produces a number of difficulties. Firstly, it slows down the evaluation of trees. Secondly, it becomes harder to interpret the trees and assess their behavioural plausibility. Finally, it is descriptively unsatisfactory. We do not expect real human decision processes to contain pointless operations (although bureaucratically specified production processes might, for example). Unfortunately, the obvious solution (the exogenous penalisation of long solutions) lacks precision. It is not possible to establish how long solutions to a particular problem "ought" to be without making arbitrary assumptions. The result is an ungrounded trade-off between length and quality. An interesting alternative is to introduce "purely syntactic" Genetic Operators. These take no account of tree fitness but simply look for redundant material within trees. For example, a Genetic Operator which replaced instances of the pattern (* constant 0) with 0 would be very simple to implement.

[2] There are other ways to solve this, e.g., implementing a strong typing (Montana 1995) or ensuring that even mixed expressions have a coherent interpretation.

This approach allows firms (for example) to apply plausible syntactic knowledge to the structure of their decision processes ("rationalisation" in the non pejorative sense) without compromising the assumption (opposed to extreme economic rationality) that they cannot evaluate the fitness of a strategy without trying it in the market.

It also suggests a possible solution to another persistent problem with Genetic Programmes, that of interpretation. Even quite small trees are often hard to interpret and thus to evaluate behaviourally. Application of syntactic Genetic Operators may reduce the tree to a form in which it can be more easily interpreted. Another approach might be to use a Genetic Programming instrumentally to interpret trees, giving the greatest fitness to the shortest tree which can predict the output of a decision process tree to within a certain degree of accuracy. Thus, in principle at least, the Genetic Programming approach can be extended to include processes that are behaviourally similar to abstraction and refinement of the decision process itself.

As in the discussion of Genetic Algorithms above, we have kept this discussion of Genetic Programming relatively technical with some digressions about its general relevance to modelling social behaviour. In the final section of this chapter, we will present some evolutionary models in social science specifically based on Evolutionary Algorithms. This discussion allows us to move from general to specific issues about the applicability of biological analogies to social systems. In particular, we will try to show why models based on Genetic Programming and some Classifier Systems are more behaviourally plausible than those based on Genetic Algorithms.

## 18.6 Example Applications of Evolutionary Algorithms

### *18.6.1 Example Using Genetic Algorithms: The Arifovic "Cobweb" Model*

Arifovic (1994) is probably responsible for the best-known simulation of this type representing the quantity setting decisions of firms to show convergence in a cobweb model.[3] She argues that the Genetic Algorithm both produces convergence over a wider range of model parameters than various forms of rational and adaptive learning, but also that it mimics the convergence behaviour of humans in experimental cobweb markets. Arifovic draws attention to two different interpretations of the Genetic Algorithm and explores the behaviour of both. In the "single population interpretation", each firm constitutes a single genotype and the Genetic Algorithm

[3] A cobweb model is one in which the amount produced in a market must be chosen before market prices are observed. It is intended to explain why prices might be subject to periodic fluctuations in certain types of markets.

operates over the whole market. In the "multiple population interpretation", each firm has a number of genotypes representing alternate solutions to the quantity setting decision and operates its own "internal" Genetic Algorithm to choose between them.

She shows that using a basic Holland-type Genetic Algorithm, neither interpretation leads to convergence on the Rational Expectations equilibrium for the cobweb market. When she adds her "Election" Genetic Operator however, both interpretations do so. The Election Operator involves using Crossover but then evaluating the offspring for profitability on the basis of the price prevailing in the previous period. The effect of this is to add some "direction" to the application of Genetic Operators, in fact a hill climbing component. An offspring is only added to the population if it would have performed better than its parents did in the previous period. This approach does not require any implausible knowledge as it is based on past events. However, it appears that the motivation for introducing the Election Operator is instrumental, namely to ensure perfect convergence to the Rational Expectations equilibrium (a goal of economic theory rather than a property of real markets necessarily). Interestingly, the graphs shown in the paper suggest that the Genetic Algorithm has done very well in converging to a stable (if mutation noise augmented) price fairly close to the Rational Expectations equilibrium. In fact, Arifovic shows how the Election Operator endogenously reduces the effective mutation rate to zero as the system approaches the theoretical equilibrium. She also points out that the Election Operator does not harm the ability of the Genetic Algorithm to learn a new equilibrium if the parameters of the cobweb model change. What she doesn't explain is why the goal of the model should be to produce the theoretical equilibrium.

In fact, there are problems with both of her models that serve as instructive examples in the application of evolutionary ideas. The single population interpretation seems to involve a standard Holland-type Genetic Algorithm even down to a Breeding Pool that has no behavioural interpretation in real systems. There is also a problem with the use of Genetic Operators that is general in Genetic Algorithms. The way in which the bit strings are interpreted is very precise. If one firm uses Crossover involving the price strategy of another, it is necessary to "gift" a common representation to all firms and assume that firms know precisely where bit strings should "fit" in their own strategies. Given the encoding Arifovic uses, inserting a bit string one position to the left by mistake doubles the price it produces. In descriptive terms, this seems to be the worst of both worlds. It is easy to see how one firm could charge the same price as another, or (with more difficulty) acquire a "narrative" strategy fragment like "keep investment in a fixed ratio to profit" but not how firms could come to share a very precise arbitrary representation and copy instances around exactly. More generally, encoding price in this way is just behaviourally odd. It is hard to imagine what a firm would think it was doing if it took a "bit" of one of its prices and "inserted it" into another. Of course, the effect would be to raise or lower the price, but the way of going about it is very bizarre.

We think the reason for this is that an encoding is not a procedure that is endogenously evolved. A Genetic Programme that calculates price by taking the previous price of another firm, adding unit cost and then adding 2 is telling a firm behaviourally how to determine price. These are "real" procedures given by the ontology of what a firm knows: the set of operators and terminals. By contrast there has to be reason why a firm would bother to encode its price as a bit string rather than just operating on it directly. Unless this encoding is "gifted", it is not clear how (or why) the firm would develop it.

The multiple population interpretation is much more plausible in behavioural terms since the problem representation only needs to be the same within a firm, although the strangeness of "combining" prices remains. A firm applying Genetic Operators to its own strategies can reasonably be assumed to know how they are encoded however.

However, both interpretations come up against a serious empirical problem noted by Olivetti (1994). Because the Election Operator is effectively a hill-climbing algorithm, it fails to converge under quite small changes to the assumptions of the model. In particular, Olivetti shows that the system doesn't converge when a white noise stochastic disturbance is added to the demand function. This suggests that Arifovic has not understood the main advantage of the Genetic Algorithm and her pursuit of instrumental convergence at the expense of behavioural plausibility is actually counter-productive. In a sense, this is just a reprise of the previous instrumental insight. Genetic Algorithms perform better on difficult problems precisely because they do not "hill climb" (as the Election Operator does) and can thus "jump" from one optimum to another through parallel search.

### *18.6.2 Example Using Classifier Systems: The Moss Price Setting Model*

As discussed briefly above, Classifier Systems consist of sets of "IF [condition] THEN [action]" rules that can collectively solve problems. They are evolutionary because new rules are typically generated using a Genetic Algorithm to select, recombine and mutate the most effective rules in the population. However, there is one significant (and potentially problematic) difference between Classifier Systems and Genetic Algorithms or Genetic Programming. This is the allocation of fitness to the individual rules, frequently using the so-called Bucket Brigade algorithm. This allows individual rules to "bid" fitness in order to take part in the set that is used to solve the problem in a particular instance. Rules taking part in a successful outcome then receive "recompense" also in terms of fitness. Unfortunately, the behavioural interpretation for this algorithm is not clear. In addition, the system is required to make decisions about how to "allocate" fitness between participating rules. This is the "Credit Assignment Problem" recognised in Artificial Intelligence and it is hard to produce effective general solutions. In particular, rules that are only used occasionally may nonetheless be essential under specific circumstances. (It is possible that an instrumental approach and lack of biological awareness have created this problem but that it is not actually intrinsic to this kind of modelling. In biological evolution there is no credit assignment. Phenotypic traits stand and fall together.)

That said, the Classifier System has one definite advantage over both Genetic Programming and Genetic Algorithms assuming these difficulties can be overcome. This is that the individual rules may be much simpler (and hence more easily interpreted behaviourally) than Genetic Programmes. This ease of interpretation also makes it more plausible that individual rules (rather than sub trees from Genetic Programmes or very precisely encoded bit strings from Genetic Algorithms) might be transferred meaningfully between firms either by interpretation of observable actions or "gossip". Interestingly, despite their advantages, Classifier Systems are easily the least applied Evolutionary Algorithms for understanding social behaviour and this lacuna offers real opportunities for new research.

In what appears to be one of the earliest applications to firm decision making, Moss (1992) compares a Classifier System and a (non-evolutionary) algorithm of his own design on the task of price setting in a monopoly. His algorithm hypothesises specific relationships between variables in the market and then tests these. For example, if an inverse relationship between price and profit is postulated, the firm experiments by raising price and seeing whether profit actually falls. If not, the hypothesis is rejected and another generated. If it works, but only over a range, then the hypothesis is progressively refined. The conclusion that Moss draws from this approach illustrates an important advantage of Genetic Programming over Genetic Algorithms and (some) Classifier Systems – that its solutions are explicitly based on process and therefore explanatory. Moss points out that the simple Classifier System simply evolves a price while his algorithm shows how the firm evolves a representation of the world that allows it to set a price. Although not doing it quite as explicitly as his algorithm, a Genetic Programme may incorporate a stylised representation of market relationships into the encoding of the decision process. (Of course, in certain circumstances the firm may lack the operators and terminals to deduce these relationships adequately or they may not form a reliable basis for action. In this case, simpler strategies like "price following" – simply setting the same price as another firm – are likely to result.)

To return to the point made by Moss, all the Classifier System models so far developed to study firm behaviour seem to be "flat" and "hard coded". By "flat" we mean that only a single rule is needed to bridge the gap between information received and action taken. In practice, the Classifier System paradigm is capable of representing sets of rules that may trigger each other in complex patterns to generate the final output. This set of rules may also encapsulate evolved knowledge of the environment although "hard coding" prevents this. For example, we might model the production process as a Classifier System in which the rules describe the microstructure of the factory floor: where each worker went to get raw materials, what sequence of actions they performed to transform them and where they put the results. In such a model events (the arrival of a partially assembled computer at your position on the production line) trigger actions (the insertion of a particular component). However, running out of "your" component would trigger a whole other set of actions like stopping the production line and calling the warehouse. The construction of such "thick" Classifier Systems is a task for future research.

"Hard coding" implies that each rule bridges the gap between input and output in the same way, suggesting the common representation of Genetic Algorithms with its attendant behavioural implausibility. In the models described above decision-makers do not have the option to add to the set of conditions or to change the mappings between conditions and actions: changing price on the basis of customer loyalty rather than costs for example. There is nothing in the Classifier System architecture to prevent this, but all the current models seem to implement the architecture in a simplified and behaviourally implausible way that makes it more like a Genetic Algorithm than Genetic Programming in terms of "hard coding" of representations and decision processes.

### *18.6.3 Example Using Genetic Programming: An Artificial Stock Market*

In this example (Edmonds 2002) there is a simulated market for a limited number of stocks, with a fixed number of simulated traders and a single "market maker". Each trader starts off with an amount of cash and can, in each trading period, seek to buy or sell each of the kinds of stock. Thus at any time a trader might have a mixture of cash and amounts of each stock. A single market maker sets the prices of each stock at the beginning of each trading period depending on the last price and the previous amount of buying and selling of it. The 'fundamental' is the dividend paid on each stock, which for each stock is modelled as a slowly moving random walk. There is a transaction cost for each buy or sell action by the traders. Thus there is some incentive to buy and hold stocks and not trade too much, but in general more money can be made (or lost) in short-term speculation. The market is endogenous except for the slowly changing dividend rate so that the prices depend on the buy and sell actions and a trader's success depends on "out-smarting" the other traders.

In the original Artificial Stock Market model (Arthur et al. 1997) each artificial trader had a fixed set of price prediction strategies. At each time interval they would see which of these strategies was most successful at predicting the price in the recent past (fixed number of time cycles) and rely on the best of these to predict the immediate future price movements. Depending on its prediction using this best strategy it would either buy or sell.

In the model presented here each trader has a small population of action strategies for each stock, encoded as a GP tree. In each time period each artificial trader evaluates each of these strategies for each stock. The strategies are evaluated against the recent past (a fixed number of time cycles) to calculate how much value (current value based on cash plus stock holdings at current market prices) the trader would have had if they had used this strategy (taking into account transactions costs and dividends gained), assuming that the prices were as in the recent past. The trader then picks the best strategy for each stock and (given constraints of cash and holdings) tries to apply this strategy in their next buy and sell (or hold) actions.

At the end of each trading period the set of action strategy trees are slightly evolved using the GP algorithm. That is to say that they are probabilistically "remembered" in the next trading round depending on their evaluated success, with a few of them crossed in a GP manner to produce new variations on the old strategies and very few utterly new random strategies introduced. As a result of this a lot of evolution of small populations is occurring, namely a population for *each* trader and *each* stock. Here each GP tree represents a possible strategy that the trader could think of for that stock. The Genetic Programming algorithm represents the trader's learning process for each stock, thinking up new variations of remembered strategies, discarding strategies that are currently unsuccessful and occasionally thinking up completely novel strategies. This is a direct implementation of Campbell's model of creative thinking known as "Blind Variation and Selective Attention" (Campbell 1965). A new trader will have relatively poor strategies generally and will not necessarily have the feedback to choose the most appropriate strategy for a new stock. By contrast, an expert will have both a good set of strategies to choose from and better judgement of which to choose. These aspects of social (evolutionary) learning are clearly important in domains where there is genuine novelty which many traditional approaches do not handle well (or in some cases at all.)

The nodes of the strategy trees can be any mixture of appropriate nodes and types. This model uses a relatively rich set of nodes, allowing arithmetic, logic, conditionals, branching, averaging, statistical market indices, random numbers, comparisons, time lags and the past observed actions of other traders. With a certain amount of extra programming, the trees can be strongly typed (Haynes et al. 1996), i.e. certain nodes can take inputs that are only a specific type (say numeric) and output a different type (say Boolean) – for example the comparison "greaterThan". This complicates the programming of the Genetic Operators but can result in richer and more specific trees.

Below are a couple of examples of strategies in this version of the stock market model. The output of the expression is ultimately a numeric value which indicates buy or sell (for positive or negative numbers), but only if that buy or sell is of a greater magnitude than a minimal threshold (which is a parameter, allowing for the "do nothing" – hold – option).

- [minus [priceNow 'stock-1'] [maxHistoricalPrice 'stock-1']] – *Sell if price is greater than the maximum historical price otherwise buy*;
- [lagNumeric [2] [divide [doneByLast 'trader-2' 'stock-3'] [indexNow]]] – *Buy or sell as done by trader-2 for stock-3 divided by the price index 3 time periods ago*.

The field of Evolutionary Computation is primarily concerned with the efficiency and effectiveness of its algorithms in solving explicitly posed problems. However, efficiency is not the primary consideration here but rather how to make such algorithms correspond to the behaviour of observed social actors. In this model, a large population of strategies within each individual would correspond to a very powerful ability in a human to find near-optimal strategies, which is clearly unrealistic. Thus a relatively small population of strategies is "better" since it does mean that particular traders get 'locked-in' to a narrow range of strategies for a period of time (maybe they all do so badly that a random, novel strategy does better eventually). This allows the emergence of "group think" and trading "styles" that can reasonably be anticipated in real markets.

Other relevant issues might be that traders are unlikely to ever completely discard a strategy that has worked well in the past. (Many evolutionary models fail to take account of the fact that humans are much better at recall from structured memory than they are at reasoning. Such a model might thus "file" all past strategies but only have a very small subset of the currently most effective ones in live memory. However, if things started going very badly, it would be easy to choose not from randomly generated strategies but from "past successes". It is an interesting question whether this would be a more effective strategy.) Clearly however the only ultimate tests are whether the resulting learning behaviour sufficiently matches that of observed markets and whether the set of operators and terminals can be grounded in (or at least abstracted from) the strategies used by real traders. (Either test taken alone is insufficient. Simply matching behaviour may be a coincidence while "realistic" trading strategies that don't match behaviour have either been abstracted inappropriately or don't really capture what traders do. It is quite possible that what they are able to report doing is only part of what they actually do.)

Given such a market and trader structure what transpires is a sort of learning "arms-race" where each trader is trying to "out-learn" the others, detecting the patterns in their actions and exploiting them. The fact that all agents are following some strategy at all times ensures that (potentially) there are patterns in existence to be out-learned. Under a wide range of conditions and parameter settings one readily observes many of the qualitative patterns observed in real stock markets – speculative bubbles and crashes, clustered volatility, long-term inflation of prices and so on. Based on the simulation methodology proposed by Gilbert and Troitzsch (2005) and the idea of generative social science put forward by Epstein (2007), this outcome shows how a set of assumptions about individual actions (how traders implement and evolve their strategies) can potentially be falsified against aggregate properties of the system such as price trends across the range of stocks. Such models are an active area of research; a recent Ph.D., which surveys these, is (Martinez-Jaramillo 2007).

### 18.6.4 Example: The Functional Survival of "Strict" Churches

There are clear advantages to using existing evolutionary algorithms to understand complex social processes as we hope we have shown through the examples above. Apart from an opportunity to discuss the "technicalities" of evolutionary algorithms through looking at simple cases, it is valuable to have programs that can be used "off the shelf" (rather than needing to be developed from scratch) and for which there is an active research agenda of technical developments and formal analysis which can be drawn on. However, the major downside of the approach has also been hinted at (and will be discussed in more detail in the conclusion). Great care must be exercised in choosing a domain of application for evolutionary algorithms in understanding complex social systems. The more an evolutionary algorithm is used “as is”, the smaller its potential domain of social application is likely to be. Furthermore, while it is possible, by careful choice of the exact algorithm, to relax some of the more socially unhelpful assumptions of evolutionary algorithms (the example of an external Fitness Function and a separate Breeding Pool have already been discussed), the danger is that some domains will simply require too much modification of the basic evolutionary algorithm to the point where the result becomes awkward or the algorithm incoherent. (A major problem with existing models has been the inability of their interpretations to stand up to scrutiny. In some cases, such as the Election Operator proposed by Arifovic, it appears that even the designers of these models are not fully aware of the implications of biological evolution.)

As suggested at the beginning of the chapter, the other approach, formerly rare but now increasingly popular is to start not with an evolutionary algorithm but with a social system and build a simulation that is nonetheless evolutionary based on the structure of that. The challenge of choosing domains with a clear analogy to biological evolution remains but is not further complicated by the need to unpick and redesign the assumptions of an evolutionary algorithm. Such an example of a “bespoke” evolutionary simulation is provided in this section.

Iannaccone (1994) puts forward an interesting argument to explain the potentially counter-intuitive finding that “strict churches are strong”. It might seem that a church that asked a lot of you, in terms of money, time and appropriate behaviour, would be less robust (in this consumerist era) than one that simply allowed you to attend on “high days and holidays” (choosing your own level of participation). However, the evidence suggests that it is the liberal churches that are losing members fastest. Iannaccone proposes that this can be explained by reflecting on the nature of religious experience. The satisfaction that people get out of an act of worship depends not just on their own level of involvement but also that of all other participants. This creates a free rider problem for “rational” worshippers. Each would like to derive the social benefit while minimising their individual contribution. Churches are thus constantly at the mercy of those who want to turn up at Christmas to a full and lively church but don’t want to take part in the everyday work (like learning to sing the hymns together) that makes this possible.

Iannaccone then argues that an interesting social process can potentially deal with this problem. If we suppose that churches do things like demanding time, money and appropriate behaviour from their worshippers, this affects the satisfaction that worshippers can derive from certain patterns of activity. If the church can somehow make non-religious activities less possible and less comfortable, it shifts the time allocations of a “rational” worshipper towards the religious activities and can simultaneously reward him or her with the greater social benefit that comes from the church “guiding” its members in this way. To take a mildly contrived example, Muslims don’t drink alcohol. They also dress distinctively. A Muslim who wanted to drink couldn't safely ask his friends to join him, could easily be seen entering or leaving a pub by other Muslims and would probably feel out of place and uncomfortable once inside (quite apart from any guilt the church had managed to instill). The net effect is that Muslims do not spend much time in pubs (while many others in the UK do) and have more time for religious activity. Of course, it is easy to pick holes in the specifics of Iannaccone's argument. Why would the Muslim not dress up in other clothes? (That itself might need explanation though.) Why not engage in another non religious activity that was not forbidden? Why assume that only religious activities are club goods? (Isn't a good night at the pub just as much a result of collective effort?)

However, regardless of the details, the basic evolutionary point is this. Religious groups set up relatively fixed "creeds" that tell members when and how to worship, what to wear and eat, how much money must be given to the church and so on. Given these creeds worshippers join and leave churches. To survive, churches need worshippers and a certain amount of "labour" and income to maintain buildings, pay religious leaders and so on. Is it in fact the case, as Iannaccone argues that the dynamics of this system will result in the differential survival of strict churches at the expense of liberal ones? This is in, fact, a very general framework for looking at social change. Organisations like firms depend on the ability to sell their product and recruit workers in a way that generates profit. Organisations like hospitals are simultaneously required to meet external goals set by their funders and honour their commitments to their "customers": On one hand, the budget for surgery may be exhausted. On the other, you can't turn away someone who is nearly dead from a car crash knowing they will never survive to the next nearest accident and emergency department. This evolutionary interplay between organisations facing external constraints and their members is ubiquitous in social systems.

Before reporting the results and discussing their implications, two issues must be dealt with. Because this is a "two sided" process (involving worshippers and churches) we must attend to the assumptions made about the behaviour of these groups. In the model discussed here, it was assumed that churches were simply defined by a fixed set of practices and did not adapt themselves. This is clearly a simplification but not a foolish one. Although creeds do adapt, they often do so over very long periods and this is a risky process. If worshippers feel that a creed is just being changed for expedience (rather than in a way consistent with doctrine) they may lose faith just as fast as in a church whose creed is clearly irrelevant to changed circumstances. Speculatively, the great religions are those that have homed in on the unchanging challenges and solutions that people face in all times and all places while the ephemeral ones are those that are particular to a place or set of circumstances. Conversely, the model assumes that worshippers are strictly rational in choosing the allocations of time to different activities that maximise their satisfaction. Again, this assumption isn't as artificial as it may seem. Although we do not choose religions like we choose baked beans, there is still a sense in which a religion must strike a chord in us (or come to do so). It is hard to imagine that a religion that someone hated and disbelieved in could be followed for long merely out of a sense of duty. Thus, here, satisfaction is being used in a strictly subjective sense without inquiring into any potential objective correlates. This life, for me, is better than that life. In terms of predicting individual behaviour, this renders satisfaction a truism but in the context of the model (and explaining the survival of different kinds of churches) what matters is not what people happen to like but the fact that they pursue it. To sum up, we could have represented the churches as more adaptive and the worshippers as less adaptive but since we are interested in the *interplay* of their behaviours (and, incidentally, this novel approach reveals a shortage of social science data about how creeds change and worshippers participate in detail), there is no definite advantage to doing so.

In a nutshell, the model works as follows (more details can be found in Chattoe 2006a). Each agent allocates their time to activities generating satisfaction (and different agents like different things to different extents). They can generate new time allocations in two main ways. One is by introspection, simply reflecting that a bit more of this and a bit less of that might be nicer. The other is by meeting other agents and seeing if their time allocations would work better. This means, for example, that an agnostic who meets a worshipper from church A may suddenly realise that leading their life in faith A would actually be much more satisfying than anything they have come up with themselves. Conversely, someone "brought up in" church B (and thus mainly getting ideas from other B worshippers about "the good life") may suddenly realise that a life involving no churchgoing at all is much better for him or her (after meeting an agnostic). Of course, who you meet will depend on which church you are in and how big the churches (and agnostic populations) are. It may be hard to meet agnostics if you are in a big strict church and similarly, there are those whom a more unusual religion might suit very well who will simply not encounter its creed. Churches are created at a low rate and each one comes with a creed that specifies how much time and money members must contribute and how many non-religious activities are "forbidden". Members can only have time allocations that are compatible with the creed of the church. These allocations determine the social benefits of membership discussed above. If a church cannot meet minimum membership and money constraints, it disappears. Thus, over time, churches come and go, differing in their "strictness", and their survival is decided by their ability to attract worshippers and contributions. Worshippers make decisions that are reasonable (but not strictly rational in that they are not able instantaneously to choose the best time allocation and church for them – which may include no church – for any state of the environment). This system reproduces some stylised facts about religion. New churches start small and are often (but not always) slow to grow. Churches can appear to fade and then experience resurgences. There are a lot of small churches and very few large ones.

What happens? In fact, there is almost no difference between the lifetimes of liberal churches and mildly strict ones. What is clear however is that very strict churches (and especially cults – which proscribe all non-religious activities) do not last very long at all. It is important to be clear about this as people often confuse membership with longevity. It is true that strict churches can grow very fast and (for a while) very large but the issue at stake here is whether they will survive in the long term. To the extent that the assumptions of the simulation are realistic, the answer would appear to be no. Thus we have seen how it is possible to implement a reasonably coherent biological analogy in a social context without using a pre-existing evolutionary algorithm.

## 18.7 Conclusion: Using Biological Analogies to Understand Social Systems

Having presented a number of case studies of evolutionary algorithms in different application areas, we are now in a position to draw some general conclusions about the design and use of evolutionary simulations. Despite the fact that some have claimed that a generalised version of evolution (Blind Variation and Selective Retention) *is* the basic template for human creativity (Campbell 1965) and that it is plausible that some processes similar to biological evolution do occur in human societies, it is unlikely that these processes will be direct translations of biological evolution in all its details. For this reason, we would propose that research into evolutionary models proceeds as follows (although it is inevitable that there will be some backward and forward interplay between the stages for reasons discussed below):

1. Start with your substantive research domain of interest (linguistics, stock markets, the rise and fall of religious groups) and consider the general arguments for representing these (or parts of them) in evolutionary terms. While it is seldom spelt out explicitly, there are actually rather few candidate "general social theories" to explain the dynamic interaction of choice and change. Unless one believes that individuals have the power and knowledge required for rational action to benefit them (and note that this condition isn't met in situations as simple as the two person one shot Prisoner's Dilemma), evolution is really the only coherent and completely specified theory available.[4] Thus (and obviously the authors are biased in this regard) if you believe that agents act on imperfect knowledge in an independently operating[5] environment (such that there often is a gap between what you expect to happen and what happens, however effectively you collect and process data about your environment), it is worth considering an evolutionary approach. We would argue that these conditions are met in most social settings but economists would disagree.
2. Consider the explicit specification of an evolutionary process for your particular domain of research (perhaps using the four process specification above as a guide). The key choice made in this context is a "coherent" object of selection (OOS) whose presence or absence is empirically accessible. This makes

[4] In fact, it might be argued that it is the *only* one. Rational choice cannot contend with novelty or the origin of social order. By focusing on *relative* performance, no matter how absolutely poor, evolution can produce order from randomness.

[5] This independence comes *both* from other social actors and physical processes like climate and erosion.

organisations and firms with any kind of formal status particularly suitable. For informal groups like families, for example, it is much less clear what constitutes a "unit". (Is it, in a traditional society setting, that they physically survive, or, in a modern setting, that they still cohabit or are still on speaking terms? The problems here are evident.) Interestingly, individuals (while obviously "physically" coherent) are still problematic as objects of selection. Unless the model involves "bare" survival, it is less obvious what happens when an agent is "selected". However, examples still exist, such as who is trading in particular markets. Most of the rest of the evolutionary process specification follows naturally from the choice of an OOS. It then becomes fairly clear what the resource driving selection is (food for tribal groups, profit for firms, membership for voluntary organisations, attention for memes), what causes the birth and death of OOS (sexual reproduction, merger, religious inspiration, bankruptcy, lack of interest or memorability and so on) and what variation occurs between OOS.

This last is an interesting area and one where it is very important to have a clearly specified domain of application. For example, consider industrial organisation. Textbook economic theory creates in the mind an image of the archetypal kettle factory (of variable size), selling kettles "at the factory gates" directly to customers and ploughing profits straight back into growth and better technology. In such a world, a firm that is successful early on can make lots of poor judgements later because it has efficient technology, market dominance, retained profit and so on. As such, evolutionary pressure rapidly ceases to operate. Further, this kind of firm does not "reproduce" (it merely gets larger) and even imitation of its strategy by other firms (that are smaller and poorer) may not cause the effective "spread" of social practices required by an evolutionary approach. (What works for the dominant firm may actually be harmful to smaller "followers".)

By contrast, we can see the more modern forms of competition by franchises and chains (Chattoe 1999) or the more realistic detail of "supply chain production" as much more naturally modelled in evolutionary terms. In the first case, firms do directly "reproduce" a set of practices (and style of product, décor, amount of choice and so on) from branch to branch. More successful chains have more branches. Furthermore, the "scale" of competition is determined by the number of branches and it is thus reasonable to say that successful business practices proliferate. Wimpy may drive out "Joe Smith's Diner" from a particular town but Joe Smith is never a real competitor with the Wimpy organisation even if he deters them from setting up a branch in that town. This means that selection pressure continues to operate with chains at any scale competing with other chains at similar scales. Short of outright monopoly, there is never a dominant market position that is stable.[6]

[6] This is probably because the market is spatially distributed and the only way of making additional profits is by opening more branches (with associated costs). There are no major economies of scale to be exploited as when the kettle factory simply gets bigger and bigger with all customers continuing to bear the transport costs.

In the second case, we can see how open ended evolution may create new opportunities for business and that supply chains as a whole constitute "ecologies" (Chattoe-Brown 2009). Initially, each firm may transport its own goods to market but once markets are sufficiently distant and numerous, there may be economies of scale in offering specialist transport and logistics services (for example, all goods going from Bristol to Cardiff in 1 week may be carried by a single carter or a firm may create a distribution infrastructure so not all goods are transported directly from origin to destination but via cost saving looped routes.) Again, it is clear how the organisations here must operate successful practices that satisfy both suppliers (those who want to deliver goods) and customers (those who want to receive them) and, further, how the nature of the business environment may change continuously as a consequence of innovation (whether technical or social). The creation of the refrigerated ship or the internal combustion engine may foreclose some business opportunities (like raising animals in the city or harness making) and give rise to others which may or may not be taken up (spot markets, garages).

These examples show several things. Firstly, it is necessary to be very clear what you are trying to understand as only then can the fitness of the evolutionary analogy be assessed. Secondly, it is useful to have a systematic way (Chattoe 1998, 2006b) of specifying evolutionary models since these stand or fall on their most implausible assumptions (particularly in social sciences which aren't very keen on this approach).[7] Thirdly, there are a lot more opportunities for evolutionary modelling than are visible to the "naked eye", particularly to those who take the trouble to develop both domain knowledge and a broad evolutionary perspective. Considering the ubiquity of branch competition and intermediate production in the real world, the economic literature is amazingly distorted towards the "autonomous kettle factory view" and simulation models of realistic market structures are scarcer still (though this is just starting to change). The price of adopting a novel method is scepticism by one's peers (and associated difficulties in "routine" academic advancement) but the rewards are large domains of unexplored research opportunities and the consequent possibility for real innovation. Finally, whilst it is always possible to use an evolutionary algorithm as a "black box learning system" within the "mind" of an agent or organisation, there is a design issue about interpreting this kind of model discussed previously. The choice of learning algorithm can make a crucial difference in simulations (Edmonds and Moss 2001) and one cannot simply assume that any learning algorithm will do.

3. Explore whether data for your chosen domain is available (or can readily be got using standard social science methods).[8] If it is available, does it exist at both the

[7] More informally, "the assumptions you don't realise you are making are the ones that will do you in".

[8] In a way, it is a black mark against simulation that this needs to be said. Nobody would dream of designing a piece of statistical or ethnographic work without reference to the availability or accessibility of data!

individual level and in aggregate? For example, is there observational data about firm price setting practices (in board meetings for example) and long term historical data about the birth, death and merger of firms in a particular industry and their prices over time? Because simulation is a relatively new method, it is still possible to build and publish exploratory (or less flatteringly "toy") models of evolutionary processes but it is likely to get harder and may become impossible unless the evolutionary model or the application domain is novel. It is almost certainly good scientific practice to make the accessibility of data part of the research design but it does not follow from this that only models based on available data are scientific. The requirement of falsifiability is served by the data being collectable "in principle", not already collected. The best argument to support claims of scientific status for a simulation is to consider (as a design principle) how each parameter could be calibrated using existing data or existing research methods. (The case is obviously weaker if someone has to come up with a new data collection method first although it helps if its approach or requirements can be sketched out a priori.)

This aspect of research design also feeds into the decision about whether to use an existing evolutionary algorithm and if so, which one. The emerging methodology of social simulation (Gilbert and Troitzsch 2005, pp. 15–18, Epstein 2007) is to make a set of empirically grounded hypotheses at the micro level (firms set prices thus) and then to falsify this ensemble at the macro level. (The real distribution of survival times for firms is thus: It does or does not match the simulated distribution of survival times produced by the model.) A problem will arise if it is hard to interpret the simulated price setting practices. Suppose, for example, we use GP to model the evolution of trading strategies in stock markets. We may use interviews or observation of real traders to decide what terminals and operators are appropriate but, having let the simulation run and observed plausible aggregate properties, we may still not know (and find it extremely hard to work out because of the interpretation issue) whether the evolved strategies used are actually anything like those which traders would (or could) use.

Equating the empirical validation of the GP grammar with the validation of the strategies evolved from it is bit like assuming that, because we have derived a Swahili grammar from listening to native speakers that we are then qualified to decide when Swahili speakers are telling the truth (rather than simply talking intelligibly). It muddles syntax and semantics. The design principle here is then to consider how the chosen evolutionary algorithm will be interpreted to establish the validity of evolved practices. (Creative approaches may be possible here like getting real traders to design or choose GP trees to trade for them or getting them to "critique" what are effectively verbal "translations" of strategies derived from apparently successful GP trees as if they were from real traders.) In this regard, it is their potential ease of interpretation that makes the relative neglect of CS models seem more surprising in evolutionary modelling.

4. Having first got a clear sense of what needs to be modelled, it is then possible to choose a modelling technique in a principled way. As the analysis of case studies

suggest, the danger with a "methods-led" approach is that the social domain will be stylised (or simply falsified) to fit the method. A subsidiary difficulty with the methods-led approach is that even if the researcher is wise enough to use a modified evolutionary algorithm to mirror a social process accurately (rather than distorting or abstracting the domain to fit the method), inadequate technical understanding may render the modified algorithm incoherent or ineffective. It is thus very important to understand fully any methods you plan to apply particularly with regard to any instrumental assumptions they contain. (In making convergence her goal for the GA cobweb model, Arifovic introduced an election operator which actually rendered the GA *less* effective in solving hard problems. This issue would probably have been foreseen in advance by a competent instrumental user of the GA technique. The muddle arose from the *interface* between social description and the GA as a highly effective instrumental optimisation device.)

Having chosen a modelling technique, all its supporting assumptions must also be examined in the light of the application domain. For example, it is very important not to confuse single and multiple population interpretations of a GA: Do firms each have multiple candidate pricing strategies and choose them by an evolutionary process or is there one overall evolutionary process, in which single pricing strategies succeed and fail with the associated firms "carrying" them? Each model (or some combination) might be justified on empirical grounds but only if the difference in interpretation is kept clearly in mind. Although we are sceptical that systems of realistic social complexity would allow this, the principled choice of methods means that it is even possible that some domains would not require simulation at all but could be handled by mathematical models of evolution like replicator dynamics (Weibull 1995) or stochastic models (Moran 1962).

By contrast, however, if the properties of the chosen social domain are too far from a standard evolutionary algorithm (such that it can neither be used wholesale or deconstructed without collapsing into incoherence), the best solution is to build a bespoke evolutionary model as was done for the "strict churches" case study. (At the end of the day, evolutionary algorithms were themselves "evolved" in a completely different engineering environment and we would not therefore expect them to apply widely in social systems. Thus, great care needs to be taken to use them only where they clearly do apply and thus have real value.) With free and widely used agent based modelling packages like NetLogo[9] and associated teaching materials (Gilbert and Troitzsch 2005; Gilbert 2007), this is now much easier than it was. Ten years ago, one reason to use an existing algorithm was simply the significant cost of building your own from scratch. To sum up this strategy of research, the decision to use, modify or build an evolutionary algorithm from scratch should be a conscious and principled one based on a clear understanding of the domain and existing social science data about it.

[9] http://ccl.northwestern.edu/netlogo/

The final piece of advice is not technical or methodological but presentational. In applying a novel method, be prepared to suffer equally at the hands of those who don't understand it and those who do! One of the hardest things to do in academia is to strike a balance between rejecting ill founded criticisms or those that translate to "I just don't like this" without also rejecting real objections that may devalue months (or even years) of your effort (and still, frustratingly for you, be part of "good science"). To judge criticisms in a novel area, you must be especially well informed and thus confident of your ground. For example, there is no clear cut evidence for Lamarckism (modification of the genotype by the phenotype during the life of the organism in a way that can then be transmitted by reproduction) in biology but in social systems such processes are ubiquitous. (Someone discovers a good way to discipline children. Those who were thus disciplined do the same thing to their children. This is an acid test because, with hindsight, the "victims" have to see it as beneficial, and thus not have been warped by it, even if it was hateful at the time. Punishments so nasty that the victims won't inflict them, or so ineffective that the parents stop bothering, will die out.) Failure to understand this issue may either set you on the path of non-Lamarckian (and thus quite possibly implausible) evolutionary models of social systems or of apologising mistakenly for building Lamarckian models which don't "truly" reflect biological evolution (when that was never the design criterion for using biological analogies in social science anyway).

The best way to address these issues is hopefully to follow the systematic procedure outlined above. This minimises the chances that you will miss things which critics can use to reject your models (and if they are hostile enough, your whole approach) and ensures that by justifying the models to yourself, you can actually justify them to others. In popular scientific folklore Darwin (still the greatest evolutionist) spent a considerable period trying to anticipate all possible objections to his theory and see how valid they were (and what counters he could provide) before he presented his work. Given how fraught the acceptance of his theory has been anyway, imagine if he had not troubled to take that step!

We hope we have shown, by the use of diverse case studies and different evolutionary modelling techniques, both the considerable advantages and (potentially avoidable) limitations of this approach and encourage interested readers to take these ideas forward both in developing new kinds of models and applying evolutionary models to novel domains. The field is still wide open and we are always pleased to hear from potential students, co-workers, collaborators, supporters or funders!

**Acknowledgements** Edmund Chattoe-Brown acknowledges the financial support of the Economic and Social Research Council as part of the SIMIAN (http://www.simian.ac.uk) node of the National Centre for Research Methods (http://www.ncrm.ac.uk).

## Further Reading

(Gilbert and Troitzsch 2005) is a good general introduction to social science simulation and deals with evolutionary techniques explicitly, while (Gilbert 2007) is recommended as an introduction of this kind of simulation for studying evolution in social systems. For deeper introductions to the basic techniques see (Goldberg 1989), which is still an excellent introduction to GA despite its age (for a more up-to-date introduction see: Mitchell 1996), and (Koza 1992a, 1994) for a very accessible explanation of GP with lots of examples. (Forrest 1991) is a good introduction to techniques in Classifier Systems.

More details about the four example models are given in the following: (Chattoe 2006a) shows how a simulation using an evolutionary approach can be related to mainstream social science issues, (Edmonds 2002) gives an example of the application of a GP-based simulation to an economic case, and (Moss 1992) is a relatively rare example of a classifier based model.

## References


Antoinisse H (1991) A grammar based genetic algorithm. In: Rawlins G (ed) Foundations of genetic algorithms: proceedings of the first workshop on the foundations of genetic algorithms and classifier systems, Indiana University. Morgan Kaufmann, San Mateo, 15–18 July 1990, pp 193–204

Arifovic J (1994) Genetic algorithm learning and the cobweb model. J Econ Dyn Control 18:3–28

Arthur WB, Holland JH, LeBaron B, Palmer R, Tayler P (1997) Asset pricing under endogenous expectations in an artificial stock market. In: Arthur WB, Durlauf SN, Lane DA (eds) The economy as a complex evolving system II (Santa Fe Institute studies in the science of complexity, proceedings volume XXVII) Addison-Wesley, Reading, pp 15–44

Becker G (1976) Altruism, egoism and genetic fitness: economics and sociobiology. J Econ Lit 14:817–826

Belew R (1989) When both individuals and populations search: adding simple learning to the genetic algorithm. In: Schaffer J (ed) Proceedings of the third international conference on genetic algorithms, George Mason University. Morgan Kaufmann, San Francisco, 4–7 June 1989, pp 34–41

Belew R (1990) Evolution, learning and culture: computational metaphors for adaptive search. Complex Syst 4:11–49

Blackmore S (1999) The meme machine. Oxford University Press, Oxford

Boorman S, Levitt P (1980) The genetics of altruism. Academic, St Louis

Boyd R, Richerson PJ (1985) Culture and the evolutionary process. University of Chicago Press, Chicago

Buss DM (1998) Evolutionary psychology: the new science of the mind. Allyn and Bacon, Boston

Calvin W (1996a) How brains think: evolving intelligence, then and now. Basic Books, New York

Calvin W (1996b) The cerebral code: thinking a thought in the mosaics of the mind. MIT Press, Cambridge, MA

Campbell DT (1965) Variation and selective retention in socio-cultural evolution. In: Barringer HR, Blanksten GI, Mack RW (eds) Social change in developing areas: a reinterpretation of evolutionary theory. Schenkman, Cambridge, MA, pp 19–49

Campbell DT (1974) Evolutionary epistemology. In: Schlipp PA (ed) The philosophy of Karl R. Popper, vol XIV, The library of living philosophers. Open Court, LaSalle, pp 412–463

Cavalli-Sforza L, Feldman M (1973) Cultural versus biological inheritance: phenotypic transmission from parents to children. Hum Genet 25:618–637

Chattoe E (1998) Just how (un)realistic are evolutionary algorithms as representations of social processes? J Artif Soc Soc Simul 1(3). http://www.soc.surrey.ac.uk/JASSS/1/3/2.html

Chattoe E (1999) A co-evolutionary simulation of multi-branch enterprises. Paper presented at the European meeting on applied evolutionary economics, Grenoble, 7–9 June. http://webu2.upmf-grenoble.fr/iepe/textes/chatoe2.PDF

Chattoe E (2002) Developing the selectionist paradigm in sociology. Sociology 36:817–833

Chattoe E (2006a) Using simulation to develop and test functionalist explanations: a case study of dynamic church membership. Br J Sociol 57:379–397

Chattoe E (2006b) Using evolutionary analogies in social science: two case studies. In: Wimmer A, Kössler R (eds) Understanding change: models, methodologies and metaphors. Palgrave Macmillan, Basingstoke, pp 89–98

Chattoe E, Gilbert N (1997) A simulation of adaptation mechanisms in budgetary decision making. In: Conte R, Hegselmann R, Terna P (eds) Simulating social phenomena (Lecture notes in economics and mathematical systems), vol 456. Springer, Berlin, pp 401–418

Chattoe-Brown E (2009) The implications of different analogies between biology and society for effective functionalist analysis (Draft paper). Department of Sociology, University of Leicester, Leicester

Cloak FT (1975) Is a cultural ethology possible? Hum Ecol 3:161–182

Costall A (1991) The meme meme. Cult Dyn 4:321–335

Csányi V (1989) Evolutionary systems and society: a general theory of life, mind and culture. Duke University Press, Durham

Darwin CR (1859) On the origin of species by means of natural selection. John Murray, London

Dautenhahn K, Nehaniv CL (eds) (2002) Imitation in animals and artifacts. MIT Press, Cambridge, MA

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Dawkins R (1982) Organisms, groups and memes: replicators or vehicles? In: Dawkins R (ed) The extended phenotype. Oxford University Press, Oxford

Dawkins R (1993) Viruses of the mind. In: Dahlbohm B (ed) Dennett and his critics. Blackwell, Malden, pp 13–27

Dennett D (1990) Memes and the exploitation of imagination. J Aesthet Art Critic 48:127–135

Dobzhansky T, Ayala FJ, Stebbins GL, Valentine JW (1977) Evolution. W.H. Freeman, San Francisco

Dosi G, Marengo L, Bassanini A, Valente M (1999) Norms as emergent properties of adaptive learning: the case of economic routines. J Evol Econ 9:5–26

Edelman G (1992) Bright air, brilliant fire: on the matter of the mind. Basic Books, New York

Edmonds B (2002) Exploring the value of prediction in an artificial stock market. In: Butz VM, Sigaud O, Gérard P (eds) Anticipatory behaviour in adaptive learning systems (Lecture notes in computer science), vol 2684. Springer, Berlin, pp 285–296

Edmonds B, Moss S (2001) The importance of representing cognitive processes in multi-agent models. In: Dorffner G, Bischof H, Hornik K (eds) Artificial neural networks: ICANN 2001, international conference Vienna, 21–25 Aug 2001, proceedings (Lecture notes in computer science), vol 2130. Springer, Berlin, pp 759–766

Epstein JM (2007) Generative social science: studies in agent-based computational modelling. Princeton University Press, Princeton

Fagin R, Halpern J, Moses Y, Vardi M (1995) Reasoning about knowledge. MIT Press, Cambridge, MA

Forrest S (1991) Parallelism and programming in classifier systems. Pitman, London

Friedman DP, Felleisen M (1987) The little LISPER. MIT Press, Cambridge, MA

Gilbert N (2007) Agent-based models, vol 153, Quantitative applications in the social sciences. Sage, London
Gilbert N, Troitzsch KG (2005) Simulation for the social scientist, 2nd edn. Open University Press, Buckingham
Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Boston
Goldberg DE, Deb K, Korb B (1990) Messy genetic algorithms revisited: studies in mixed size and scale. Complex Syst 4:415–444
Grefenstette J, Gopal R, Rosmaita B, Van Gucht D (1985) Genetic algorithms for the travelling salesman problem. In: Grefenstette J (ed) Proceedings of the first international conference on genetic algorithms and their applications, Carnegie Mellon University, Pittsburgh. Lawrence Erlbaum, Hillsdale, 24–26 July 1985, pp 160–168
Hannan MT, Freeman J (1993) Organizational ecology. Harvard University Press, Cambridge, MA
Harvey I (1993) Evolutionary robotics and SAGA: the case for hill crawling and tournament selection. In: Langton CG (ed) Artificial life III: proceedings of the workshop on artificial life, Santa Fe, June 1992. Addison-Wesley, Boston, pp 299–326
Haynes T, Schoenefeld D, Wainwright R (1996) Type inheritance in strongly typed genetic programming. In: Angeline PJ, Kinnear JE (eds) Advances in genetic programming 2. MIT Press, Boston, pp 359–376
Heyes CM, Plotkin HC (1989) Replicators and interactors in cultural evolution. In: Ruse M (ed) What the philosophy of biology is: essays dedicated to David Hull. Kluwer, Amsterdam
Hodgson G (1993) Economics and evolution: bringing life back into economics. Polity Press, Cambridge, MA
Hoenigswald HM, Wiener LS (1987) Biological metaphor and cladistics classification. Francis Pinter, London
Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor
Hughes A (1988) Evolution and human kinship. Oxford University Press, Oxford
Hull DL (1982) The naked meme. In: Plotkin HC (ed) Learning development and culture: essays in evolutionary epistemology. Wiley, New York
Hull DL (1988) Interactors versus vehicles. In: Plotkin HC (ed) The role of behaviour in evolution. MIT Press, Cambridge, MA
Iannaccone L (1994) Why strict churches are strong. Am J Sociol 99:1180–1211
Kampis G (1991) Self-modifying systems in biology: a new framework for dynamics, information and complexity. Pergamon Press, Oxford
Kauffman SA (1993) The origins of order, self-organization and selection in evolution. Oxford University Press, Oxford
Koza JR (1991) Evolving a computer program to generate random numbers using the genetic programming paradigm. In: Belew R, Booker L (eds) Proceedings of the fourth international conference on genetic algorithms, UCSD, San Diego. Morgan Kaufmann, San Francisco, 13–16 July 1991, pp 37–44
Koza JR (1992a) Genetic programming: on the programming of computers by means of natural selection. A Bradford Book/MIT Press, Cambridge, MA
Koza JR (1992b) Genetic evolution and co-evolution of computer programmes. In: Langton C, Taylor C, Farmer J, Rassmussen S (eds) Artificial life II: proceedings of the workshop on artificial life, Santa Fe, February 1990. Addison-Wesley, Redwood City, pp 603–629
Koza JR (1992c) Evolution and co-evolution of computer programs to control independently acting agents. In: Meyer J-A, Wilson S (eds) From animals to animats: proceedings of the first international conference on simulation of adaptive behaviour (SAB 90), Paris. A Bradford Book/MIT Press, Cambridge, MA, 24–28 Sept 1990, pp 366–375
Koza JR (1992d) A genetic approach to econometric modelling. In: Bourgine P, Walliser B (eds) Economics and cognitive science: selected papers from the second international conference on economics and artificial intelligence, Paris. Pergamon Press, Oxford, 4–6 July 1990, pp 57–75
Koza JR (1994) Genetic programming II: automatic discovery of reusable programs. A Bradford Book/MIT Press, Cambridge, MA

Kuhn TS (1970) The structure of scientific revolutions. University of Chicago Press, Chicago
Kummer H, Daston L, Gigerenzer G, Silk J (1997) The social intelligence hypothesis. In: Weingart P, Richerson P, Mitchell SD, Maasen S (eds) Human by nature: between biology and the social sciences. Lawrence Erlbaum, Hillsdale, pp 157–179
Lomborg B (1996) Nucleus and shield: the evolution of social structure in the iterated prisoner's dilemma. Am Sociol Rev 61:278–307
Lynch A (1996) Thought contagion, how belief spreads through society: the new science of memes. Basic Books, New York
Macy M (1996) Natural selection and social learning in the prisoner's dilemma: co-adaptation with genetic algorithms and artificial neural networks. Sociol Meth Res 25:103–137
Martinez-Jaramillo S (2007) Artificial financial markets: an agent based approach to reproduce stylized facts and to study the red queen effect (PhD thesis). Centre for Computational Finance and Economic Agents (CCFEA), University of Essex
Metcalfe J (ed) (1994) Metacognition: knowing about knowing. A Bradford Book/MIT Press, Cambridge, MA
Mitchell M (1996) An introduction to genetic algorithms. A Bradford Book/MIT Press, Cambridge, MA
Montana DJ (1995) Strongly typed genetic programming. Evol Comput 3(2):199–230
Moran PAP (1962) The statistical processes of evolutionary theory. Clarendon, Oxford
Moss S (1992) Artificial Intelligence models of complex economic systems. In: Moss S, Rae J (eds) Artificial intelligence and economic analysis: prospects and problems. Edward Elgar, Cheltenham, pp 25–42
Nelson RR, Winter SG Jr (1982) An evolutionary theory of economic change. Belknap Press of Harvard University Press, Cambridge, MA
North DC (1990) Institutions, institutional change and economic performance. Cambridge University Press, Cambridge
Oliphant M (1996) The development of Saussurean communication. Biosystems 37:31–38
Olivetti C (1994) Do genetic algorithms converge to economic equilibria? (Discussion paper, 24). Department of Economics, University of Rome "La Sapienza", Rome
Popper KR (1979) Objective knowledge: an evolutionary approach. Clarendon, Oxford
Reader J (1970) Man on earth. Collins, London
Runciman WG (1998) The selectionist paradigm and its implications for sociology. Sociology 32:163–188
Schraudolph N, Belew R (1992) Dynamic parameter encoding for genetic algorithms. Mach Learn 9:9–21
Smith R, Forrest S, Perelson A (1992) Searching for diverse co-operative populations with genetic algorithms (TCGA report, 92002). The clearinghouse for genetic algorithms, Department of Engineering Mechanics, University of Alabama, Tuscaloosa
Tarde G (1884) Darwinisme naturel et Darwinisme social. Revue Philosophique XVII:607–637
Tarde G (1903) The laws of imitation. Henry Holt, New York
Vega-Redondo F (1996) Evolution, games and economic behaviour. Oxford University Press, Oxford
Weibull J (1995) Evolutionary game theory. MIT Press, Cambridge, MA
Westoby A (1994) The ecology of intentions: how to make memes and influence people: culturology. http://ase.tufts.edu/cogstud/papers/ecointen.htm
Whitley D (1989) The GENITOR algorithm and selection pressure: why rank-based allocation of reproductive trials is best. In: Schaffer J (ed) Proceedings of the third international conference on genetic algorithms, George Mason University. Morgan Kaufmann, San Francisco, 4–7 June 1989, pp 116–121
Wilson EO (1975) Sociobiology: the new synthesis. Harvard University Press, Cambridge, MA
Windrum P, Birchenhall C (1998) Developing simulation models with policy relevance: getting to grips with UK science policy. In: Ahrweiler P, Gilbert N (eds) Computer simulations in science and technology studies. Springer, Berlin, pp 183–206

# Part IV
# Applications

# Chapter 19
# Agent-Based Modelling and Simulation Applied to Environmental Management

**Christophe Le Page, Didier Bazile, Nicolas Becu, Pierre Bommel, François Bousquet, Michel Etienne, Raphael Mathevet, Véronique Souchère, Guy Trébuil, and Jacques Weber**

**Why Read This Chapter?** To understand the recent shift of paradigms prevailing in both environmental modelling and renewable resources management that led to the emerging rise in the application of ABMS. Also, to learn about a practical way to characterize applications of ABMS to environmental management and to see this framework applied to review a selection of recent applications of

C. Le Page (✉) • D. Bazile • F. Bousquet • G. Trébuil
CIRAD, UPR47 Green, F-34398 Montpellier, France
e-mail: christophe.le_page@cirad.fr; didier.bazile@cirad.fr; francois.bousquet@cirad.fr; guy.trebuil@cirad.fr

N. Becu
CNRS, UMR8586 Prodig, 75005 Paris, France
e-mail: nicolas.becu@univ-paris1.fr

P. Bommel
PUC University, Rio de Janeiro, Brazil

CIRAD, UPR47 Green, F-34398 Montpellier, France
e-mail: pierre.bommel@cirad.fr

M. Etienne
INRA, Unité d'Ecodéveloppement, Site Agroparc, 84914 Avignon cedex 9, Avignon, France
e-mail: etienne@avignon.inra.fr

R. Mathevet
Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175 – CNRS 1919 Route de Mende, F34293 Montpellier, France
e-mail: raphael.mathevet@cefe.cnrs.fr

V. Souchère
UMR SADAPT INRA AgroParisTech, B.P. 1, 78850 Thiverval Grignon, France
e-mail: souchere@grignon.inra.fr

J. Weber
CIRAD, Direction Régionale Ile de France, 42 rue Scheffer, F-75116 Paris, France
e-mail: jaques.weber@cirad.fr

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_19,

ABMS from various fields related to environmental management including the dynamics of land use changes, water, forest and wildlife management, agriculture, livestock productions and epidemiology.

**Abstract** The purpose of this chapter is to summarize how agent-based modelling and simulation (ABMS) is being used in the area of environmental management. With the science of complex systems now being widely recognized as an appropriate one to tackle the main issues of ecological management, ABMS is emerging as one of the most promising approaches. To avoid any confusion and disbelief about the actual usefulness of ABMS, the objectives of the modelling process have to be unambiguously made explicit. It is still quite common to consider ABMS as mostly useful to deliver recommendations to a lone decision-maker, yet a variety of different purposes have progressively emerged, from gaining understanding through raising awareness, facilitating communication, promoting coordination or mitigating conflicts. Whatever the goal, the description of an agent-based model remains challenging. Some standard protocols have been recently proposed, but still a comprehensive description requires a lot of space, often too much for the maximum length of a paper authorized by a scientific journal. To account for the diversity and the swelling of ABMS in the field of ecological management, a review of recent publications based on a lightened descriptive framework is proposed. The objective of these descriptions is not to allow the replication of the models but rather to characterize the types of spatial representation, the properties of the agents, the features of the scenarios that have been explored, and also to mention which simulation platforms were used to implement them (if any). This chapter concludes with a discussion of recurrent questions and stimulating challenges currently faced by ABMS for environmental management.

## 19.1 Introduction

In this chapter, we state that there is a combined shift in the way of thinking in both ecosystem management and ecological modelling fields. Over the last 20 years, the status of computer simulation in the field of renewable resources management has changed. This chapter investigates how agent-based modelling and simulation (ABMS) may have contributed to this evolution and what are the challenges it has to face for such a combination to remain fruitful.

Biosphere 2, an artificial closed ecological system built in Arizona (USA) in the late 1980s, was supposed to test if and how people could live and work in a closed biosphere. It proved to be sustainable for eight humans for 2 years, when low oxygen level and wild fluctuations in carbon dioxide led to the end of the experience. Biosphere 2 represents the quest for "engineering Nature" that has fascinated a lot of people (including a non-scientific audience) during the second part of the last century. The human aspect of this "adventure" mainly dealt with the psychological impact on a few people living in enclosed environments. In the real world, the relationships between human beings and the biosphere are based on tight linkages between cultural and biological diversity. Launched around 20 years before the Biosphere 2 project, the Man and Biosphere Program (MAB) of UNESCO is seeking to improve the global relationship between people and their environment. This is now the kind of approach – in line with the Millennium Development Goal #7 from the United Nations – that is attracting more and more interest.

In ecological management, the place of people directly involved in the management scheme is now widely recognized as central, and the impact of their activities has both to be considered as promoting and endangering different types of biodiversity. At the same time, ABMS has progressively demonstrated its ability to explicitly represent the way people are using resources, the impact of this management on plant and animal dynamics and the way ecosystems adapt to it. The next section discusses how both trends have been reinforcing each other in more detail.

The third section of this chapter gives a review of recent applications of ABMS in the field of environmental management. To avoid confusion due to the co-existence of multiple terms not clearly distinguishable, we use ABMS here as an umbrella term to refer indifferently to what authors may have denominated "agent-based modelling", "multi-agent simulation", or even "multi-agent based simulation" (also the name of an international workshop where applications dealing with environmental management are regularly presented). Our review is covering the dynamics of land use changes, water, forest and wildlife management, but also agriculture, livestock productions and epidemiology. We are focusing here on models with explicit consideration of the stakeholders (in this chapter this is how we will denominate people directly concerned by the local environmental management system). Bousquet and Le Page (2004) proposed a more extensive review of ABMS in ecological modelling. For a specific review of ABMS dealing with animal social behaviour, see Chap. 22 in this handbook (Hemelrijk 2013).

## 19.2 A Shift in Intertwined Paradigms

During the last two decades, evidences accumulate that the interlinked fields of ecosystem management and environmental modelling are changing from one way of thinking to another. This is a kind of paired dynamics where agents of change from one field are fostering the evolution of conceptual views in the other one. A survey of the articles published in "Journal of Environmental Management" and "Ecological Modelling" – just to refer to a couple of authoritative journals in those fields – clearly reveals this combined shift of paradigms. Another indication from the scientific literature was given when the former "Conservation Ecology" journal was renamed "Ecology and Society" in June 1997.

Among ecologists, it has become well accepted that classical equilibrium theories are inadequate and that ecosystems are facing cycles of adaptive change made of persistence and novelty (Holling 1986). Concepts from the sciences of complexity are now widely adopted in ecology (Levin 1998), and the perception of ecosystems as complex adaptive systems, in which patterns at higher levels emerge from localized interactions and selection processes acting at lower levels, has begun to affect the management of renewable resources (Levin 1999).

Beyond the standard concept of "integrated renewable resource management", the challenge is now to develop a new "integrative science for resilience and sustainability" focusing on the interactions between ecological and social components and taking into account the heterogeneity and interdependent dynamics of these components (Berkes and Folke 1998). The relationships between stakeholders dealing with the access and use of renewable resources are the core of these intertwined ecological and social dynamics that are driving the changes observed in many ecosystems.

Panarchy is a useful concept to understand how renewable resources management is affected by this new paradigm in ecology. It has been formalized as the process by which ecological and social systems grow, adapt, transform, and abruptly collapse (Gunderson and Holling 2002). The back loop of such changes is a critical time when uncertainties arise and when resilience is tested and established (Holling 2004). This new theoretical background is making sense to social scientists working on renewable resources management (Abel 1998) and to interdisciplinary groups expanding ecological regime shifts theory to dynamics in social and economic systems (Kinzig et al. 2006).

For a long period, the mainstream postulate in ecological modelling has been that science should first help to understand the "natural functioning" of a given ecosystem, so that the impacts of external shocks due to human activities ("anthropic pressures") could be monitored. Models were mainly predictive, oriented towards decision makers who were supposed to be supported by powerful tools (expert systems, decision-support systems) in selecting the "best", "optimal" management option. Nowadays, command-and-control approaches are seen as "being worse than inadequate" (Levin 1999).

Evidently, there is a growing need for more flexible (usable and understandable by diverse participants) and adaptive (easily modified to accommodate unforeseen situations and new ideas) models that should allow any involved stakeholders (ecosystem and resource managers among others) to gain insights through exploration of simulation scenarios that mimic the challenges they face. Similar to the role of metaphor in narratives, such simulation models do not strive for prediction anymore, but rather aim at sparking creativity, facilitating discussion, clarifying communication, and contributing to collective understanding of problems and potential solutions (Carpenter et al. 1999). To underline the change of status of simulation models used in such a way, the term "companion modelling" has been proposed (ComMod 2003; Etienne 2011).

In recent years, ABMS has attracted more and more attention in the field of environmental management (Bousquet and Le Page 2004; Hare and Deadman

2004). Recent compilations of experiences have been edited (Gimblett 2002; Janssen et al. 2002; Bousquet et al. 2005; Perez and Batten 2006). We propose to review recent ABMS applications in the field of environmental management based on a simplified framework presented in the next section.

## 19.3 A Framework for Characterizing Applications of ABMS to Environmental Management

To standardize the description of ecological models based on the interactions between elementary entities (individual-based models and agent-based models), Grimm and others (Grimm et al. 2006) have recently proposed a protocol based on a three blocks sequence: overview, design concepts, details (ODD). It is a kind of guideline for authors wishing to publish their model whose fulfilment corresponds to an entire article devoted to communicating the details of the model. Hare and Deadman (2004) also proposed a first classification scheme from the analysis of 11 case studies. Revisiting some elements from these two contributions, we propose here to successively give some insights about: (1) the purpose of the model; (2) the way the environment is represented; (3) the architecture of the different agents; (4) the implementation (translation of the conceptual model into a computer programme); (5) the simulation scenarios.

### *19.3.1 What Is the Model's Purpose?*

As recommended by Grimm and the 27 other participants to the collective design of the ODD protocol (2006), a concise formulation of the model's purpose has to be stated first: it is crucial to understand why some aspects of reality are included while others are ignored. The reasons leading to start a modelling process are not always clearly given. Is it mainly to gain understanding and increase scientific knowledge? Is it more about raising awareness of stakeholders who do not have a clear picture of a complex system? Does it aim at facilitating communication or supporting decision? The more the information about the model's purpose will be precise, the less confusion and disbelieving about its real usefulness will remain.

### *19.3.2 How Is the Environment Represented?*

Applications of ABMS to investigate environmental management issues are relying on a fundamental principle: they represent interacting social and ecological dynamics. On one hand, agents represent some sort of stakeholders, at the level of individual people or/and at more aggregated levels some groups of individuals defining (*lato* *sensu*) institutions (social groups such as families; economic groups such as farmers' organizations; political groups such as non-governmental organizations). On the other hand, the environment, holding some sort of renewable resources, stands for the landscape. The renewable resources are contributing to define the landscape, and in turn the way the landscape is structured and developed influences the renewable resources. Typically, the resources are modified by direct actions of agents on their environment, whereas the resources also exhibit some intrinsic natural dynamics (growth, dispersal, etc.. . .). At the same time, agents' decisions are somehow modified by the environment as the state of resources evolves. In such situations, the implementation of the social dynamics is performed through defining the behaviours of agents, whereas the implementation of the natural dynamics is commonly ascribed to the spatial entities defining the environment. Furthermore, spatial entities viewed as "management entities" can support the reification of the specific relationships between a stakeholder using the renewable resource and the renewable resource itself.

Yet some applications of ABMS to environmental management do not represent any spatially-explicit natural dynamics. The data related to the environmental conditions (i.e. the overall quantity of available resource), used by the agents to make their decisions, are managed just like any other kind of information. But even when the environmental conditions are not spatially-explicit, the explicit representation of space can help to structure the interactions among the agents. For instance, in the simulation of land use changes, the cognitive reasoning of agents can be embedded in cellular automata (CA) where each cell (space portion) represents a decisional entity that considers its neighbourhood to evaluate the transition function determining the next land use. Typically, acquaintances are then straightforwardly set from direct geographical proximity. It is still possible to stick on CA with more flexible "social-oriented" ways to define the acquaintances, but as soon as decisional entities control more than a single space portion, they have to be disembodied from the spatial entities. The FEARLUS model proposed by Polhill et al. (2001) and described in Sect. 19.4 is a good illustration of such a situation.

Applications of ABMS explicitly representing some renewable resources have to deal with the fact that renewable resources are usually heterogeneously scattered over the landscape being shaped by their patterns. Any irregularities or specificities in the topological properties of the environment legitimate to incorporate a spatially-explicit representation of the environment. The combined use of geographic information systems (GIS) and ABMS is a promising approach to implement such integration (Gimblett 2002), particularly when there is a need to refer explicitly to an actual landscape in a realistic way. More generally, the relationship between an artificial landscape and a real one can be pinpointed by referring to 3 levels of proximity: (1) none, in case of theoretical, abstract landscape; (2) intermediate, when the reference to a given landscape is implicit; (3) high, when the reference to a given landscape is explicit. Theoretical ABMS applications frequently use purely abstract landscapes, such as in the well-known sugar and spice virtual world of Sugarscape (Epstein and Axtell 1996). In the intermediate case, the implicit reference to a given landscape may exist through matching proportions in the composition of the landscape and similar patterns in its spatial configuration. When the reference to an actual landscape is explicit, the use of GIS is required to design the environment. An example of such realistic representation of a given landscape is given by Etienne et al. (2003). Characterizing the relationship between the simulated environment and the reality is a good way to estimate to what extent the model may provide a wide scope: the rule “the more realistic, the less generic” is hardly refutable.

From a technical point of view, the representation of space in ABMS applications with spatially-explicit representation of the environment could be either continuous or discrete. Most of the time, the representation of space is based on a raster grid (the space is regularly dissected into a matrix of similar elementary components), less frequently it is made of a collection of vector polygons. The continuous mode is quite uncommon in ABMS. This is related to the standard scheduling of ABMS that relies on either a discrete-time approach or on a discrete-event approach. Therefore, dealing with time (regular or irregular) intervals, the spatial resolution of the virtual landscape can be chosen so that the elementary spatial entity (defined as the smallest homogeneous portion of space) can be used as the unit to characterize distances or neighbourhood. Using a discrete mode to represent the space allows to easily define aggregated spatial entities that directly refer to different ecological scales relevant to specific natural or social dynamics, as well as to the management units specifically handled by the stakeholders. The corresponding spatial entities are interrelated according to a hierarchical organization, through aggregations.

## *19.3.3 How Are the Agents Modelled?*

As we restrict our study to the sole applications with explicit consideration of the stakeholders, by “agent” we mean a computer entity representing a kind of stakeholder (individual) or a group of stakeholders. We stick with that operational definition even when the authors opt for another terminology and propose to characterize each kind of agent by considering two aspects: internal reasoning leading to decision-making and interactions with the other agents (coordination).

### 19.3.3.1 Internal Reasoning

Decision-making is the internal process that specifies how an agent behaves. It encompasses two dimensions: sophistication (from reactive to cognitive) and adaptiveness (through evaluation). What is called the “behaviour” of an agent refers to a wide range of notions. In some situations, the behaviour of an agent is simply and straightforwardly characterized by the value of a key parameter, as in the theoretical exploration of the tragedy of the commons by Pepper and Smuts (2000) where agents are either restrained (intake rate of resource is set to 50 %) or

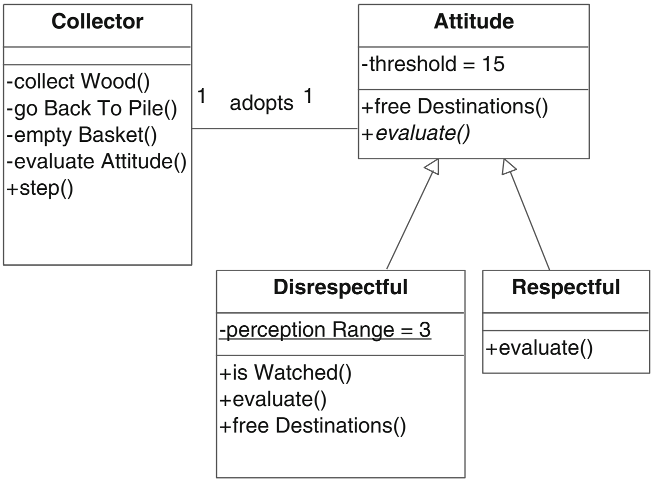



**Fig. 19.1** Design pattern of the driftwood collector agents (Thébaud and Locatelli 2001)

unrestrained (intake rate of resource is set to 99 %) in their foraging activity, when all the other biological functions (perception, movement, reproduction, mortality) are the same. In some other cases, the behaviour of a given agent does not only depend on internal characteristics, like the driftwood collector agents proposed by Thébaud and Locatelli (2001) who are stealing wood collected by other agents only when their attitude is still disrespectful (internal property) and when they cannot be observed (no peer pressure). Whatever the factors determining the behaviour of an agent are, this behaviour may or may not change over time. When the behaviour is simply characterized by a value of a key parameter, the adaptiveness can be taken into account without any particular architectural design. For more sophisticated behaviours, it becomes necessary to use a design pattern linking the agent to its behavioural attitude. With such a design pattern, the different behavioural attitudes are made explicit through corresponding subclasses, as shown in Fig. 19.1 with an example taken from the Dricol model (Thébaud and Locatelli 2001). The adoption of a particular attitude is updated according to some evaluation function.

Regarding the degree of sophistication of the decision-making process, the so-called *reactive* agents implement a direct coupling between perception (often related to little instantaneous information) and action. The forager agents of Pepper and Smuts (2000) mentioned above are typical reactive agents. On the opposite side, *cognitive* agents implement more complex decision-making processes by explicitly deliberating about different possibilities of action and by referring to specific representations of their environment, which is of particular importance for applications of ABMS to environmental management (Bousquet and Le Page 2004). An example of such agents is given by Becu et al. (2003): farmer agents evaluate direct observations and messages received from others (social network), update their knowledge base, and evaluate options according to their objectives (see the corresponding architecture in Fig. 19.2).

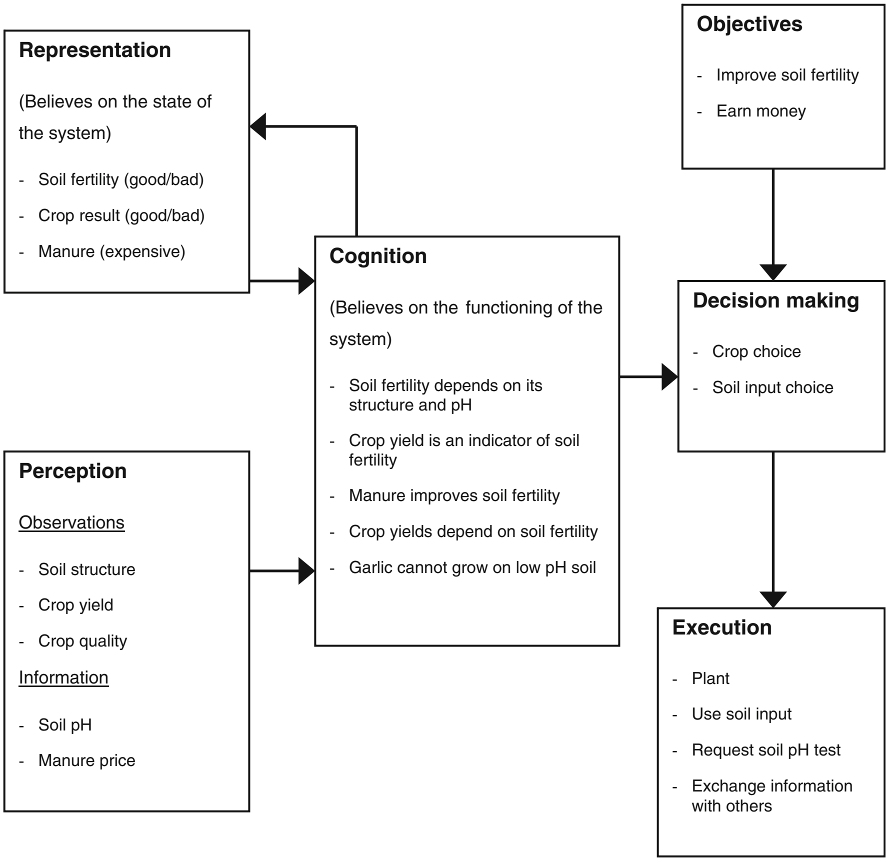



**Fig. 19.2** Architecture of farmer agents from the Catchscape model (Becu et al. 2003)

#### 19.3.3.2 Interactions with Other Agents (Coordination)

Bousquet (2001) synthesized his general approach of multi-agent systems to study environmental management issues with a diagram (see Fig. 19.3). We will refer here to the three kinds of interactions depicted in Fig. 19.3 to describe the types of agents implemented in applications of ABMS to environmental management.

The deliberative process of one agent is quite often influenced by some other closely related agents. The proximity may be either spatial (local neighbourhood) or social (acquaintances). In situations like the Dricol model (Thébaud and Locatelli 2001), what matters is just the presence of other agents in the surroundings. When more information about the related agents is needed, then the rules to access this information have to be specified. It is often assumed that the information is directly accessible through browsing the agents belonging to a given network. This corresponds to "Ie" in Fig. 19.3: other agents perceived through the environment are considered as part of the environment of one agent.

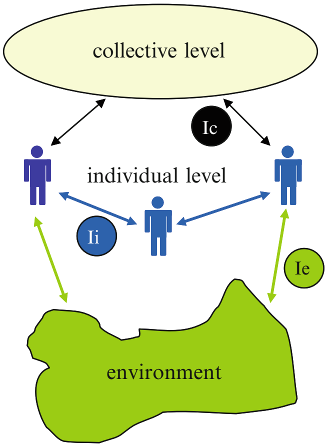



**Fig. 19.3** Interactions between agents via the environment (Ie), through peer to peer communication (Ii) and via the collective level (Ic) (Bousquet 2001)

Agents may strictly control the access to their internal information unless they intentionally decide to communicate it ("Ii" in Fig. 19.3). Then the sharing of information has to go through direct exchanges of peer-to-peer messages, with a specified protocol of communication.

Relating agents directly to the collective level ("Ic" in Fig. 19.3) is most often achieved via the notion of *groups* to which they can belong and for which they can be representative to outsiders. Inspired by the Aalaadin meta-model proposed by Ferber and Gutknecht (1998), recent applications of ABMS to agricultural water (Abrami 2004) and waste water (Courdier et al. 2002, described in the next section) management as well as to epidemiology (Muller et al. 2004, described in the next section) illustrate how both notions of *group* and *role* are useful to handle levels of organisation, and behaviours within levels of organisation. Even when «Ic» are not implemented through specific features of the agents' architecture, the mutual influence of both collective and individual levels is fundamental in renewable resources management. On one hand, individuals' behaviours are driven by collective norms and rules; on the other hand, these norms and rules are evolving through agents' interactions. This individuals-society dynamics linkage, introducing the notion of institution, relies on the representation of common referents (Bousquet 2001). Such "mediatory objects" are for instance the water temples involved in the coordination of a complex rice-terraces system in Bali (Lansing and Kremer 1993; Janssen 2007).

### *19.3.4 Implementation*

We believe it is useful to indicate whether a simulation platform was used or not to implement the model. Nowadays, some established generic tools such as Ascape, Cormas, Mason, (Net)(Star)Logo, Repast, or Swarm are being used by large communities of users. Intending to release researchers from low-level technical-operational issues, their development is boosted by their comparisons performed through the implementation of simple benchmark models (Railsback et al. 2006) and the analysis of their abilities to fulfil identified requirements (Marietto et al. 2003). The maintainers and developers of such generic platforms have also taken into consideration some sensitive technical aspects (floating point arithmetic, random numbers generators, etc.) recently pointed out (Polhill et al. 2006), and provide some elements to help users to escape these numerical traps. Additionally to the benefit of not having to re-implement basic functionalities from scratch, it may also happen that a previous model, made available online in the library of existing models, presents some similarities with the new model to be developed.

Nevertheless, using a generic platform is not a panacea. It may incidentally lead to poorer presentation of the developed models if the authors (wrongly) assume that any reader is aware of the platform's general principles. A new research stream (model-to-model comparison) recently emerged from the fact that it is very difficult to replicate simulation models from what is reported in publications (Hales et al. 2003). Reproducing results, however, is a *sine qua non* condition for making ABMS a more rigorous tool for science. It may be achieved through a better description of individual models, but also through the maintenance and development of strong communities of users sharing the same tools for implementation. This kind of stimulating diversity of the simulation platforms may contribute to identify some generic 'shorthand' conventions that could minimize the effort to describe the model rigorously and completely (Grimm et al. 2006).

### *19.3.5 Simulation Scenarios*

Some ABMS platforms like NetLogo, for simplicity purposes, merge in the same (unique) implementation file the definition of the domain entities with the specification of the simulation scenario. In their ODD protocol, Grimm et al. (2006) suggest to state the elements pertaining to the scheduling of a "standard" simulation scenario in the first part (overview) of the sequential protocol, then come back to some specific design concepts characterizing the domain entities (discussed here in Sect. 19.3.3) in the second part (design) of the sequential protocol, and finally to describe the initialization of a standard scenario in the third and last part (details). Yet, a clear separation between model and simulation should be promoted when seeking genericity. At the level of the agents, focusing on the description of their internal structure and potential behaviour may help to identify some modules of their architecture that could be re-used in other contexts. At the level of the initialization and scheduling of the simulation, the same benefit can be expected: for instance, generating parameter values for a population of agents from a statistical distribution, or creating an initial landscape fitting some schematic patterns (Berger and Schreinemachers 2006).

The notion of "standard" scenario is not always very easily recognizable. Some authors prefer to start by presenting what they call "reference" scenarios that correspond to "extreme" situations. For instance, whenever the structure of a given model makes sense to mention it, a "*no agents*" simulation scenario should be available without any modifications, i.e. just initializing the number of agents to zero. These scenarios can be used either as a verification step in the modelling process (to test that the implementation is a faithful translation of the conceptual model), or as a reference to compare the outputs obtained from more plausible simulation scenarios. More generally, simulation scenarios have to address validation by questioning the results through looking back at the system under study. In ABMS, validation is a multi-dimensional notion. Depending on the purpose assigned to the model, the focus will be mainly set on: (1) checking if the simulated data are fitting available real datasets; (2) looking for comparable processes observed in other case studies; (3) evaluating to what extent the stakeholders accept the model and its outputs as a fair representation of their system. For a more detailed discussion of validation see Chap. 8 in this volume (David 2013).

Another essential dimension of simulation scenarios relates to the model output used to observe them. Confronting the interpretations of the same simulated results built from specific stakeholders' viewpoints may be an effective way to share the different opinions and then highlight the need to improve the agents coordination mechanisms, or even to achieve a compromise (Etienne et al. 2003).

## 19.4 A Review of Recent Applications of ABMS to Environmental Management

To classify the recent applications of ABMS in environmental management is not an easy task, as the range of covered topics is wide: dynamics of land use changes, water, forest and wildlife management, but also agriculture, livestock productions and epidemiology. Some topics like epidemiology can easily be treated separately. Some others are likely to be appearing simultaneously in some case studies, especially for those dealing with multiple uses of the same renewable resource and/or representing landscape with several land-use types. This latter situation frames an entire research field in human geography that is focusing on the dynamics of land-use/cover changes (LUCC). In the classification proposed below, some applications clearly related to LUCC are listed in other subsections (mainly in "agriculture" and in "forest"). Conversely, the "LUCC" subsection contains applications that are related to some other topics specifically addressed later on. Finally, some other topics like biodiversity are multi-dimensional and thus the related case studies can be split into several other topics (for instance, biodiversity related to endangered species are reversed into "wildlife"). Whenever it undoubtedly exists for an application, we are mentioning the relevance to other topics. A specific category has been added to group the examples of ABMS addressing theoretical issues in ecological management.

As the number of publications related to the use of ABMS in ecological management is booming, it is almost impossible to analyze all of them. So for each of the categories presented above, we had to select only a few representative case studies to be briefly described by referring as much as possible to the elements discussed in previous sections (the case studies that were not selected to be analyzed are just mentioned in the introduction paragraph of each category). Following Hare and Deadman (2004) who proposed a taxonomy of ABMS in environmental management as a first step to provoke discussion and feedback, our purpose here is to contribute to the framing of a practical bibliographic survey by proposing some key characteristics useful for comparing applications of ABMS in ecological management. See Appendix for a table recording the key characteristics of the selected case studies.

### *19.4.1 Theoretical Issues in Environmental Management*

Thébaud and Locatelli (2001) have designed a simple model of driftwood collection to study the emergence of resource sharing conventions; Pepper and Smuts (2000) have investigated the evolution of cooperation in an ecological context with simple reactive agents foraging either almost everything or just half of a renewable resource. Schreinemachers and Berger (2006) have compared respective advantages of heuristic and optimizing agent decision architectures; Rouchier et al. (2001) have compared economic and social rationales of nomad herdsmen securing their access to rangelands; Evans and Kelley (2004) have compared experimental economics and ABMS results to explore land-use decision-making dynamics; Soulié and Thébaud (2006) represent a virtual fishery targeting different species in different areas to analyze the effects of spatial fishing bans as management tools.

### *19.4.2 Dynamics of Land-Use/Cover Changes*

Parker et al. (2003) have recently reviewed the application of multi-agent systems to better understand the forces driving land-use/cover change (MAS/LUCC). Their detailed state of the art presents a wide range of explanatory and descriptive applications. Since this authoritative paper has been published, new applications related to LUCC have continued to flourish. For instance, Caplat et al. (2006) have simulated pine encroachment in a Mediterranean upland, Matthews (2006) has proposed a generic tool called PALM (People and Landscape Model) for simulating resource flows in a rural livestock-based subsistence community; LUCITA (Lim et al. 2002), an ABM representing colonist household decision-making and land-use change in the Amazon Rainforest, has been developed further (Deadman et al. 2004); Bonaudo et al. (2005) have designed an ABM to simulate the pioneers fronts in the same Transamazon highway region, but at a lower scale; Manson (2006; 2005) has continued to explore scenarios of population and institutional changes in the Southern Yucatan Peninsular Region of Mexico; Huigen (2004, 2006) has developed MameLuke to simulate settling decisions and behaviours in the San Mariano watershed, the Philippines. Below we describe in more detail a selection of applications that are also characterized in the overview table presented in the Appendix.

#### 19.4.2.1 FEARLUS, Land-Use and Land Ownership Dynamics

FEARLUS, an abstract model of land use and land ownership implemented with Swarm, has been developed to improve the understanding of LUCC in rural Scotland by simulating the relative success of imitative versus non-imitative process of land use selection in different kinds of environment (Polhill et al. 2001). An abstract regional environment is defined as a toroidal raster grid made out of 8-connex land parcels, each being characterized by fixed biophysical conditions. The same external conditions that vary over time apply in the same way to all land parcels. These two factors are determining the economic return of a given land use at a particular time and place. The land manager agents decide about the land uses of the land parcels they own (initially a single one) according to a specific selection algorithm. During the simulation, they can buy and sell land parcels (landless managers leave the simulation; new ones may enter it by buying a land parcel). Simulation scenarios were defined on several grids by pairing selection algorithms from the predefined sets of 5 imitative and 5 non-imitative selection algorithms.

#### 19.4.2.2 Greenbelt to Control Residential Development

This ABM, the simplest version of which being strictly equivalent to a mathematical model, has been developed to investigate the effectiveness of a greenbelt located beside a developed area for delaying residential development outside the greenbelt (Brown et al. 2004). The environment is represented as an abstract cellular lattice where each cell is characterized by two values: a constant one to account for aesthetic quality and a variable one to denote the proximity to service centres. Service centres are called agents but actually they are more passive entities as they do not exhibit any decision making. Residential agents, all equipped with the same aesthetic and service centre preferences, decide their location among a set of randomly selected cells according to a given utility function. The Swarm platform was used for implementation. Scenarios, scheduled with periodic introductions of new agents, are based on the values of residential agents' preferences and on the spatial distribution of aesthetic quality.

#### 19.4.2.3 LUCC in the Northern Mountains of Vietnam

Castella et al. (2005a) developed the SAMBA model under the Cormas simulation platform to simulate the land use changes during the transition period of decollectivisation in a commune of the northern Vietnam uplands (Castella et al. 2005a, b; Castella and Verburg 2007). This simple and adaptable model with heuristic value represented the diversity of land use systems during the 1980s as a function of household demographic composition and paddy field endowment in the lowland areas. The environment in which agents make decisions was made of a 2,500 cell grid, and 6 different land use types could be attributed to each cell, representing a plot of 1,000 $m^2$, also characterized by its distance to the village. While there were no coordination among farmer agents with reactive behaviour in the early version of the model, interactions among them was added later and the model coupled to a GIS to extrapolate the dynamics to the regional landscape level (Castella et al. 2005b). The simulated scenarios tested the effects of the size of the environment, the overall population and household composition, and the rules for the allocation of the paddy fields on the agricultural dynamics and differentiation among farming households. More recently, this process-oriented model was compared to a spatially explicit statistical regression-based pattern-oriented model (CLUE-s) implemented at the same site. While SAMBA better represented the land use structure related to villages, CLUE-s captured the overall pattern better. Such complementarity supports a pattern-to- process modelling approach to add knowledge of the area to empirically calibrated models (Castella and Verburg 2007).

#### 19.4.2.4 Competing Rangeland and Rice Cropping Land-Uses in Senegal

To test the direct design and use of Role-Playing Games (RPG) and ABMS with farmers and herders competing for land-use in the Senegal River Valley, participatory simulation workshops were organized in several villages (D'Aquino et al. 2003). The ABM used during the last day of the workshops was straightforwardly implemented with the Cormas platform from the characteristics and rules collectively agreed the day before when crafting and testing a RPG representing stakeholders' activities related to agriculture and cattle raising. The environment is set as a raster grid incorporating soil, vegetation and water properties of the village landscape as stated by the stakeholders (a GIS was used only to clear ambiguities). The same crude rules defined and applied during the RPG were used to implement the autonomous reactive farmer agents. After displaying the scenario identified during the RPG, new questions emerged and were investigated by running the corresponding simulation scenarios. The hot debates that emerged demonstrate the potential of these tools for the improvement of collective processes about renewable resources management.

#### 19.4.2.5 Landscape Dynamics in the Méjan Plateau, Massif Central, France

Etienne et al. (2003) developed a multi-agent system in order to support a companion modelling approach on landscape dynamics in the Méjan plateau of the Massif Central, the mountain range of central France (Etienne et al. 2003). The purpose of the model is to support the coordination process among stakeholders concerned with pine encroachment. The environment is a cellular automaton coming from the rasterisation of a vector map. Several procedures account for vegetation changes due to pine encroachment according to natural succession trends and range, timber or conservation management decisions. The three agents types (sheep farmers, foresters and the National Park) are concerned by this global biological process but it affects their management goals in a very different way (sheep production, timber production, nature conservation). The model is used to simulate and compare collectively contrasting management scenarios arising from different agreements. Simulation results were used to support the emergence of collective projects leading to a jointly agreed management plan.

#### 19.4.2.6 GEMACE: Multiple Uses of the Rhone River Delta, Southern France

This ABM developed with the Cormas platform simulates the socio-economic dynamic between hunting managers and farmers in the Camargue (Rhone river delta, southern France), through the market of the wildfowling leasing system, in interaction with ecological and spatial dynamics (Mathevet et al. 2003a). A CA represents an archetypal region based on a spatial representation of the main types of estates, distributed around a nature reserve. Each cell is characterized by water and salt levels through land relief, land-use history, infrastructure, spatial neighbourhood, and current land use. A wintering duck population, heterogeneously distributed in its habitats, is affected by various factors such as land-use changes, wetland management, hunting harvest, and disturbance. Land-use decisions are made at farmland level by farmers and hunting managers that are communicating agents. Their strategy, farming or hunting oriented, is based on crop rotation, allocation of land use and water management, and may change according to some specific representations and values related to farming and hunting. Scenario runs allowed discussing the structuring of the waterfowl hunting area resulting from the individual functioning of farms in conjunction with a nature reserve and other hunting units and the conservation policy.

### *19.4.3 Water Management*

In the field of sustainable development, water management resources are an issue of major importance. ABMS dealing with water management are used to simulating the management of irrigated ecosystems, to represent the interactions among stakeholders by capturing their views and formalizing the decision-making mechanisms (especially negotiation processes), to capture the socioeconomic aspects of potable water management and evaluate scenarios based on alternative control measures, etc. For instance, Haffner and Gramel (2001) have investigated strategies for water supply companies to deal with nitrate pollution; Janssen (2001) has simulated the effects of tax rates related to the intensive use of phosphorus on lake eutrophication; Becu et al. (2003) have developed CATCHSCAPE to simulate the impact of upstream irrigation management on downstream agricultural viability in a small catchment of Northern Thailand; Krywkow et al. (2002) have simulated the effects of river engineering alternatives on the water balance of the Meuse river in the Netherlands, and have related this hydrological module to stakeholders' negotiations and decisions. Below we describe in more detail a selection of applications that are also characterized in the table presented in the Appendix.

#### 19.4.3.1 SHADOC: Viability of Irrigated Systems in the Senegal River Valley

To examine how existing social networks affect the viability of irrigated systems in the Senegal River Valley, the SHADOC ABM focuses on rules used for credit assignment, water allocation and cropping season assessment, as well as on organization and coordination of farmers in an irrigation scheme represented as a place of acquisition and distribution of two resources: water and credit (Barreteau and Bousquet 2000; Barreteau et al. 2004). The model used a spatially non-explicit representation: all plots are subject to the same hydrological cycle regardless of their exact geographical position. The societal model is structured with three types of group agents in charge of credit management, watercourse and pumping station management. As far as individual agents (farmers) are concerned, the model employs a four-level social categorization with different types of farmers according to their own cultivation objective. Each agent acts according to a set of rules local to him. Each agent also has its own point of view about the state of the system and especially its potential relations with other agents. SHADOC was first designed as a tool for simulating scenarios of collective rules and individual behaviours.

#### 19.4.3.2 MANGA: Collective Rules of Water Allocation in a Watershed

MANGA has been developed to test the economics, environmental, and ethical consequences of particular water rules in order to improve the collective management of water resources according to agricultural constraints, different actors' behaviours, and confrontation of decision rules of each actor (Le Bars et al. 2005). Their modelling approach takes into account cognitive agents (farmers or water supplier) trying to obtain the water they need via negotiation with the others as a result of its individual preferences, rationality, and objectives. The MANGA model used a spatially non-explicit representation for coupling social and environmental models. To implement the decision-making process of the cognitive agents, the authors used the BDI formalism and more particularly the PRS architecture. During simulations, MANGA allows to test several water allocation rules based on water request, irrigated corn area or behaviour evolution.

#### 19.4.3.3 SINUSE: Water Demand Management in Tunisia

Sinuse is a simulator conceived to simulate the interactions between a water table and the decisions of farmers in Tunisia (Feuillette et al. 2003). The farmers' decisions are driven by economic objectives, but the dynamics of the system is mainly dependent on the interactions among agents. The agents interact through message sending to exchange land, and to team up to build wells. They also interact through imitation and influence on the land price. They interact through the environment as they share a common resource, the water table which has its own dynamics and depends on the number of active wells. The model was developed with Cormas platform. Simulations study the influence of various policies such as subsidies for improved irrigation equipment.

#### 19.4.3.4 Water Management and Water Temple Networks in Bali

Do irrigation systems necessarily need a centralized authority to solve complex coordination problems? An ancestral Balinese system of coordination based on villages of organized rice farmers (subaks) linked via irrigation canals has served as a case study to investigate this question (Janssen 2007; Lansing and Kremer 1993). Actions to be done on each specific date for each subak are traditionally related to offerings to temples. The original model was recently re-implemented to deeper investigate why the temple level would be the best level for coordination. The environment is set as a network of 172 subaks, together with a network of 12 dams allocating the water to the subaks. Each subak has up to 4 neighbouring subaks. It selects one cropping plan out of 49 predefined ones. The corresponding water demand is affecting the runoff between dams. Harvests are affected by water stresses and pest outbreaks. The densities of pest in subaks are changing due to local growth (related to the presence of rice) and migration (based on a diffusion process). Six simulation scenarios based on the level of social coordination were explored by Lansing and Kremer. Additionally, to the two extreme scenarios defined with a single group of all 172 subaks (full synchronization) and 172 separate groups (no synchronization), 4 intermediate scenarios were tested, based on groups defined from the existing system of temples.

#### 19.4.3.5 Sharing Irrigation Water in the Lingmuteychu Watershed, Bhutan

Raj Gurung and colleagues used ABMS, following the companion modelling approach, to facilitate water management negotiations in Bhutan (Gurung et al. 2006). A conceptual model was first implemented as a role-playing game to validate the proposed environment, the behavioural rules, and the emergent properties of the game. It was then translated into a computerized multi-agent system under the Cormas platform, which allowed different scenarios to be explored. Communicating farmer-agents exchanged water and labour, either within a kinship network or among an acquaintance network. Different modes of communication (intra-village and inter-village) were simulated and a communication observer displayed the exchange of water among farmers.

### *19.4.4 Forestry*

Applications of ABMS in forestry are either focusing on LUCC issues or on management issues. For instance, Moreno et al. (2007) have simulated social and environmental aspects of deforestation in the Caparo Forest Reserve of Venezuela; Nute et al. (2004) have developed NED-2, an agent-based decision support system that integrates vegetation growth, wildlife and silvi-culture modules to simulate forest ecosystem management plans and perform goal analysis on different views of the management unit. Below we describe in more detail a selection of applications that are also characterized in the table presented in the Appendix.

#### 19.4.4.1 Deforestation and Afforestation in South-Central Indiana

Hoffmann et al. (2002) propose an original way of using ABMS to improve scientific knowledge on the interactions between human activities and forest patterns in Indiana, during the last 200 years (Hoffmann et al. 2002). The environment is a raster artificial landscape randomly generated from the 1820's land-cover ratio between crops, fallows and forests, and randomly calculated slopes. Farmer is the only type of agent identified but they can behave differently according to two potential goals (utility maximizing or learning reinforcement) and two actions: deforestation or afforestation. Simulations are used to check through statistical analysis of a high number of runs, the impact of ecological (slope), social (stakeholders goals) or economic (agricultural prices, returns) factors in changing land-use patterns.

#### 19.4.4.2 Forest Plantation Co-management

This ABMS modeling approach links social, economic and biophysical dynamics to explore scenarios of co-management of forest resources in Indonesia (Purnomo and Guizol 2006; Purnomo et al. 2005). The purpose is to create a common dynamic representation to facilitate negotiations between stakeholders for growing trees. The environment is a simplified forest landscape (forest plots, road, agricultural land) represented on a cellular automaton. Each stakeholder has explicit communication capacities, behaviours and rationales from which emerge specific actions that impact landscape dynamics. The model is used to simulate different types of collaboration between stakeholders and both biophysical and economic indicators are provided to measure the impact of each scenario on forest landscape and smallholders incomes. Simulations results are supposed to support the selection of the system of governance providing the best pathway to accelerate plantation development, local community poverty alleviation and forest landscape improvement.

### *19.4.5 Wildlife*

Understanding how human activities impact on the population of animals in the wild is a concern shared by conservationists, by external harvesters (hunters, fishermen) and by local people. Viewed as a source of food or as an emblem of biodiversity, management schemes first have to ensure the viability of the population. Viewed as competitors for the living space of local people, management schemes have to control the population. For instance, Zunga et al. (1998) have simulated conflicts between elephants and people in the Mid-Zambezi Valley; Galvin et al. (2006) have used ABMS to analyse how the situation in the Ngorongoro Conservation Area (NCA) in northern Tanzania could be modified to improve human welfare without compromising wildlife conservation value; Jepsen et al. (2005) have investigated the ecological impacts of pesticide use in Denmark. Below we describe in more detail a selection of applications that are also characterized in the table presented in the Appendix.

#### 19.4.5.1 Water Management in Mediterranean Reedbeds

Using the Cormas platform, the authors have developed an ABM to be used in environmental planning and to support collective decision-making by allowing evaluation of the long-term impact of several water management scenarios on the habitat and its fauna of large Mediterranean reedbeds (Mathevet et al. 2003b). A hydro-ecological module (water level, reedbed, fish, common and rare bird populations) is linked to a socio-economic module (reed market and management). Each cell was assigned a type of land use, land tenure and topography from a GIS to create a virtual reedbed similar to the studied wetland. Five types of interacting agents represent the users of the wetland. They are characterized by specific attributes (satisfaction, cash amount, estates etc.) and exploit several hydro-functional units. The behaviour of the agents depends on their utility function based on their evaluation of the access cost to the reedbed and on their beliefs. The ecological and socio-economic consequences of individual management decisions go beyond the estates, and relate to the whole system at a different time scale.

#### 19.4.5.2 Giant Pandas in China

Using data from Wolong Nature Reserve for giant pandas (China), this ABM simulates the impact of the growing rural population on the forests and panda habitat (An et al. 2005). The model was implemented using Java-Swarm 2.1.1 and IMSHED that provides a graphical interface to set parameters and run the program. It has three major components: household development, fuelwood demand, and fuelwood growth and harvesting. The simulated landscape was built from GIS data. Two resolutions were identified for sub-models requiring extensive human demographic factors and for landscape sub-models. Both person and household are cognitive agents that were defined from socioeconomic survey. They allowed simulating the demographic and household dynamics. Agents interact with each other and their environment through their activities according to a set of rules. The main interaction between humans and the environment is realized through fuelwood collection according to demand. This model was used to test several scenarios and particular features of complexity, to understand the roles of socioeconomic and demographic factors, identifying particular areas of special concern, and conservation policy.

#### 19.4.5.3 Traditional Hunting of Small Antelopes in Cameroon

To investigate the viability of populations of blue duikers, a small antelope traditionally hunted by villagers in the forests of Eastern Cameroon, an ABM has been developed with the Cormas platform (Bousquet et al. 2001). The raster spatial grid was defined by reading data from a GIS map corresponding to a village that was surveyed during several months. Each cell represents 4 ha (the size of the blue duiker habitat) and is characterized by a cover (river, road or village) and a reference to a hunting locality. The population dynamics of the blue duiker is simulated through the implementation of biological functions (growth, age-dependent natural mortality, migration, reproduction) applied to all the individual antelope agents. Hunter agents decide on the location of their traps by selecting one hunting locality out of the four they use. A first set of simulation scenarios was based on unilateral decisions of hunter agents, all of them following the same general rule. Coordination among kinship groups of hunters was introduced in a second set of experiences.

#### 19.4.5.4 Whale-Watching in Canada

To investigate the interactions between whale-watching boats and marine mammals in the Saguenay St. Lawrence Marine Park and the adjacent Marine Protected Area in the St. Lawrence estuary, in Quebec, this ABM was implemented with the RePast platform (Anwar et al. 2007). A raster grid defined from a GIS database represents the landscape of the studied area. The boats are cognitive agents and whales are simple reactive agents. Several simulations were run to explore various decision strategies of the boat agents and how these strategies can impact on whales. For each simulation, the happiness factor was used as an indicator of how successful the boat agents were in achieving their goals. Results showed that cooperative behaviour that involves a combination of innovator and imitator strategies based on information sharing yields a higher average happiness factor over non-cooperative and purely innovators behaviours. However, this cooperative behaviour creates increased risk for the whale population in the estuary.

### *19.4.6 Agriculture*

ABMS applied to agriculture is mainly focussing on decision-making processes at the farm level (typically, agents represent households). Economic aspects usually play a pivotal role and standard procedures like linear programming are often used to represent individual choices among available production, investment, marketing alternatives, etc. This economic module is then embedded into a more integrated framework to explicitly represent spatial and social aspects. For instance, to investigate technology diffusion, resource use changes and policy analysis in a Chilean region, Berger (2001) has connected an economic sub-model based on recursive linear programming to an hydrological sub-model; Ziervogel et al. (2005) have used ABMS to assess the impact of using seasonal forecasts among smallholder farmers in Lesotho; Sulistyawati et al. (2005) have analyzed the consequence at the landscape level of swidden cultivation of rice and the planting and tapping of rubber by Indonesian households whose demography and economic welfare are simulated. Below we describe in more detail a selection of applications that are also characterized in the table presented in the Appendix.

#### 19.4.6.1 Agricultural Pest and Disease Incursions in Australia

To analyse the effectiveness and regional economic implications of alternative management strategies for a range of different scenarios of a disease incursion in the agricultural sector, an ABM has been developed with the Cormas platform and applied to the case of the wheat disease Karnal bunt in a region of south eastern Queensland, Australia (Elliston and Beare 2006). A cellular spatial grid allows representing the spread of the pest across neighbouring paddocks and a range of potential transmission pathways including the wind, farm inputs and agents (farmers, contractors and quarantine officers) through their movement over the spatial grid. Farmers make cropping decisions about planting, spraying for weeds, harvesting and the use of contract labour. They can directly identify and report signs of a Karnal bunt incursion on their property. The incursion can also be detected from quality inspection when farm production reaches the collective storage unit. Then a quarantine response, based on a recursive checking in the neighbourhood of infected farms, is implemented by officer agents. Simulation scenarios are based on one hand on levels of farmer detection and reporting and on the other hand on the way the disease was first introduced into the system (limited and slowly expanding incursion versus potentially rapid expansion from a wide use of contaminated fertiliser).

#### 19.4.6.2 AgriPolis: Policy Impact on Farms Structural Changes in Western Europe

Agripolis is the evolution of a model developed by Balmann (1997). It describes the dynamics of an agricultural region composed of farms managed by farmers (an agent represents both of these concepts) (Happe et al. 2006). The Landscape and the Market are the other agents. The farm agent has a cognitive decision-making process: this process corresponds to the traditional modelling in agricultural economics, where agents try to maximise their income. The land market is the central interaction institution between agents in Agripolis. Farm agents extend their land by renting land from farm landowners. The allocation of land is done through auctions. The Agripolis model was used to study a region in southwest Germany. A sensitivity analysis is done to analyze the relationship between policy change and determinants of structural changes such as the interest rate, managerial abilities and technical change.

#### 19.4.6.3 Adaptive Watershed Management in the Mountains of Northern Thailand

This companion modelling experiment aims at facilitating a learning process among Akha highlanders about the socio-economic aspects (i.e. allocation of formal & informal credit, on & off-farm employment) of the expansion of plantation crops to mitigate soil erosion risk on steep land (Barnaud et al. 2007). Farmers' individual decision-making regarding investment in perennial crops, assignment of family labour to off-farm wage earning activities and search for credit were modelled. The simulated scenarios looked at the effects of the duration of the grace period of the loans, the distribution of formal credit amongst the 3 types of farms, and different structures of the networks of acquaintances for informal credit. Two main indicators were used to analyze the results of the simulations for each type of farm: (1) the total area under plantation crops (ecological indicator) and (2) the proportion of bankrupt farms leaving the agricultural sector (socio-economic indicator).

## *19.4.7 Livestock Management*

Janssen et al. (2000) have used adaptive agents to study the co-evolution of management and policies in a complex rangeland system; another work lead by Janssen (2002) has investigated the implications of spatial heterogeneity of grazing pressure on the resilience of rangelands; Bah et al. (2006) have simulated the multiple uses of land and resources around drillings in Sahel under variable rainfall patterns; Milner-Gulland et al. (2006) have built an ABM of livestock owners' decision-making, based on data collected over 2 years in five villages in southeast Kazakhstan. Below we describe in more detail a selection of applications that are also characterized in the table presented in the Appendix.

### 19.4.7.1 Rangeland Patterns in Australia

To evaluate general behaviours of rangeland systems in Australia, this ABM represents a landscape made of enterprises, which are cognitive agents that represent a commercial grazing property (Gross et al. 2006). Each property is defined by an area, the quality of land in each patch, and its livestock. Behaviours of the enterprise agents are defined by a strategy set comprised of a set of rules, which evolves over time to represent learning. A government agent has an institutional strategy set that also varies through time. Its main roles are to collect taxes and deliver drought relief in the form of interest payment subsidies. The biophysical sub-models allow simulating plant and livestock dynamics. Pastoral decisions are made by the enterprise agents according to a set of rules. The variation in the level of financial weakness leads to the adoption of a new strategy by an enterprise. Each one is randomly associated with a rate of learning. Implemented in the C++ programming language, the model is fed by inputs of historical data. The simulations emphasize consequences of interactions between environmental heterogeneity and learning rate.

### 19.4.7.2 Collective Management of Animal Wastes in La Reunion

To investigate the collective management of livestock farming wastes on La Reunion Island, an ABM called Biomas has been developed with the Geamas platform (Courdier et al. 2002). Biomas simulates the organization of transfers of organic materials between two kinds of agents: surplus-producing farms (i.e. farms where livestock raising activity dominates) and deficit farms (i.e. predominantly crop production). The environment is represented as a network of "situated objects". Their association with the Geamas agents enable the agents to act on the environment (e.g. "crop" agents are linked to "plot" situated objects). Some situated objects like "road sections" are only related to other situated objects.

Graphs of connected situated objects allow representing itineraries. The agents in Biomas are interacting through direct exchanges of messages; they are also linked to a "Group" agent through a membership process. The "Group" agent is responsible for imposing the management constraints on all its members or individual agents by means of contracts, and implements a penalty system in the case of disregard of the regulations. The simulation scenarios are based on the constraints and regulations defined at this "Group" level.

### *19.4.8 Epidemiology*

Models developed for the spread of infectious diseases in human populations are typically implemented assuming homogeneous population mixing, without a spatial dimension, social (and network) dimension, or symptom-based behaviour. ABMS offers great potential to challenge these assumptions. Recently Ling Bian (2004) proposed a conceptual framework for individual-based spatially explicit epidemiological modelling, discussing four aspects: (1) population segments or unique individuals as the modelling unit, (2) continuous process or discrete events for disease development through time, (3) travelling wave or network dispersion for transmission of diseases in space, and (4) interactions within and between night time (at home) and day time (at work) groups. As an illustration, she compares a simple population-based model of influenza to an equivalent schematic individual-based one. This abstract model has been utilised by Dunham (2005) to develop a generic ABMS tool. Recently, the fear of bioterrorism has also stimulated intensive studies in the USA, see for instance BioWar, developed by Carley et al. (2006) to simulate Anthrax and smallpox attacks on the scale of a city.

#### 19.4.8.1 Bovine Leukemia

Two methodologies (system dynamics and agent-based) used to simulate the spread of a viral disease (Bovine Leukemia) are compared (Bagni et al. 2002). The purpose is, through "what-if" analysis, to assess the system's behaviour under various conditions and to evaluate alternative sanitary policies. Based from the same set of Unified Modelling Language (UML) diagrams, Vensim and Swarm are the two platforms that have been used to implement the conceptual model. The environment represents in an abstract way a dairy farm segmented into sectors. "Cow" and "Farm sector" are the two types of autonomous agents in this model. The integration at the farm level is directly achieved through the "model swarm". Scenarios focus particularly on the number of cows detected as positive at sanitary controls (as opposed to the total number of infected cows).

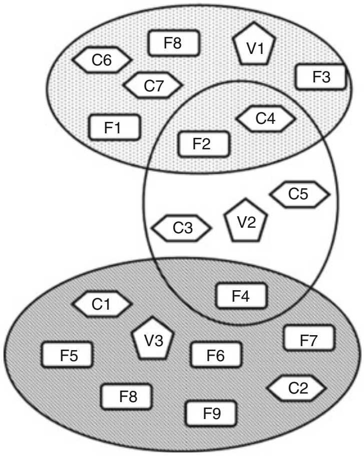



**Fig. 19.4** The representation of space as clustering of three kinds of "location" agents: village (*pentagons*), cocoa plantations (*hexagons*) and forest (*rectangles*) (Muller et al. 2004)

#### 19.4.8.2 Malaria in Haiti

To assess the impact of education on malaria healthcare in Haiti, an ABM with a realistic representation of Haiti has been designed (Rateb et al. 2005). The environment is set as a raster-grid with cells characterized by land-covers (sea, road, land, mountain, city, school, and hospital) associated with specific contamination probabilities (this is how mosquitoes are represented in the model). Apart from an epidemiological status, autonomous agents (representing individual people) are characterized by a mobility capability and an education score which value corresponds to the time agents take to attribute existing symptoms to malaria and therefore to go to hospital. Implemented in StarLogo, three scenarios based on the number of schools and hospitals have been discussed.

#### 19.4.8.3 Sleeping Sickness in Cameroon

To understand the spread of Human African Trypanosomiasis, and ultimately to elaborate a tool to evaluate risk and test control strategies, an ABM has been developed with the MadKit platform and tested with data from one village in Southern Cameroon (Muller et al. 2004). The space is not explicitly represented in this model. This is due to the meta-model associated with the MadKit platform: the system under study has to be described through "agent-group-role" interactions (Ferber and Gutknecht 1998). Hence, surprisingly, locations can only be depicted as agents here (see Fig. 19.4). They are characterized by a proportional surface area and a number of animals.

Location agents, as "groups", are responsible for "enrolling" tsetse and human agents that will, as members of the same group, be able to interact through the sending of "bite" messages. The probability for a human agent to be bitten is inversely proportional to the number of animals. Simulation scenarios are based on the organisation of space, to investigate the effect of the size and number of transmission areas.

### *19.4.9 General Considerations*

To fill in the table presented in the Appendix from the description of the models found in publications was not always easy. This is partly due to the fact that the elements to be detailed in the columns of the table require further refinements and more precise definitions. But this can also be attributed to the heterogeneity in the way model contents are detailed by the authors. The lack of a general framework to document such kind of models is patent, and all designers of ABM should become aware and refer to the framework proposed by Grimm et al. (2006). Even when the code of the implemented model is published (in appendices of articles or on a website), it is quite challenging and time-consuming to dive into it to retrieve specific information. This difficulty has triggered a bias: we have tended to select the applications we know better. As the co-authors of this chapter all belong to the same scientific network, the representativeness of the selected case studies may be questioned. This kind of task – a systematic survey based on a set of unambiguously defined characteristics – should be undertaken at the whole scientific community level, in a continuous way. Ideally, it should use an effective tool for mass collaborative authoring like wiki.

The environment, abstract or realistic, is most often represented as a raster grid. The spatial resolution, when it makes sense to define it precisely (for realistic simulated landscapes), is always clearly related to a key-characteristic of one of the model's components.

The number of applications with interactions involving the collective level is rather low. This does not necessarily imply cognitive agents. In the model of Bousquet et al. (2001), for instance, the collective level is related to the kinship structure of the small antelopes' population; when a young individual becomes mature, it leaves the parental habitat territory and starts to move around to look for a potential partner to establish a new family group in an unoccupied habitat. The group left by the young adult is affected in such a way that the reproduction can be activated again.

In our review, theoretical case studies are less numerous than empirical case studies. The prevalence of theoretical case studies is only significant for the LUCC category. It suggests that the proportion of empirical applications of ABMS is gaining ground compared to theoretical and abstract contributions. As analyzed by Janssen and Ostrom (2006), this could be explained by the fact that theoretical models, more frequent at the beginning, have demonstrated that ABMS can provide novel insights to scientific inquiry. The increased availability of more and more relevant ecological and socio-economics data then paved the way to the rise of empirically-based ABMS.

## 19.5 Why ABMS Is More and More Applied to Environmental Management

If ABMS is becoming more and more popular in environmental modelling, it is mainly because it demonstrates a potential to overcome the limitations of other kinds of models to take into account elements and processes that can hardly be ignored to consider the underlying research questions. Another aspect has to be stressed: ABMS is structurally an integrative modelling approach. It can easily be expressed with other modelling formalisms and tools. Additionally, to the evidential use of CA as a way to represent the space in ABMS applications dealing with environmental management, several other fruitful associations with complementary tools (GIS to handle spatial requests, Linear Programming modules directly used by agents to perform maximization of utility functions, etc.) have already been explored. Beyond technical aspects, ABMS can also be seen as a methodological step of a wider approach, like in companion modelling when it is jointly used with role-playing games to allow stakeholders' participation in the design of the tools used during participatory simulation workshops (Bousquet et al. 2002).

### *19.5.1 Getting Rid of Empirically Implausible Assumptions*

In ecology, traditional general population models are assuming that: (1) individuals are identical; (2) the interaction between individuals is global; (3) the spatial distribution of individuals is uniform. Required to ensure analytical tractability, these overly simplified assumptions significantly limit the usefulness of such population-based approaches. The assumption of "perfect mixing" on which population-based modelling approaches rely (two individuals randomly picked can be inter-changed) is only valid when the environment is homogeneous or when all individuals facing the same environmental conditions react in exactly the same way. One way to account for heterogeneity is to define subpopulations as classes of similar individuals (for instance based on their age). Then a distribution function of the individual states is sufficient. But when interactions between individuals are depending on the local configuration of the environment (including the other individuals), the spatial heterogeneity and the inter-individual variability (two key drivers of evolution) can not be left out anymore. Spatially-explicit individual-based models (IBM) allow representing any kind of details critical to the system under study, thus relaxing assumptions distorting the reality in an unacceptable manner. This is the main reason why for more than two decades now (DeAngelis and Gross 1992; Huston et al. 1988; Judson 1994; Kawata and Toquenaga 1994), IBM has been more and more widely used in ecological modelling (for a recent guideline to make IBM more coherent and effective, see Grimm and Railsback 2005).

When it comes to include human decision-making processes into models, the standard way consists in assuming that all individuals equally informed (perfect information sharing) exhibit a standard behaviour based on rationality to achieve optimization. It is well known that renewable resources management addresses self-referential situations. The success of an individual strategy highly depends on the ability to "best-guess" what the other individuals may do over time (Batten 2007). This is closely related to the notion of "representation" defined by Rouchier et al. (2000) as the understanding an agent has of what it perceives, and that enables it to evaluate and then to choose the actions it can undertake on its environment. Do all agents agree, or do they have very different approaches to the same object or agent? One of the main interests of ABMS is to offer the possibility to explore the second option.

More generally, ABMS is a valid technical methodology to take into account heterogeneity in parameter values and in behaviours. In abstract and theoretical ABMS, standard statistical distribution functions can be used to assign particular parameter values to the different instances of agents created from the same class. Railsback et al. (2006) have compared how the main generic ABMS simulation platforms handle the initialization of one attribute of a population of agents from a random normal distribution (see version #14 of their benchmark "StupidModel"). For more realistic applications of ABMS, Berger and Schreinemachers (2006) recently introduced a straightforward approach to empirical parameterization using a common sampling frame to randomly select observation units for both biophysical measurements and socioeconomic surveys. The heterogeneity in behaviours is usually considered with each agent having to select one behavioural module from a set of existing ones. From a conceptual design point of view, heterogeneity of behaviours is easier to represent with a hierarchy of classes. Subclasses of a generic agent class are a proper design when a given agent does not update its behaviour over time. To account for such an adaptive ability, the agent class has to be linked to a hierarchy of behaviours, as shown in Fig. 19.1. Beyond selecting an alternative out of a predefined set of options, it is even possible to define innovative agents equipped with some evolutionary programming to drive the creation of new behavioural patterns by recombining elementary behavioural components.

### 19.5.2 Dealing with Multiple Nested Levels

The seminal paper of Simon (1973) envisions hierarchical organizations as adaptive structures and not only as top-down sequences of authoritative control. This view was instilled in ecology by Allen and Starr (1982), who promoted the idea that biotic and abiotic processes at work in ecosystems are developing mutually reinforcing relationships over distinct ranges of scales. Each level, made from components interacting at the same time scale, communicates some information to the next higher and slower level. Reciprocally, any level can contribute to maintain the stability of faster and smaller levels. In the field of environmental management, both social and biophysical systems are characterized by hierarchical, nested structures. For example, family members interact to form a household, which may interact with other households in a village through political and economic institutions. Populations formed of individual species members aggregate to form communities, which, in turn, collectively define ecosystems. Holling (2001) nicely illustrates this with two mirroring examples: on one side the components of the boreal forest represented over time and space scales (from needle to landscape); on the other side the institutional hierarchy of rule sets (from the decisions of small groups of individuals to constitution and culture) represented along dimensions of the number of people involved and the turnover times.

These ideas have been conceptualized to frame the emerging paradigm of "panarchy" (Gunderson and Holling 2002): the hierarchical structure of socio-ecological systems is exhibiting never-ending cycles of growth, accumulation, restructuring and renewal. The key-concepts of this heuristic model are undoubtedly expanding the theoretical understanding of environmental management. What is their concrete contribution to the evolution of ecological modelling? To what extent can ABMS claim to represent them in a better way than other kinds of models?

The aggregation between hierarchical levels is very difficult to model in a purely analytical or statistical framework. In ecology, aggregation methods are applicable for models involving two levels of organisation (individual and population) and their corresponding time scales (fast and slow) to reduce the dimension of the initial dynamical system to an aggregated one governing few global variables evolving at the slow time scale (Auger et al. 2000). The reverse way is much more difficult to integrate to models. How to account for the influence of changes at the global level on transitions at the microscopic level? Moreover, how to simulate both ways simultaneously at work? The main challenge deals with the coordination and scheduling of the different processes running at different levels: at the collective level, explicit decisions about temporarily giving back the control to lower-level component entities, and conversely decisions from lower-level entities to create a group and to give the control to it. In the scientific community of ABMS, these ideas have directly inspired the production of conceptual organizational meta-models like Aalaadin (Ferber and Gutknecht 1998), specific features in generic simulation platforms like the threaded scheduling of agents in Swarm, as well as applications like simulating hydrological processes (runoff, erosion and infiltration on heterogeneous soil surfaces) with "waterball", pond and river agents (Servat et al. 1998).

### *19.5.3 Beyond Decision Support Systems: Exploring New Dimensions in the Way to Use Models*

As they represent complex adaptive systems which are unpredictable as a whole, ABMS applied to environmental management should caution about the large uncertainties related to their predictive abilities (Bradbury 2002). Still, empirically-based ABMS can be used as a decision-support system, for instance to assist policymakers in prioritizing and targeting alternative policy interventions, as Berger et al. (2006) did in Uganda and Chile. Nevertheless, when multiple perceptions of the reality coexist, the statement “everything is defined by the reality of the observed phenomena” can be questioned. Therefore, ABMS in the field of ecological management should take some distance with the positivist posture that designates the scientific knowledge as the only authentic one. Relating empirical observations of phenomena to each other, in a way which is consistent with fundamental theory, phenomenological modelling is a means to represent the phenomena in a formalized and synthetic way. Descriptive rather than explanatory, this approach does not truly gain understanding of the phenomena, but can claim, in simple cases, to predict them. Parker et al. (2003) refer to these two distinct explanatory and descriptive approaches to clarify the potential roles of ABMS in LUCC. Anyway, the general rule “the more realistic the application, the more descriptive the approach” may not necessarily always apply. Explanatory goals can be assigned to models closely related to a real situation as well.

In contrast to the positivist approach, the constructivist approach refers to “constructed” knowledge, contingent on human perception and social experience and not necessarily reflecting any external “transcendent” realities. Starting “from scratch” to collectively design a model is a straightforward implementation of the constructivist approach. Among scientists, it will integrate within and between disciplines. By involving stakeholders, instead of showing them a simplification of their knowledge, the collective design of the model is seeking a mutual recognition of everyone’s representation. In such a context, ABMS is more a communication platform to facilitate collective learning than a turnkey itinerary for piloting renewable resources management (Bousquet et al. 1999; ComMod 2003; Etienne et al. 2003; Gurung et al. 2006).

## 19.6 Drawbacks, Pitfalls, and Remaining Challenges

### *19.6.1 Verification and Validation of ABMS*

This is a problem challenging ABMS in general that is addressed in Chaps. 6 (Galán et al. 2013) and 8 (David 2013) of this book. In the field of ecological management, as in other fields of applications, some authors claim to intentionally bridle the development of their agent-based model to design a strict equivalent to a mathematical equations-based model, as a means to verify it (Brown et al. 2004). The same process has been tested with mathematical representations of discrete distributed systems like Petri nets (Bakam et al. 2001).

### 19.6.2 Capturing the Meta-rules Governing the Adoption of Alternative Strategies

Nowadays, a set of tested and reliable tools and methods is available to better understand decision rules of actors and integrate them in computer agents (Janssen and Ostrom 2006). What remains much more challenging is to capture the rules governing the changes in agents' behaviours. An example of such a kind of "meta-rule" is the "evaluateAttitude" of the Collector agent (see Fig. 19.1) defined by (Thébaud and Locatelli 2001). In such a stylized model, the meta-rule is simply based on a threshold value of a specific parameter (the size of the pile of collected driftwood). When it comes to making the rules explicit to governing the changes of behavioural rules of human beings in real situations, methods are still weak. The meta-rules, if they exist, that control changes of strategies are difficult to grasp and elicit, and by consequence to implement in an empirical-based ABM. One reason is the time scale which is greater for these meta-rules than for decision rules, and which makes direct observation and verification harder to carry.

### 19.6.3 Improving the Representation of Space

Representing space with a CA, by far the most frequent way in current applications of ABMS in environmental management, is easy. But, as recently pointed out by Bithell and Macmillan (2007), imposition of a fixed grid upon the dynamics may cause important phenomena to be misrepresented when interactions between individuals are mediated by their size, and may become too consumed by computer resources when the system scale exceeds the size of individuals by a large factor. How to handle discrete spatial data that is potentially completely unstructured, and how to discover patterns of neighbourhood relationships between the discrete individuals within it? New directions like particle-in-cell are suggested.

In the next few years, we can also expect more applications based on autonomous agents moving over a GIS-based model of the landscape, with rendering algorithms determining what an individual agent is able to "see". Already used to simulate recreational activities (see for instance Bishop and Gimblett 2000), behavioural responses to 3D virtual landscape may become more common.

## Further Reading

1. The special issue of JASSS in 2001[1] on "ABM, Game Theory and Natural Resource Management issues" presents a set of papers selected from a workshop held in Montpellier in March 2000, most of them dealing with collective decision-making processes in the field of natural resource management and environment.
2. Gimblett (2002) is a book on integrating GIS and ABM, derived from a workshop held in March 1998 at the Santa Fe Institute. It provides contributions from computer scientists, geographers, landscape architects, biologists, anthropologists, social scientists and ecologists focusing on spatially explicit simulation modelling with agents.
3. Janssen (2002) provides a state-of-the-art review of the theory and application of multi-agent systems for ecosystem management and addresses a number of important topics including the participatory use of models. For a detailed review of this book see Terna (2005).
4. Paredes and Iglesias (2008) advocate why agent based simulations provide a new and exciting avenue for natural resource planning and management: researches and advisers can compare and explore alternative scenarios and institutional arrangements to evaluate the consequences of policy actions in terms of economic, social and ecological impacts. But as a new field it demands from the modellers a great deal of creativeness, expertise and "wise choice", as the papers collected in this book show.

## Appendix

### *Topic and Issue*

When multiple topics are covered by a case study, the first in the list indicates the one we used to classify it. Within each topic we have tried to order the case studies from the more abstract and theoretical ones to the more realistic ones. This information can be retrieved from the issue: only case studies representing a real system mention a geographical location.

[1] http://jasss.soc.surrey.ac.uk/4/2/contents.html

| Publications | Model name | Topic | Issue | Environment | Agents | Software |
|---|---|---|---|---|---|---|
| Polhill et al. 2001 | FEARLUS | LUCC | Land-use and land ownership dynamics | Raster 7,7 region | Land manager (*49*) HeB (10) Ie | Swarm |
| Brown et al. 2004 | | LUCC | Greenbelt to control residential development | Raster *X*,80 suburb | Resident (*10*) Ho Ie R | Swarm |
| Castella et al. 2005a | SAMBA (Generic) | LUCC | LU changes under decollectivization in north Vietnam | Raster 50,50 (25 Ha) commune level (Abstract) | Household (50) HeP Ie R | Cormas |
| Castella et al. 2005b | SAMBA-GIS (More realistic) | LUCC | LU changes under decollectivization in north Vietnam | Raster regional level (More realistic) | Household (x) HeP Ie R (x = nb household in each village) | Cormas ArcView |
| D'Aquino et al. 2003 | SelfCormas | LUCC; agriculture; livestock | Competing rangeland and rice cropping land-uses in Senegal | Raster 20,20 (5Ha) village | Farmer (20) *HeB*(2) Ie R | Cormas |
| Etienne et al. 2003 | Mejan | LUCC; livestock; forestry; wildlife | Pine encroachment on original landscapes in France | Raster 23793 (1 ha) plateau | Farmer (40) *HeB* Ie,Ii,Ic C forester (2) HeB Ie,Ii,Ic R national park (1) Ho Ie,Ii,Ic C | Cormas MapInfo |
| Mathevet et al. 2003a | GEMACE | LUCC; agriculture; wildlife | Competing hunting and hunting activities in Camargue (South of France) | Raster region | Farmer Ie, Ii hunting manager Ie, Ii | Cormas |
| Barreteau and Bousquet 2000; Barreteau et al. 2004 | SHADOC | Water; agriculture | Viability of irrigated systems in the Senegal river valley | | Farmer | |
| Le Bars et al. 2005 | Manga | Water | | | Farmer (n) Ie,Ii C water supplier (1) Ic C | |
| Feuillette et al. 2003 | Sinuse | Water; agriculture | Water table level | Raster 2400 (1Ha) watershed | Farmer HeP Ie,Ii,Ic | Cormas |
| Lansing and Kremer 1993; Janssen 2007 | | Water; agriculture | Water management and water temple networks in Bali | Network 172 watershed | Village (172) HeB(49) Ie,Ic R | |
| Raj Gurung et al. 2006 | Limbukha | Water management | Negotiation of irrigation water sharing between two Bhutanese communities | Grid 8,13 (10.4 Ha) (Abstract) Village | Farmer (12) *HeB* Ii | Cormas |
| Hoffmann et al. 2002 | LUCIM | Forestry; LUCC | Deforestation and afforestation in Indiana USA | Raster 100 state | Farmer (10) HeB Ie R | |
| Purnomo and Guizol 2006 | | Forestry | Forest plantation comanagement in Indonesia | Raster 50,50 () Forest massif | Developer Ie,Ii,Ic Smallholder Ie,Ii,Ic C Broker Ie,Ii,Ic government Ie,Ii,Ic | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Mathevet et al. 2003b | ReedSim | Wildlife | Water management in Mediterranean reedbeds | | | Cormas |
| An et al. 2005 | | Wildlife | Impact of the growing rural population on the forests and panda habitat in China | | | |
| Bousquet et al. 2001 | Djemiong | Wildlife | Traditional hunting of small antelopes in Cameroon | Raster 2042 (4Ha) village | Hunter (90) HeP Ie,Ic R antelop *(7350) HeB*(3) Ie,Ic R | Cormas |
| Anwar et al. 2007 | | Wildlife | Interactions between whale-watching boats and whales in the St. Lawrence estuary | Raster | Boat C whale R | Repast |
| Elliston and Beare 2006 | | Agriculture; epidemiology | Agricultural pest and disease incursions in Australia | | | Cormas |
| Happe et al. 2006 | AgriPolis | Agriculture; lifestock | Policy impact on structural changes of W. Europe farms | Raster 73439 (1Ha) region | Farms (2869) HeP Ic | |
| Barnaud et al. (forthcoming) | MaeSalaep 2.2 | Agriculture; LUCC | LU strategies in transitional swidden agricultural systems, Thailand | Raster (300 Ha) (Realistic) village catchment | Farmer (12) HeP Ii R credit sources (3) | Cormas |
| Gross et al. 2006 | | Lifestock | Evaluation of behaviours of rangeland systems in Australia | | Entreprise C government | |
| Courdier et al. 2002 | Biomas | Livestock; agriculture | Collective management of animal wastes in La Reunion | Network region | Livestock farm (48) HeP Ie, Ii,Ic crop farm (59) HeP Ie,Ii,Ic shipping agent (34) HeP Ii | Geamas |
| Bagni et al. 2002 | | Epidemiology; livestock | Evaluation of sanitary policies to control the spread of a viral pathology of bovines | Abstract farm | Farm sector Cow | Swarm |
| Rateb et al. 2005 | | Epidemiology | Impact of education on malaria healthcare in Haiti | Raster country | Inhabitant *HeP* R | StarLogo |
| Muller et al. 2004 | | Epidemiology; agriculture; livestock | Risk and control strategies of African Trypanosomiasis in Southern Cameroon | Network village | Villager Ic R | MadKit |

### *Environment*

- First line: mode of representation, with the general following pattern:

$$[\text{none; network; raster, vector}]\ N(x)$$

  N indicates the number of elementary spatial entities (nodes of network, cells or polygons), when raster mode, N is given as number of lines *x* number of columns, unless some cells have been discarded from the rectangular grid because they were out of bound (then only the total number is given), and (x) indicates the spatial resolution.
- Second line: level of organization at which the issue is considered (for instance village; biophysical entity (watershed, forest massif, plateau, etc.); city; conurbation; province, country, etc.)

### *Agents*

One line per type of agent (the practical definition given in this paper applies, regardless of the terminology used by the authors). The general pattern of information looks like:

$$\text{name}(x)[\text{Ho; HeP; HeB}(y)][\text{Ie; Ii; Ic}]\ [\text{R; C}]$$

- (x) indicates the number of instances defined when initializing a standard scenario, italic mentions that this initial number change during simulation.
- When $x > 1$, to account for the heterogeneity of the population of agents, we propose the following coding: "Ho" stands for a homogeneous population (identical agents), "He" stands for a heterogeneous population. "HeP" indicates that the heterogeneity lies only in parameter values, while "HeB" indicates that the heterogeneity lies in behaviours. In such a case, each agent is equipped with one behavioural module selected from a set of (y) existing ones. Italic points out adaptive agents updating either parameter value (*HeP*) or behaviour (*HeB*) during simulation.
- [Ie, Ii, Ic] indicates the nature of relationships as defined in the text and shown in Fig. 19.3.
- [R; C] indicates if agents are clearly either reactive or cognitive

## References

Abel T (1998) Complex adaptive systems, evolutionism, and ecology within anthropology: interdisciplinary research for understanding cultural and ecological dynamics. Georgia J Ecol Anthropol 2:6–29

Abrami G (2004) Niveaux d'organisation dans la modélisation multi-agent pour la gestion de ressources renouvelables. Application à la mise en oeuvre de règles collectives de gestion de l'eau agricole dans la basse-vallée de la Drôme. Ecole Nationale du Génie Rural, des Eaux et Forêts, Montpellier, France

Allen TFH, Starr TB (1982) Hierarchy: perspectives for ecological complexity. University of Chicago Press, Chicago

An L, Linderman M, Qi J, Shortridge A, Liu J (2005) Exploring complexity in a human–environment system: an agent-based spatial model for multidisciplinary and multiscale integration. Ann Assoc Am Geogr 95(1):54–79

Anwar SM, Jeanneret C, Parrott L, Marceau D (2007) Conceptualization and implementation of a multi-agent model to simulate whale-watching tours in the St. Lawrence estuary in Quebec, Canada. Environ Model Software 22(12):1775–1787

Auger P, Charles S, Viala M, Poggiale J-C (2000) Aggregation and emergence in ecological modelling: integration of ecological levels. Ecol Model 127:11–20

Bagni R, Berchi R, Cariello P (2002) A comparison of simulation models applied to epidemics. J Artif Soc Soc Simul 5(3), 5

Bah A, Touré I, Le Page C, Ickowicz A, Diop AT (2006) An agent-based model to understand the multiple uses of land and resources around drilling in Sahel. Math Comp Model 44(5–6):513–534

Bakam I, Kordon F, Le Page C, Bousquet F (2001) Formalization of a multi-agent model using colored petri nets for the study of an hunting management system. In: Rash JL, Rouff C, Truszkowski W, Gordon DF, Hinchey MG (eds) Formal approaches to agent-based systems, first international workshop, FAABS 2000 (Lecture notes in computer science, 1871). Springer, Greenbelt, pp 123–132

Balmann A (1997) Farm based modelling of regional structural change: a cellular automata approach. Eur Rev Agric Econ 24(1):85–108

Barnaud C, Promburom T, Trébuil G, Bousquet F (2007) An evolving simulation and gaming process for adaptive resource management in the highlands of northern Thailand. Simulat Gaming 38(3):398–420

Barreteau O, Bousquet F (2000) SHADOC: a multi-agent model to tackle viability of irrigated systems. Ann Oper Res 94:139–162

Barreteau O, Bousquet F, Millier C, Weber J (2004) Suitability of multi-agent simulations to study irrigated system viability: application to case studies in the Senegal River valley. Agr Syst 80:255–275

Batten D (2007) Are some human ecosystems self-defeating? Environ Model Software 22:649–655

Becu N, Perez P, Walker A, Barreteau O, Le Page C (2003) Agent based simulation of a small catchment water management in northern Thailand: description of the CATCHSCAPE model. Ecol Model 170(2–3):319–331

Berger T (2001) Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. Agr Econ 25:245–260

Berger T, Schreinemachers P (2006) Creating agents and landscapes for multiagent systems from random samples. Ecol Soc 11(2):19, http://www.ecologyandsociety.org/vol11/iss2/art19/

Berger T, Schreinemachers P, Woelcke J (2006) Multi-agent simulation for the targeting of development policies in less-favored areas. Agr Syst 88:28–43

Berkes F, Folke C (eds) (1998) Linking ecological and social systems: management practices and social mechanisms for building resilience. Cambridge University Press, Cambridge, UK

Bian L (2004) A conceptual framework for an individual-based spatially explicit epidemiological model. Environ Plann B Plann Des 31(3):381–395

Bishop I, Gimblett R (2000) Management of recreational areas: GIS, autonomous agents, and virtual reality. Environ Plann B Plann Des 27(3):423–435

Bithell M, Macmillan WD (2007) Escape from the cell: spatially explicit modelling with and without grids. Ecol Model 200:59–78

Bonaudo T, Bommel P, Tourrand JF (2005) Modelling the pioneers fronts of the transamazon highway region. In: Proceedings of CABM-HEMA-SMAGET 2005, Joint conference on multi-agent modeling for environmental management, Bourg St Maurice, Les Arcs, 21–25 Mar 2005. http://smaget.lyon.cemagref.fr/contenu/SMAGET%20proc/PAPERS/BonaudoBommelTourrand.pdf

Bousquet F (2001) Modélisation d'accompagnement. Simulations multi-agents et gestion des ressources naturelles et renouvelables (Unpublished Habilitation à Diriger des Recherches). Lyon I University, Lyon

Bousquet F, Le Page C (2004) Multi-agent simulations and ecosystem management: a review. Ecol Model 176:313–332

Bousquet F, Barreteau O, Le Page C, Mullon C, Weber J (1999) An environmental modelling approach: the use of multi-agents simulations. In: Blasco F, Weill A (eds) Advances in environmental and ecological modelling. Elsevier, Paris, pp 113–122

Bousquet F, Le Page C, Bakam I, Takforyan A (2001) Multiagent simulations of hunting wild meat in a village in eastern Cameroon. Ecol Model 138:331–346

Bousquet F et al (2002) Multi-agent systems and role games: collective learning processes for ecosystem management. In: Janssen MA (ed) Complexity and ecosystem management: the theory and practice of multi-agent systems. Edward Elgar, Cheltenham, pp 248–285

Bousquet F, Trébuil G, Hardy B (eds) (2005) Companion modeling and multi-agent systems for integrated natural resources management in Asia. International Rice Research Institute, Los Baños

Bradbury R (2002) Futures, prediction and other foolishness. In: Janssen MA (ed) Complexity and ecosystem management: the theory and practice of multi-agent systems. Edward Elgar, Cheltenham, pp 48–62

Brown DG, Page SE, Riolo R, Rand W (2004) Agent-based and analytical modeling to evaluate the effectiveness of greenbelts. Environ Model Software 19:1097–1109

Caplat P, Lepart J, Marty P (2006) Landscape patterns and agriculture: modelling the long-term effects of human practices on Pinus sylvestris spatial dynamics (Causse Mejean, France). Landscape Ecol 21(5):657–670

Carley KM et al (2006) BioWar: scalable agent-based model of bioattacks. IEEE T Syst Man Cy A: Syst Hum 36(2):252–265

Carpenter S, Brock W, Hanson P (1999) Ecological and social dynamics in simple models of ecosystem management. Conserv Ecol 3(2), 4

Castella J-C, Verburg PH (2007) Combination of process-oriented and pattern-oriented models of land use change in a mountain area of Vietnam. Ecol Model 202(3–4):410–420

Castella J-C, Boissau S, Trung TN, Quang DD (2005a) Agrarian transition and lowland–upland interactions in mountain areas in northern Vietnam: application of a multi-agent simulation model. Agr Syst 86:312–332

Castella J-C, Trung TN, Boissau S (2005b) Participatory simulation of land-use changes in the northern mountains of Vietnam: the combined use of an agent-based model, a role-playing game, and a geographic information system. Ecol Soc 10(1), 27

ComMod (2003) Our companion modelling approach. J Artif Soc Soc Simul 6(2). http://jasss.soc.surrey.ac.uk/6/2/1.html

Courdier R, Guerrin F, Andriamasinoro FH, Paillat J-M (2002) Agent-based simulation of complex systems: application to collective management of animal wastes. J Artif Soc Soc Simul 5(3). http://jasss.soc.surrey.ac.uk/5/3/4.html

D'Aquino P, Le Page C, Bousquet F, Bah A (2003) Using self-designed role-playing games and a multi-agent system to empower a local decision-making process for land use management: the

SelfCormas experiment in Senegal. J Artif Soc Soc Simul 6(3). http://jasss.soc.surrey.ac.uk/6/3/5.html
David N (2013) Validating simulations. Chapter 8 in this volume
Deadman P, Robinson D, Moran E, Brondízio E (2004) Colonist household decision making and land-use change in the Amazon rainforest: an agent-based simulation. Environ Plann B Plann Des 31(5):693–709
DeAngelis DL, Gross LJ (eds) (1992) Individual-based models and approaches in ecology. Chapman and Hall, New York
Dunham JB (2005) An agent-based spatially explicit epidemiological model in MASON. J Artif Soc Soc Simulat 9(1). http://jasss.soc.surrey.ac.uk/9/1/3.html
Elliston L, Beare S (2006) Managing agricultural pest and disease incursions: an application of agent-based modelling. In: Perez P, Batten D (eds) Complex science for a complex world: exploring human ecosystems with agents. ANU E Press, Canberra, pp 177–189
Epstein JM, Axtell RL (1996) Growing artificial societies: social science from the bottom up. Brookings Institution Press, Washington, DC
Etienne M (ed) (2011) Companion modelling: a participatory approach to support sustainable development. Quae, Versailles
Etienne M, Le Page C, Cohen M (2003) A step-by-step approach to building land management scenarios based on multiple viewpoints on multi-agent system simulations. J Artif Soc Soc Simul 6(2). http://jasss.soc.surrey.ac.uk/6/2/2.html
Evans TP, Kelley H (2004) Multi-scale analysis of a household level agent-based model of landcover change. J Environ Manage 72:57–72
Ferber J, Gutknecht O (1998) A meta-model for the analysis and design of organizations in multi-agent systems. In: Demazeau Y (ed) Proceedings of the 3rd international conference on multi-agent systems (ICMAS'98), Cité des sciences – La Villette, Paris, 4–7 July 1998. IEEE, Los Alamos, pp 128–135
Feuillette S, Bousquet F, Le Goulven P (2003) SINUSE: a multi-agent model to negotiate water demand management on a free access water table. Environ Model Software 18:413–427
Galán J et al. (2013) Checking simulations: detecting and avoiding errors and artefacts. Chapter 6 in this volume
Galvin KA, Thornton PK, Roque de Pinho J, Sunderland J, Boone RB (2006) Integrated modeling and its potential for resolving conflicts between conservation and people in the rangelands of East Africa. Hum Ecol 34(2):155–183
Gimblett R (ed) (2002) Integrating geographic information systems and agent-based modeling techniques for understanding social and ecological processes (Santa Fe Institute studies in the sciences of complexity). Oxford University Press, Oxford
Grimm V, Railsback SF (2005) Individual-based modeling and ecology. Princeton University Press, Princeton
Grimm V et al (2006) A standard protocol for describing individual-based and agent-based models. Ecol Model 198:115–126
Gross JE, McAllister RRJ, Abel N, Stafford Smith DM, Maru Y (2006) Australian rangelands as complex adaptive systems: a conceptual model and preliminary results. Environ Model Software 21(9):1264–1272
Gunderson LH, Holling CS (2002) Panarchy: understanding transformations in human and natural systems. Island Press, Washington, DC
Gurung TR, Bousquet F, Trébuil G (2006) Companion modeling, conflict resolution, and institution building: sharing irrigation water in the Lingmuteychu watershed, Bhutan. Ecol Soc 11(2), 36
Haffner Y, Gramel S (2001) Modelling strategies for water supply companies to deal with nitrate pollution. J Artif Soc Soc Simul 4(3). http://jasss.soc.surrey.ac.uk/4/3/11.html
Hales D, Rouchier J, Edmonds B (2003) Model-to-model analysis. J Artif Soc Soc Simul 6(4). http://jasss.soc.surrey.ac.uk/6/4/5.html

Happe K, Kellermann K, Balmann A (2006) Agent-based analysis of agricultural policies: an illustration of the agricultural policy simulator AgriPoliS, its adaptation and behavior. Ecol Soc 11(1), 49

Hare M, Deadman P (2004) Further towards a taxonomy of agent-based simulation models in environmental management. Math Comput Simul 64:25–40

Hemelrijk C (2013) Simulating complexity of animal social behaviour. Chapter 22 in this volume

Hoffmann M, Kelley H, Evans T (2002) Simulating land-cover change in south-central Indiana: an agent-based model of deforestation and afforestation. In: Janssen MA (ed) Complexity and ecosystem management: the theory and practice of multi-agent systems. Edward Elgar, Cheltenham, pp 218–247

Holling CS (1986) The resilience of terrestrial ecosystems; local surprise and global change. In: Clark WC, Munn RE (eds) Sustainable development of the biosphere. Cambridge University Press, Cambridge, pp 292–317

Holling CS (2001) Understanding the complexity of economic, ecological, and social systems. Ecosystems 4:390–405

Holling CS (2004) From complex regions to complex worlds. Ecol Soc 9(1), 11

Huigen MGA (2004) First principles of the MameLuke multi-actor modelling framework for land use change, illustrated with a Philippine case study. J Environ Manage 72:5–21

Huigen MGA, Overmars KP, de Groot WT (2006) Multiactor modeling of settling decisions and behavior in the San Mariano watershed, the Philippines: a first application with the MameLuke framework. Ecol Soc 11(2), 33

Huston M, DeAngelis DL, Post W (1988) New computer models unify ecological theory. Bioscience 38(10):682–691

Janssen MA (2001) An exploratory integrated model to assess management of lake eutrophication. Ecol Model 140:111–124

Janssen MA (ed) (2002) Complexity and ecosystem management: the theory and practice of multi-agent systems. Edward Elgar, Cheltenham

Janssen MA (2007) Coordination in irrigation systems: an analysis of the Lansing–Kremer model of Bali. Agr Syst 93:170–190

Janssen MA, Ostrom E (2006) Empirically based, agent-based models. Ecol Soc 11(2)

Janssen MA, Walker BH, Langridge J, Abel N (2000) An adaptive agent model for analysing co-evolution of management and policies in a complex rangeland system. Ecol Model 131 (2–3):249–268

Janssen MA, Anderies JM, Stafford Smith M, Walker BH (2002) Implications of spatial heterogeneity of grazing pressure on the resilience of rangelands. In: Janssen MA (ed) Complexity and ecosystem management: the theory and practice of multi-agent systems. Edward Elgar, Cheltenham, pp 103–126

Jepsen JU, Topping CJ, Odderskær P, Andersen PN (2005) Evaluating consequences of land-use strategies on wildlife populations using multiple-species predictive scenarios. Agr Ecosyst Environ 105:581–594

Judson OP (1994) The rise of the individual-based model in ecology. Trends Ecol Evol 9(1):9–14

Kawata M, Toquenaga Y (1994) From artificial individuals to global patterns. Trends Ecol Evol 9 (11):417–421

Kinzig AP et al (2006) Resilience and regime shifts: assessing cascading effects. Ecol Soc 11(1), 20

Krywkow J, Valkering P, Rotmans J, van der Veen A (2002) Agent-based and integrated assessment modelling for incorporating social dynamics in the management of the Meuse in the Dutch province of Limburg. In: Rizzoli AE , Jakeman AJ (eds) Integrated assessment and decision support: proceedings of the first biennial meeting of the international environmental modelling and software society iEMSs, 24–27 June 2002, University of Lugano, Switzerland, Vol 2. iEMSs, Manno, pp 263–268

Lansing JS, Kremer JN (1993) Emergent properties of Balinese water temple networks: coadaptation on a rugged fitness landscape. Am Anthropol 95(1):97–114

Le Bars M, Attonaty J-M, Ferrand N, Pinson S (2005) An agent-based simulation testing the impact of water allocation on farmers' collective behaviors. Simulation 81(3):223–235

Levin SA (1998) Ecosystems and the biosphere as complex adaptive systems. Ecosystems 1:431–436

Levin SA (1999) Towards a science of ecological management. Conserv Ecol 3(2), 6

Lim K, Deadman PJ, Moran E, Brondizio E, McCracken S (2002) Agent-based simulations of household decision making and land use change near Altamira, Brazil. In: Gimblett R (ed) Integrating geographic information systems and agent-based modeling techniques for understanding social and ecological processes (Santa Fe Institute studies in the sciences of complexity). Oxford University Press, Oxford, pp 277–310

López Paredes A, Hernández Iglesias C (eds) (2008) Agent-based modelling in natural resource management. INSISOC, Valladolid

Manson S (2005) Agent-based modeling and genetic programming for modeling land change in the southern Yucatan Peninsular region of Mexico. Agr Ecosyst Environ 111:47–62

Manson S (2006) Land use in the southern Yucatan peninsular region of Mexico: scenarios of population and institutional change. Comput Environ Urban Syst 30:230–253

Marietto MB, David N, Sichman JS, Coelho H (2003) Requirements analysis of agent-based simulation platforms: state of the art and new prospects. In: Sichman JS, Bousquet F, Davidsson P (eds) MABS'02, Proceedings of the 3rd international conference on multi-agent-based simulation (Lecture notes in artificial intelligence, 2581). Springer, Berlin, pp 125–141

Mathevet R, Bousquet F, Le Page C, Antona M (2003a) Agent-based simulations of interactions between duck population, farming decisions and leasing of hunting rights in the Camargue (Southern France). Ecol Model 165(2/3):107–126

Mathevet R, Mauchamp A, Lifran R, Poulin B, Lefebvre G (2003b) ReedSim: simulating ecological and economical dynamics of Mediterranean reedbeds. In: Proceedings of MODSIM03, International congress on modelling and simulation, integrative modelling of biophysical, social, and economic systems for resource management solutions, Townsville, 14–17 July 2003. Modelling & Simulation Society of Australia & New Zealand. http://mssanz.org.au/MODSIM03/Volume_03/B02/01_Mathevet.pdf

Matthews R (2006) The people and landscape model (PALM): towards full integration of human decision-making and biophysical simulation models. Ecol Model 194:329–343

Milner-Gulland EJ, Kerven C, Behnke R, Wright IA, Smailov A (2006) A multi-agent system model of pastoralist behaviour in Kazakhstan. Ecol Complex 3(1):23–36

Moreno N et al (2007) Biocomplexity of deforestation in the Caparo tropical forest reserve in Venezuela: an integrated multi-agent and cellular automata model. Environ Model Software 22:664–673

Muller G, Grébaut P, Gouteux J-P (2004) An agent-based model of sleeping sickness: simulation trials of a forest focus in southern Cameroon. C R Biol 327:1–11

Nute D et al (2004) NED-2: an agent-based decision support system for forest ecosystem management. Environ Model Software 19:831–843

Parker DC, Manson SM, Janssen MA, Hoffmann MJ, Deadman P (2003) Multi-agent systems for the simulation of land-use and land-cover change: a review. Ann Assoc Am Geogr 93 (2):314–337

Pepper JW, Smuts BB (2000) The evolution of cooperation in an ecological context: an agent-based model. In: Kohler TA, Gumerman GJ (eds) Dynamics in human and primate societies (Santa Fe Institute studies in the sciences of complexity). Oxford University Press, New York, pp 45–76

Perez P, Batten D (eds) (2006) Complex science for a complex world: exploring human ecosystems with agents. ANU E Press, Canberra

Polhill JG, Gotts NM, Law ANR (2001) Imitative versus non-imitative strategies in a land use simulation. Cybern Syst 32(1–2):285–307

Polhill JG, Izquierdo LR, Gotts NM (2006) What every agent-based modeller should know about floating point arithmetic. Environ Model Software 21:283–309

Purnomo H, Guizol P (2006) Simulating forest plantation co-management with a multi-agent system. Math Comput Model 44:535–552

Purnomo H, Mendoza GA, Prabhu R, Yasmi Y (2005) Developing multi-stakeholder forest management scenarios: a multi-agent system simulation approach applied in Indonesia. Forest Policy Econ 7(4):475–491

Railsback SF, Lytinen SL, Jackson SK (2006) Agent-based simulation platforms: review and development recommendations. Simulation 82(9):609–623

Rateb F, Pavard B, Bellamine-BenSaoud N, Merelo JJ, Arenas MG (2005) Modeling malaria with multi-agent systems. Int J Intell Info Technol 1(2):17–27

Rouchier J, Bousquet F, Barreteau O, Le Page C, Bonnefoy J-L (2000) Multi-agent modelling and renewable resources issues: the relevance of shared representation for interacting agents. In: Moss S, Davidsson P (eds) Multi-agent-based simulation, second international workshop, MABS 2000, Boston, July; revised and additional papers (Lecture notes in artificial intelligence, 1979). Springer, Berlin, pp 341–348

Rouchier J, Bousquet F, Requier-Desjardins M, Antona M (2001) A multi-agent model for describing transhumance in North Cameroon: comparison of different rationality to develop a routine. J Econ Dyn Control 25:527–559

Schreinemachers P, Berger T (2006) Land-use decisions in developing countries and their representation in multi-agent systems. J Land Use Sci 1(1):29–44

Servat D, Perrier E, Treuil J-P , Drogoul A (1998) When agents emerge from agents: introducing multi-scale viewpoints in multi-agent simulations. In: Sichman JS, Conte R, Gilbert N (eds) Proceedings of the first international workshop on multi-agent systems and agent-based simulation (Lecture notes in artificial intelligence, 1534). Springer, London, pp 183–198

Simon HA (1973) The organization of complex systems. In: Pattee HH (ed) Hierarchy theory: the challenge of complex systems. Braziller, New York, pp 1–27

Soulié J-C, Thébaud O (2006) Modeling fleet response in regulated fisheries: an agent-based approach. Math Comput Model 44:553–554

Sulistyawati E, Noble IR, Roderick ML (2005) A simulation model to study land use strategies in swidden agriculture systems. Agric Syst 85:271–288

Terna P (2005) Review of "complexity and ecosystem management: the theory and practice of multi-agent systems". J Artif Soc Soc Simul 8(2). http://jasss.soc.surrey.ac.uk/8/2/reviews/terna.html

Thébaud O, Locatelli B (2001) Modelling the emergence of resource-sharing conventions: an agent-based approach. J Artif Soc Soc Simul 4(2). http://jasss.soc.surrey.ac.uk/4/2/3.html

Ziervogel G, Bithell M, Washington R, Downing T (2005) Agent-based social simulation: a method for assessing the impact of seasonal climate forecast applications among smallholder farmers. Agric Syst 83:1–26

Zunga Q et al (1998) Coupler Systèmes d'Information Géographiques et Systèmes Multi-Agents pour modéliser les dynamiques de transfotmation des paysages. Le cas des dynamiques foncières de la Moyenne Vallée du Zambèze (Zimbabwe). In: Ferrand N (ed) Modèles et systèmes multi-agents pour la gestion de l'environnement et des territoires. Cemagref Editions, Clermont-Ferrand, pp 193–206

# Chapter 20
# Assessing Organisational Design

**Virginia Dignum**

**Why Read This Chapter?** To understand the requirements of an adequate formal model of organisations; in particular, to understand the main components of an organisational model and how they might be formalised. Also to be aware of how organisational change might be represented and assessed using formal models.

**Abstract** This chapter highlights the use of formal models for the analysis and change of organisational designs. In order to survive, organisations must be able to maintain a good fit with the environment, and therefore be able to determine when and how changes should take place. Current work in Organisational Theory is taken as a basis for the development of a theoretical model in which organisational performance and change measures can be defined and analyzed. In addition, this chapter presents a prototype framework for the simulation of reorganization.

## 20.1 Introduction

Organisations are complex dynamic goal-oriented processes, with outcomes dependent both on the contextual situation as on the coordinated and purposeful action of human beings in the environment. Providing organisation designers with knowledge and tools that enable them to effectively identify, assess and respond to environment changes is therefore of utmost importance for the survival of the organisation. One of the main reasons for creating organisations is to provide the means for coordination that enable the achievement of global goals. Organisations can be seen as open systems in that they have to interact with their environment in order to realize those goals (Aldrich and Pfeffer 1976). The design of organisations must consider both structural and human components. Organisational structure has

V. Dignum (✉)
Faculty Technology Policy and Management, section ICT, Delft University of Technology, Delft, The Netherlands
e-mail: m.v.dignum@tudelft.nl

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_20,

essentially two objectives (Duncan 1979): Firstly, it facilitates the flow of information within the organisation in order to reduce the uncertainty of decision making. Secondly, the structure of the organisation should integrate organisational behavior across the parts of the organisation so that it is coordinated. This raises two challenges: division of labor and coordination (Mintzberg 1993). The design of organisational structure determines the allocation of resources and people to specified tasks or purposes, and the coordination of these resources to achieve organisational goals (Galbraith 1977). Ideally, the organisation is designed to fit its environment and to provide the information and coordination needed for its strategic objectives.

Organisational design mostly adopts a multi-contingency view, which says that an organisation's design should be chosen based on the multi-dimensional characteristics of a particular environment or context. These characteristics include structural (e.g. goals, strategy and structure) and human (e.g. work processes, people, coordination and control) components. Contingency theory states that even if there is not one best way to organize, some designs perform better than others. In fact, Roberts states that the central problem of organisation design is finding a strategy and an organisation structure that are consistent with each other. This requires finding a dynamic balance in the face of strategic and organisational changes (Roberts 2004). Given that there are no exact recipes to construct the optimal organisation, organisational assessment is about evaluating a certain design and determine its appropriateness given the organisation's aims and constraints.

In order to access organisational design and performance, the fact that organisations and their members are independent entities must be considered. Taking this point of view, agents participating in organisations are not merely fulfilling some roles of an organisation to realize global objectives set out by the organisational structures. They have an independent existence (like in human organisations people have an existence independent from the organisation). Following this approach, it becomes very important to describe the exact relation between the organisational entities on the one hand and the agents on the other hand.

The above considerations provide two different motivations for this work. On the one hand, the need for a formal representation of organisations, with their environment, objectives and agents in a way that enables to analyze their partial contributions to the performance of the organisation in a changing environment. These formalisms will enable the analysis of abstract organisation models and the verification of formal properties such as flexibility or robustness (Grossi et al. 2007) in dynamic environments (Dignum and Dignum 2007). On the other hand, the need for such a model to be realistic enough to incorporate the more 'pragmatic' considerations faced by real organisations. Most existing formal models lack this realism, e.g. either by ignoring temporal issues, or by taking a very restrictive view on the controllability of agents, or by assuming complete control and knowledge within the system (cf. van der Hoek and Wooldridge 2005; Santos et al. 1997). Such instruments must moreover be both formal and realistic.

Organisational assessment requires that there is a set of parameters against which organisational performance is to be evaluated. Organisation design processes usually start with an assessment of what is the current state and where there are gaps between desired and actual outcomes. Design strategies will determine what is the focus of the design process. Some organisations place a higher priority on efficiency, preferring designs that minimize the costs of producing goods or services. Other organisations emphasize effectiveness, focusing on generating revenues or seizing leading-edge innovation in the marketplace (Burton et al. 2006). In practice, the process of organisational assessment relies on individual experience and skill of the assessor. Even if best practices exist, these are often not formal or sufficiently detailed to enable an objective comparison and uniform application. In this chapter, we will introduce an abstract model to describe organisations and their environments that allows for formal analysis of the fit between organisational design and environment characteristics. This enables the a priori comparison of designs and their consequences and therefore supports the decision making process on the choice of design. The chapter is organised as follows: In Sect. 20.2, we will describe in detail the characteristics and requirements of the organisational design process. In Sect. 20.3 we present the formal model for organisation specification. The model is further extended in Sect. 20.4 to support the analysis of organisational change. Finally, in Sect. 20.5 we describe a simulation framework that implements this formalism. Conclusion and directions for future work are given in Sect. 20.6.

## 20.2 Organisational Design

Organisations are inherently computational and (distributed) computational systems are inherently organisational (Carley 2002). As such, the use of agent based models and simulation to formally capture and understand organisations has already yielded positive results and promises further tremendous potential (Davidsson 2002). Also, in the Multi-Agent Systems (MAS) research area there is an increased realization that concepts from Organisation Theory are well applicable to understand and structure MAS. Agent organisations can be understood as complex entities where a multitude of agents interact, within a structured environment aiming at some global purpose. Models for the assessment of organisational design must be able to separate organisational from individual concerns.

Agent-based models provide an abstract representation of organisations, their environment, objectives and participating agents that enables the analysis of their partial contributions to the performance of the organisation in a changing environment. The agent-based approach views an organisation as a set of mechanisms of social order that regulate autonomous actors (agents) to achieve common goals (Dignum 2004). The performance of an organisation in such framework is determined both by the structures of interactions between agents, and by the individual characteristics of the agents. Organisational structures define the formal lines of communication, allocation of information processing tasks, distribution of decision-making authorities, and the provision of incentives. From an organisational perspective, the main function of an individual agent is the enactment of a role that contributes to the global aims of the organisation. That is, organisational goals determine agent roles and interaction norms. Agents are then seen actors that perform role(s) described by the organisation design. However, the very notion of agent autonomy refers to the ability of individual agents to determine their own actions, plans and beliefs. From the agent's perspective, its own capabilities and aims determine the reasons and the specific way an agent will enact its role(s), and the behavior of individual agents is motivated from their own goals and capabilities (Dastani et al. 2003; Weigand and Dignum 2004). That is, agents bring their own ways into the society as well.

Role-oriented approaches to agent systems assume this inherent dichotomy between agent interests and global aims (Castelfranchi 1995; Dignum 2004; Zambonelli et al. 2001; Artikis and Pitt 2001). This indicates two levels of autonomy:

- **Internal autonomy**: interaction and structure of the society must be represented independently from the internal design of the agents
- **Collaboration autonomy**: activity and interaction in the society must be specified without completely fixing in advance the interaction structures.

The first requirement relates to the fact that in open environments, organisations allow the participation of multiple heterogeneous entities, the number, characteristics, aims and architecture of which are unknown to the designer, the design of the organisation cannot be dependent on their individual designs. With respect to the second requirement, fundamentally, a tension exists between the goals of the organisation decision makers and the autonomy of the participating entities. On the one hand, the more detail the organisation can use to specify its aims and processes, the more requirements are possible to check and guarantee at design time. This allows, for example, to ensure the legitimacy of the interactions, or that certain rules are always followed. On the other hand, there are good reasons to allow the agents some degree of freedom, basically to enable their freedom to choose their own way of achieving collaboration, and as such increase flexibility and adaptability.

Existing computational simulations of organisations, based on formal models provide a precise theoretical formulation of relationships between variables can inform strategic decisions concerning reorganization (Dignum et al. 2006) but are often limited to a specific domain and difficult to validate (Harrison et al. 2007).

Another aspect to consider is the fact that organisations and their environments are not static. Agents can migrate, organisational objectives can change, or operational behavior can evolve. That is, as circumstances change, organisational structures must also be able to change, disappear or grow. In fact, organisations are active entities, capable not only of adapting to the environment but also of changing that environment, which leads to the question of how and why change decisions are made in organisations (Gazendam and Simons 1998). Models for organisations must therefore be able to describe dynamically adapting organisations to changes in the environment. Such models will enable to understand how different organisations can be designed from different populations of agents, performing different tasks in changing environments, to meet various performance goals (So and Durfee 1998).

To sum up, formal models for organisations that are able to deal with realistic situations, must meet the following requirements (Dignum and Dignum 2007):

1. Represent notions of ability and activity of agents, without requiring knowledge about the specific actions available to a specific agent (internal autonomy).
2. Accept limitedness of agent capability.
3. Represent the abilities and activities of a group of agents.
4. Deal with temporal issues, and especially the fact that activity takes time.
5. Represent the concept of 'being responsible' for the achievement of a given state of affairs.
6. Represent organisational (global) goals and their dependency on agents' activity, by relating activity and organisational structure.
7. Represent organisations in terms of organisational roles that enable individuals to act in accordance with their own capabilities and requirements (collaboration autonomy).
8. Relate roles and agents (role enacting agents).
9. Deal with resource limitedness and the dependency of activity on resources.
10. Represent organisational dynamics (evolution of organisation over time, changes in agent population).
11. Deal with normative issues (representation of boundaries for action and the violation thereof).

The first six requirements above are related to the structural properties of an organisation. The last requirements are related to the operational aspects of an organisation, such as the notion that agent activity has a cost (that is, choosing one or the other course of action is not only dependent on agent capabilities but also the costs of an action must compare positively to its benefits), which is related to actual performance of an agent within the organisational structure.

## 20.3 A Model for Organisation Design

As discussed in the previous sections, one of the main reasons for creating organisations is efficiency, that is, to provide the means for coordination that enable the achievement of global goals in an efficient manner. Such global objectives are not necessarily shared by any of the individual participants, that can only be achieved through combined action. In order to achieve its goals, it is thus necessary that an organisation employs the relevant agents, and assures that their interactions and responsibilities enable an efficient realization of the objectives. In its most simple expression, an organisation consists of a set of agents and a set of objectives in a given environment. Contingency theory states that organisational performance can be seen as a dependent variable on these factors (So and Durfee 1998):

1. **Environment**: this is the space in which organisations exist. This space is not completely controllable by the organisation and therefore results of activity are can not always be guaranteed. Two dimensions of environment are unpredictability and (task) complexity. This class also includes the description of tasks and resources (such as size and frequency).
2. **Agents**: are entities with the capability to act, that is to control the state of some element in the environment. Agents can be defined by their capabilities.
3. **Structure**: describe the features of the organisation, such as roles, relationships and strategy. Differentiating dimensions in this category are size, skills, degree of centralization and formalization.

These are the main building blocks of organisational design and should be included in any model that aims at understanding or specifying organisational performance or behavior. We have developed a formal language LAO (Language for Agent Organisations) based on these concepts to describe and analyze organisational design and performance. The definition of the semantics of LAO is presented in (Dignum and Dignum 2007). In the following, we give an overview of the language.

### *20.3.1 Environment*

In Organisation Theory, environment is commonly defined as the forces outside the organisation that can impact it (Daft et al. 2010). The environment changes over time are not fully controlled by the organisation nor the individuals that populate it. For a formal representation of the environment, we use Kripke semantics to describe the environment and its changes. In LAO, the environment is formally represented as a partially ordered set of worlds $W$. A world describes an actual state of the environment, and transitions indicate the effects of possible changes (the opportunities and threats in terms of organisational theory concepts). The set of atomic facts, $\Phi$, describes the vocabulary (or ontology) of the organisation domain.

At each moment, the state of the organisation is given by a certain state of affairs in the current world. Transitions between worlds represent changes in the situation. Such changes can occur both by decision of the organisation or by external factors affecting the environmental conditions. We use a branching time structure to represent choice, that is, that there is more than one possible future for each world.

### *20.3.2 Agents and Groups*

Agents are active entities that can make change happen, that is, agents can modify the truth value of a fact. Transitions from one to a next world are partially controlled by the agents. Intuitively, the idea is that, in organisations, changes are for some part result of the intervention of (specific) agents. Formally, transitions between worlds are labeled with the set of agents that influence the changes on that transition. That is, for a transition $r = (w,w'),w,w' \in W$, $t(r)$ indicates the set of agents that indeed contribute to the changes indicated by that transition. For each world $w \in W$ and each agent $a \in A$ we can indicate the set of transitions from $w$ for which $a$ has influence. Moreover, at any moment, unexpected change can also occur, which is not result of the action of any of the agents in that world (i.e. unlabelled or partially labelled transitions are possible).

The notions of agent capability and action have been widely discussed in MAS. The intuition is that an agent possesses capabilities that make action possible. That is, in order to talk about agent activity, or, that agent $a$ possesses the ability to make proposition φ hold in some next state in a path from the current world, we need to establish the control of the agent over the (truth) value of φ. For each agent $a$ we partition the set of atomic facts Φ in any world $w$ of M in two classes: the set of atomic facts that agent $a$ can control, $C_a$, and the set of atomic facts that $a$ cannot control, $\overline{C_a}$. Control over a composite fact (i.e. not atomic fact), φ, is possible iff $C_a$ yields φ. For example, if we consider a world where the basic facts are *sand-door* and *paint-door*, an agent $a$ that only has the capability to perform *sand-door*, that is, $C_a$*(sand-door)*, will not be able to control the composite fact *(sand-door* ∧ *paint-door)*. Agents can control the state of variables in a world, represented by $C_a\varphi$ iff the agent controls the set of atomic facts that yield φ. Furthermore, LAO is defined such that no agent can control the obvious (tautologies), and if an agent controls a fact φ it also controls its negation ¬φ. Intuitively, the ability of an agent to realize a state of affairs φ in a world $w$, depends not only on the capabilities of the agent but also on the status of that world. Therefore, we define the ability of $a$, $G_a\varphi$ to represent the case in with the agent not only controls φ but is currently in a state in which it has influence over some of the possible transitions. Only when the agent influences all outgoing transitions from the current world and φ will hold in all worlds reachable from the current world, we can talk about agent activity; in the other cases, there is a (possibly unsuccessful) attempt. The attempt by agent $a$ to realize φ is represented by $H_a\varphi$. Finally, the *stit* operator, $E_a\varphi$ ('agent $a$ sees to it that φ') represents the result of successful action (that is, φ holds in all worlds following the current one). This notion of agent activity is based on that introduced by Pörn (1974) to represent the externally '*observable*' consequences of an action instead of the action itself, and as such abstracts from internal motivations of the agents. *Stit* can be seen as an abstract representation of the family of all possible actions that result in φ.

Agents are, by definition, limited on their capabilities, that is, the set of facts in the world that they can control. This implies that in MAS certain states of affairs can only be reached if two or more agents cooperate to bring that state to be. We define control and action of a group of agents based on the combined atomic capabilities of the agent in the group. For example, if a group contains two agents $a$ and $b$, such that $C_a$*(sand-door)* and $C_b$*(paint-door)*, only working as a group is possible to realize both. This is represented by $C_{\{a,b\}}$*(sand-door* ∧ *paint-door)*.

Table 20.1 gives an overview of agent and group operators. We refer the reader to (Dignum and Dignum 2007) for the formal specification of LAO.

**Table 20.1** Overview of agent and group operators

| | Agent | Group | Description |
|---|---|---|---|
| Capability | $C_a\varphi$ | $C_S\varphi$ | Control over the truth value of φ |
| Ability | $G_a\varphi$ | $G_S\varphi$ | Capability and influence over at least one outgoing transition |
| Attempt | $H_a\varphi$ | $H_S\varphi$ | Ability and φ holds in all worlds reached by an influenced transition |
| Action (stit) | $E_a\varphi$ | $E_S\varphi$ | Ability and φ holds in all next worlds |

### *20.3.3 Structure*

The organisational structure involves two basic concepts: *roles* and their *relations* in terms of which the the overall behavior of the organisation is determined. The specification of the overall behavior of the organisation concerns the optimization of agents' activities, the management and security of the information flow among agents, and the enforcement of certain outcomes. According to Selznick (1948) organisations "represent rationally ordered instruments for the achievement of stated goals", that is, organisations arise in order to achieve specific objectives, and these objectives are pursued defining a number of sub-objectives contributing to the overall purpose of the organisation. These sub-objectives identify the roles that are played in the organisation. The relation between roles and overall objectives of the organisation, i.e., the primitive decomposition of tasks within the organisation, defines the essential form of organisational structure (Grossi et al. 2005). Roles are the basic units over which this structure ranges determining the source of the "rational order" holding in the organisation. A formal definition of organisation therefore needs to include the constructs to relate roles in the organisation in a way that enables objectives to be passed to those agents that can effectively realize them. Note that in our view roles and agents are independent concepts. Role describe activities and services necessary to achieve organisations goals and enable to abstract from the individuals that eventually will enact the role. In this sense, in our model, roles can be seen as place-holders for agents and represent the behavior expected from agents by design. On the other hand, agents are active entities that are able to effectively enact roles in order to achieve organisational objectives. That is, objectives of an organisation, $D_O$ are assigned to roles but can only be achieved through agent action. In order to make this possible, an organisation must in any case employ the relevant agents, that is agents with the right capabilities. In the following, we will use the notion of role enaction agent (Dignum 2004) to refer to the combination of agent and role.

In most cases, the objectives of the organisation are only known to a few of the agents in the organisation, who may have no control over those objectives. It is therefore necessary to organize agents in the organisation in a way that enables objectives to be passed to those agents that can effectively realize them. Dependency between agents $a$ and $b$ is represented by $a \leq b$ meaning that agent $a$ is able to delegate the realization of some state of affairs to agent $b$. Agents can be made responsible for some of the objectives of the organisation, represented by $R_a\varphi$.[1] Responsibility means that either the agent attempts to achieve that objective, or it delegates the objective to another agent with whom it has a dependency relation. That is, given a proposition φ, such that $R_a\varphi$, the order relation $a \leq_O b$ indicates that $a$ is capable to pass its responsibility for φ to $b$, that is, $C_a R_b \varphi$.

### 20.3.4 Organisation Instance

Formally, given a world $w \in W$, an *organisation* $O$ is defined by a set of agents, a set of objectives (missions or desires), and a set of capabilities or assets. The capabilities of an organisation are defined by the capabilities of the agents that are part of the organisation at a given moment. Moreover, the capabilities of agents and groups can vary. Each world describes the capabilities of agents, so that by defining with agents belong to an organisation, one is able to determine the potential capabilities of that organisation. That is, given a organisation $O$ such that $A_O$ is the set of agents in $O$, organisational capability $C_O$ is defined as: $C_O = C_{A_O}$.

We are now able to give a formal definition of organisation as introduced in LAO. Given a set of worlds $W$ representing states of affairs in the environment, an organisation $O$ in a world $w \in W$ is defined as:

$$O^w = \left\{A_O^w, \leq_O^w, D_O^w, S_O^w\right\}$$

where $A_O^w = \{a_1, \ldots, a_n\}$ is the set of agents, $\left(A_O^w, \leq_O^w\right)$ is a partial order relation on $A$ reflecting the structure of the organisation, $D_O^w \subseteq \Phi$ is a set of objectives (states of affairs to achieve), and $S_O^w \subseteq \Phi$ is the set of current assets and characteristics of $O$. This abstract definition of organisations enables to analyze some theoretical properties and incorporates some realistic notions such as bounded capabilities of agents, coordination and control. As an example, consider the organisation $O = (A, \leq_O, D, S^0)$, where:

| |
|---|
| $A$ is the set of agents $\{a,b,c,d\}$ |
| $\leq_O$ are dependencies between the agents |
| $(a \leq_O b, \quad a \leq_O c \leq_O d))$ |
| D = (sand-door ∧ paint-door) ∨ tile-floor ∨ paper-wall, is the objective of O |
| $S^0 = \{R_a D,\ \ C_b\ (\textit{sand-door}),\ \ C_d\ (\textit{paint-door}),\ \ C_{\{b,c\}}\ (\textit{tile-floor})\}$, is the initial state |

Note that the initial organisational state $S^0$ indicates the capabilities of the agents in $A$ and that agent $a$ is responsible for the achievement of the organisational goal. This example also shows that organisations are dependent on the capabilities of their agents to achieve their objectives. In this case, without agent $b$, the

[1] In fact, agents are responsible for the objectives associated with the roles it enacts

**Table 20.2** Example of organisational activity

| | |
|---|---|
| $s_0$ | $R_a$((sand-door ∧ paint-door) ∨ tile-floor ∨ paper-wall) |
| $s_1$ | $E_aR_b(sand\text{-}door) \wedge E_aR_c\ (paint\text{-}door)$ |
| $s_2$ | $R_b(sand\text{-}door) \wedge R_c\ (paint\text{-}door)$ |
| ... | ... |
| $s_i$ | $R_b(sand\text{-}door) \wedge E_cR_d\ (paint\text{-}door)$ |
| $s_{i+1}$ | $R_b(sand\text{-}door) \wedge R_d\ (paint\text{-}door)$ |
| ... | ... |
| $s_{n-1}$ | $E_b(sand\text{-}door) \wedge E_d(paint\text{-}door)$ |
| $s_n$ | Sand-door ∧ paint-door |

organisation can never achieve its goals. A possible strategy for this organisation to realize its objectives is depicted in Table 20.2.

## 20.4 Performance and Change

The model presented in the previous section describes snapshots of organisations, that is, the state of the organisation at a given moment (world). In order to keep effective, organisations must strive to maintain a good fit in a changing environment. Changes in the environment lead to alterations on the effectiveness of the organisation and therefore to the need to reorganize, or in the least, to the need to consider the consequences of that change to the organisation's effectiveness and efficiency. On the other hand, organisations are active entities, capable not only of adapting to the environment but also of changing that environment. This means that organisations are in state of, to a certain degree, altering environment conditions to meet their aims and requirements, which leads to the question of how and why reorganization decisions should be reached. The flexibility of an organisation is defined as the combination of the changeability of an organisational characteristic (structure, technology, culture) and the capabilities of management to change that characteristic (Gazendam and Simons 1998).

In Organisational Theory, the concept of *adaptation* can mean different things, ranging from strategic choice to environmental determinism. Strategic choice refers to the planned pursuit of ends based on a rational assessment of available means and conditions, resulting on a explicit decision to change the organisation. Deterministic views on adaptation, on the other hand, explain organisational change as (involuntary) response to environmental requirements. In this chapter, we treat adaptation as a *design* issue that requires an (explicit) action resulting in the modification of some organisational characteristics. Such decisions can be of two kinds: proactive, preparing the organisation in advance for an expected future, and reactive, making adjustments after the environment has changed (Dignum et al. 2004).

In terms of the formal model of organisations introduced in the previous sections, changes are represented as (temporal) transitions between two different worlds. Given a world $w \in W$, many different events may happen that change some proposition in that world resulting in a different world (the relation $T$ between two worlds represents this). Because not all parameters in $w$ are controllable by the organisation, the current state of the organisation is not necessarily, and in fact in most cases not, completely controlled by the organisation itself. That is, changes are not always an effect of (planned) organisational activity. We distinguish between exogenous and endogenous change. In the exogenous situation, changes occur outside the control, and independently of the actions, of the agents in the organisation, whereas that in the endogenous case, changes that are result of activity explicitely taken by agents in the organisation.

In summary, reorganization consists basically of two activities. Firstly, the formal representation and evaluation of current organisational state and its 'distance' to desired state, and, secondly, the formalization of reorganization strategies, that is, the purposeful change of organisational constituents (structure, agent population, objectives) in order to make a path to desired state possible and efficient. In the following section, we describe how LAO can be used to support reorganization decisions. In Sect. 20.4.2, we will extend LAO to accommodate and reason about changing organisations.

### 20.4.1 Analysis of Reorganisation in LAO

In general, at any given moment, an organisation cannot fully control its state nor the way it will evolve. Formally, given organisation $O = (A, \leq_O, D, S^0)$, this is represented by: $S_O \neq C_O$, or $\exists \varphi \in S_O : \neg C_O \varphi$. That is, there are elements of the desired state that are not achievable by the capabilities currently existing in the organisation. Some parameters may be controlled by external parties, but others may be completely uncontrollable (disasters, natural forces, etc.). External agents which can control parameters that are part of the organisation state are said to have an *influence* on the organisation. Organisations are constantly trying to identify the optimal design for their environment, and will choose a change strategy that they believe will improve their current situation. Figure 20.1 shows the relation between organisation desires and states, and controllable parameters both a generic situation, and the ideal situation.

Ideally, in a successful organisation the set of desires $D_O$ will be a subset of the organisational state $S_O$, which is under the control of the agent in the organisation. Reorganization activities aim therefore at aligning these sets, either by attempting to change the current state, or by altering the set of desires.

Intuitively, performance is a value function on the environment (current world), agents and organisational capability, and on the task (desired state of affairs). The higher the performance the better organisational objectives are achieved in a given state. That is, reorganization decisions start with the evaluation of current organisational performance. In order to specify organisational performance in LAO, we define a function $\theta$ with signature $W \times P(A_{\leq}) \times \Phi \mapsto R$, such that

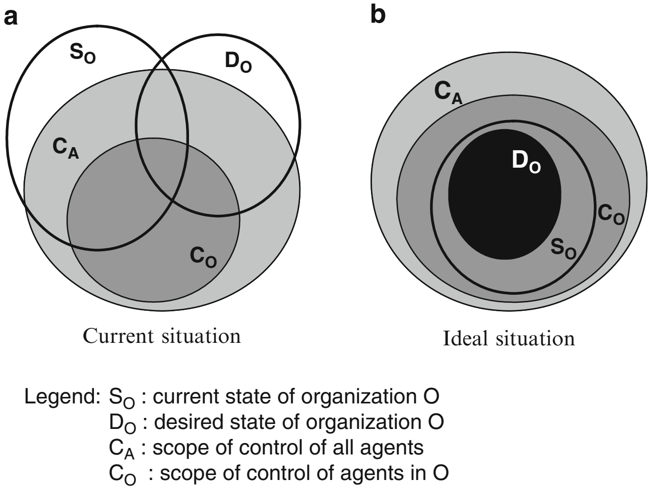



**Fig. 20.1** Organisational state and organisational control

$\theta(w, G_{\leq}, \varphi)$ returns the value of the performance in world $w$ of structured group $G_{\leq}$ for φ, indicating how well $G$ can realize φ. The function θ can be seen as the cost associated with a transition in the world $W$. We assume that for each agent and each world, the performance for each atomic proposition $p$ is fixed. That is, $\forall w,a,p$, $\exists c \in R : \theta\ (w,a,p) = c$. The function θ has the following properties:

1. $\theta(w, G_{\leq}, p) + \theta(w, G_{\leq}, q) \leq \theta(w, G_{\leq}, p \wedge q)$, for atomic propositions $p$, $q$
2. $\neg C_{G_{\leq}} \varphi \rightarrow (\theta(w, G_{\leq}, \varphi) = \infty)$
3. $\theta(w, G_{\leq}, \varphi) \leq \theta(w, G_{\leq}, \varphi \wedge \psi)$
4. $(\varphi \rightarrow \psi) \rightarrow (\theta(w, G_{\leq}, \psi) \leq \theta(w, G_{\leq}, \varphi))$

Informally, the first property above, represents the fact that an agent can get 'tired'. That is, the cost of performing a combination of activities can be higher than the sum of the costs associated with each activity, for that agent (Cholvy et al. 2005). The second property indicates that performance is impossible if the capability is non-existent. Properties 3 and 4 are associated with the logical relations between propositions. Some authors have used finite state machines to describe organisation transitions (Matson and DeLoach 2005). An important difference between our work and such approaches is the temporal dimension of our model, that is, an organisation can never move to a previous state even if conceptually equivalent, all states are different in time.

The function θ provides a way to establish the cost of achievement of a certain state of affairs, given the current state and group of agents. Based on this function, different strategies can be adopted that determine the course of action to follow in order to choose a organisation with better performance with respect to a given goal. In the next subsection, we describe the actions that can be taken in order to effectively change the organisation. In the case of endogenous reorganization processes, it must be possible for agents within the organisation to monitor the organisation performance, as part of the process of deciding on reorganization. We therefore specify the capability and responsibility of agents for checking performance as follows: $C_a$*check-perform* $(\theta(w, G_{\leq}, \varphi), \upsilon)$, and $R_a$*check-perform* $(\theta(w, G_{\leq}, \varphi), \upsilon)$. We abbreviate these expressions to $C_a$ *check-perform* and $R_a$ *check-perform* to indicate the capability, resp. responsibility to check the overall performance. Similar expressions are also defined for groups of agents.

### *20.4.2 Acting for Change*

In this section, we are concerned with planned reorganization. That is, we want to specify the activities that are consciously taken by an organisation in order to attempt to modify its characteristics. Note that, under planned reorganization, we consider both endogenous reorganization and exogenous reorganization, in which reorganization is achieved by activity outside of the system, for example by the designer (and thus off-line).

In human organisations, reorganization strategies take different forms, such as hiring new personnel, downsizing, training or reassigning tasks or personnel (Carley and Svoboda 1996). In principle, organisations can also decide to modify their mission or objectives. However, studies suggest that such second-order change, that is, a shift from one strategic orientation to another, is atypical even in times of highly dynamic environmental behavior (Fox-Wolfgramm et al. 1998). Because organisations aim at making certain states of affairs to be the case, and only agents can bring affairs to be, it is important for the organisation to make sure it 'employs' and organizes an adequate set of agents $(A_O, \leq_O)$ such that the combined action of those agents has the potentiality to bring about the desired state of affairs $D_O$. The dependency relation $\leq_O$ between agents must allow for the desired states to be achieved, that is, dependencies must be sufficient for responsibilities to be passed to the appropriate agents, that is, the agents that have the necessary capabilities. If that is not the case, the organisation should take the steps needed to decide and implement reorganization, such that the resulting organisation $O'$ is indeed able to realize its objectives $D_O$. In practice, reorganization activities can be classified in three groups:

- **Staffing**: Changes on the set of agents: adding new agents, or deleting agents from the set. This corresponds to HRM activities in human organisations (hiring, firing and training).
- **Structuring**: Changes on the ordering structure of the organisation. Corresponding to infra-structural changes in human organisations: e.g. changes in composition of departments or positions.

- **Strategy**: Changes on the objectives of the organisation: adding or deleting desired states. Corresponding to strategic (or second-order) changes in human organisations: modifications on the mission, vision, or charter of the organisation.

Given an organisation $O = (A_O, \leq_O, D_O, S_O)$ and an agent $a$, such that $C_a\varphi$, a possible initial definition for the reorganization operations over $O$ is:

- ***Staff***$^+(\mathbf{O}, \mathbf{a})$: Adding agent $a$ to organisation $O$ results in organisation $O' = (A_O \cup \{a\}, \leq_O, D_O, S_O \cup C_a\varphi)$.
- ***Staff***$^-(\mathbf{O}, \mathbf{a})$: If $a \in A_O$, removing agent $a$ from organisation $O$ results in organisation $O' = (A_O/\{a\}, \leq_O/dep_a, D_O, S_O/C_a\varphi)$ where $dep_a = \{(x{=}_O y){:}x = a \vee y = a\}$
- ***Struct***$^+(\mathbf{O}, (\mathbf{a} = \mathbf{b}))$: If $a, b \in A_O$, adding dependency $(a{\leq_O}b)$ to organisation $O$ results in organisation $O' = (A_O, \leq_O \cup \{(a \leq b)\}, D_O, S_O)$
- ***Struct***$^-(\mathbf{O}, (\mathbf{a} \leq \mathbf{b}))$: If $(a{\leq_O}b)$, removing dependency $(a{\leq_O}b)$ from organisation $O$ results in organisation $O' = (A_O, \leq_O/\{(a \leq b)\}, D_O, S_O)$
- ***Strateg***$^+(\mathbf{O}, \mathbf{d})$: If $(d \wedge D_O)$ is consistent, adding objective $d$ to organisation $O$ results in organisation $O' = (A_O, \leq_O, D_O \cup \{d\}, S_O)$
- ***Strateg***$^-(\mathbf{O}, \mathbf{d})$: If $d \in D_O$, removing objective $d$ from organisation $O$, results in organisation $O' = (A_O, \leq_O, D_O/d, S_O)$

The above definition gives a fairly simple (naive) description of model updates. This is especially the case for the strategic reorganization operations. The solution chosen above is to specify that strategic decisions cannot be realized if they yield in a contradiction. In reality, a more elaborate formal treatment of reorganization, must consider belief revision problems that result from the addition and/or deletion of model components. This is an issue for future research and will not be further discussed in this chapter.

The operations described above enable the reorganization of an agent organisation. The issue of deciding about reorganization still remains. That is, how do organisations reach the decision to reorganize? What should then be reorganized? When should one reorganize? An informal foundation for reasoning about dynamic reorganization is given in (Dignum et al. 2004). In Sect. 20.4.3 we will further discuss this issue. However, nothing is yet said about how to reach the decision to reorganize, and how to monitor the state of an organisation wrt its desires and environment conditions.

Reorganization operations are just operations, that is, effect the value of some variables. That is, reorganization can also be either endogenous or exogenous. In the *exogenous* case, reorganization lays outside the control of any agent in the world, and is often realized by action of the system designer. In the *endogenous* case, agents are able to achieve the states resulting from reorganization operations (that is, can control reorganization results). In this case, we can specify an agent $a$ such that $C_a\sigma$ where σ is one of the reorganization results specified above. For example, the fact that agent $a$ is able to hire other agents is represented by

$C_a staff^+(O,b)$. We can thus cover both the concept of engineered reorganization (exogenous) and dynamic reorganization (endogenous). Furthermore, as discussed in (Dignum et al. 2004; Barber and Martin 2001), in the endogenous case, reorganization can either be *collaborative*, that is, there is a group of agents $G \subseteq A$, such that $C_G\rho$, or *role-based*, that is, a single agent controls the reorganization result, $C_a\rho$, where ρ is one of the reorganization operations defined above.

### 20.4.3 Deciding About Change

An explicit, planned reorganization strategy must take into account the current performance and determine which characteristics of the organisation should be modified in order to achieve a better performance. The general idea behind reorganization strategies, is that one should be able to evaluate the utility of the current state of affairs (that is, what happens if nothing changes), and the utility of future states of affairs that can be obtained by performing reorganization actions. The choice is then to choose the future with the highest utility.

By applying the θ function defined in Sect. 20.4.1, one is able to determine the current and future performance of the organisation, and also calculate the cost of reorganization. This can be used to decide about when to reorganize. In order to enable endogenous reorganization, agents in the organisation must be able to evaluate the performance of the organisation, that is, have the capability $C_a$*check-perform* $(\theta(w, G_{\leq}, \varphi), \upsilon)$, and responsibility $R_a$ *check-perform* $(\theta(w, G_{\leq}, \varphi), \upsilon)$. Moreover, expressions describing different reorganization decisions can be expressed in the same way. For instance, $R_a$*minimize-perform* indicates that agent $a$ is responsible to minimize organisation performance costs.

The general idea is that one evaluates the utility of the current state, and the utility of all following states obtained by performing controlling or structuring actions. The choice is then to move into the state with the highest utility. A possible strategy to decide whether to realize reorganization operation ρ is if $\theta(w, O, \rho) + \theta(w', O', D_O) \leq \theta(w, O, D_O)$, where $O'$ is the organisation in $w'$ resulting from the realization of ρ on organisation $O$ in $w$. Informally, this strategy says that if the cost of reorganization plus the cost of achieving $D_O$ by the reorganized organisation is less than the cost of achieving $D_O$ without reorganizing, then reorganization should be chosen. Another issue to take into account on the reorganization decision is the frequency of reorganization. In particular when considering human organisations, it cannot be expected that the organisation can be reorganized every time a change occurs. Such information can be included in the calculation of the θ action by adding a 'penalty' for frequent reorganization. In (Dignum et al. 2006) we describe some experiments concerning frequency of change.

## 20.5 Evaluation Through Simulation

We have applied the model presented in this chapter to formally describe several existing agent systems that simulate organisations and organisational adaptation. Those systems were taken from the literature and include such different domains as architectures for Command and Control (Kleinman et al. 2005), RoboSoccer (Hübner et al. 2004), market regulation (Nissen 2001) and sensor networks (Matson and DeLoach 2004). The original work on those cases use different frameworks and were developed independently by other research groups. For simplicity reasons, we will use here the example of the builders organisation described in Sect. 20.3.

We have developed a simulation framework, **Organisation Ontology Simulator** (OOS) that provides an ontology for LAO concepts (Dignum and Tick 2007; Dignum 2010). Organisation instances are then specified as instances of the organisation ontology. OOS converts the ontology instances into a custom Java-based format, which can be used by the Repast[2] simulation environment to visualize and run the simulations. In addition, OOS provides a library of agent behaviors based on the 3APL agent platform[3], which allows to use software agents with cognitive BDI capabilities within the Repast environment. A simulation is created as follows. First, a specific organisation is created as instance of the organisation ontology. Agent instances are associated with standard agent behaviors existing in the 3APL library of OOS. A more detailed description of OOS can be found in (Tick 2007).

The case study described in this section aims at illustrating the framework in order to assess organisational design in a fictive situation by designing two sufficiently different environments, and two organisations that exploit the differences between environments. The idea is that in the case objectives are incongruent with organisation, i.e. in the current organisational situation objectives cannot be achieved, organisations will need to reorganize in order to fulfil their missions. For instance, organisation $O$ is incongruent with $D_O$ if $C_O \neq D_O$. Using the builders organisation introduced in Sect. 20.3 we define the following types of environments and organisations:

- **Environments**: Provide the tasks to be taken up by the organisation. We define three types of tasks: *paint-door*, *paper-wall*, *tile-floor*. Frequency of tasks, set in the simulation at 0.25, indicates how often the organisation will be requested to perform a new task. Two different types of environments are identified as follows:

  Heterogenous (**Het**): all task types occur with the same probability.
  Homogenous (**Hom**): mostly one type of task: tasks of type *paint-door* occur in 80 % of the cases; other tasks occur with 10 % probability each.

[2] http://repast.sourceforge.net/

[3] http://www.cs.uu.nl/3apl/

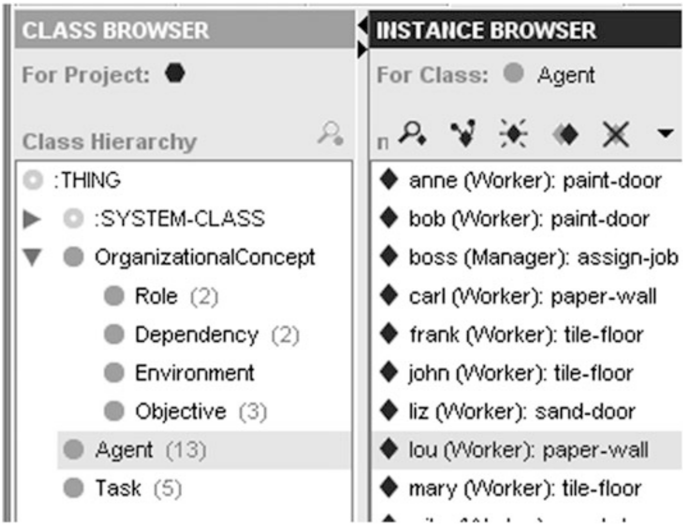



**Fig. 20.2** Organisation ontology for the builder scenario

- **Organisations**: In this simulation we assume a hierarchical structure and two different roles: manager and worker. We populate the organisation with 1 manager agent and 12 worker agents. We further specify two types of organisations as follows:

  Specialists (**Spec**): each worker can only perform tasks of one type (four workers for each type). Time to complete a task is calculated randomly as $T_s \in [50, 100]$.
  Generalists (**Gen**): all workers can perform all tasks, however take longer to complete a task: $T_g = 2^* T_s - 25$.

The idea is that a manager receives tasks to be executed and distributes those to a worker with capability to perform that task and is currently free. If no capable agent is free, the manager keeps tasks in a waiting list. There is a penalty for keeping tasks in the waiting list, resulting in less profit. Each completed task results in a profit of $x$ for the organisation. For each waiting task the organisation looses $x/10$ per simulation tick (representing working days). In total, four different scenarios are obtained, as follows: Het-Gen, Het-Spec, Hom-Gen, Hom-Spec.

OOS was used to generate simulations of the four scenarios above, each corresponding to a different instantiation of the organisation ontology depicted in Fig. 20.2. Each scenario, with a length of 400 time units, was run 10 times, resulting in 40 different simulations. For each scenario, we calculated the number of completed tasks, the number of waiting tasks, and the average completion length. Figure 20.3 shows the results obtained.

As can be expected, the performance of Generalist organisations is not influenced by the type of environment. Specialist organisations do not perform

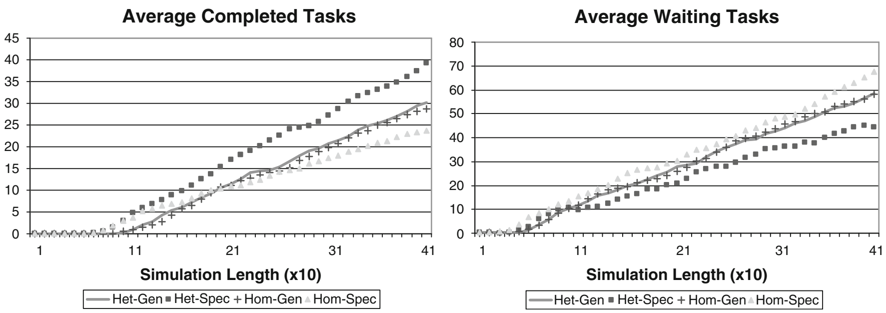



**Fig. 20.3** Organisational performance

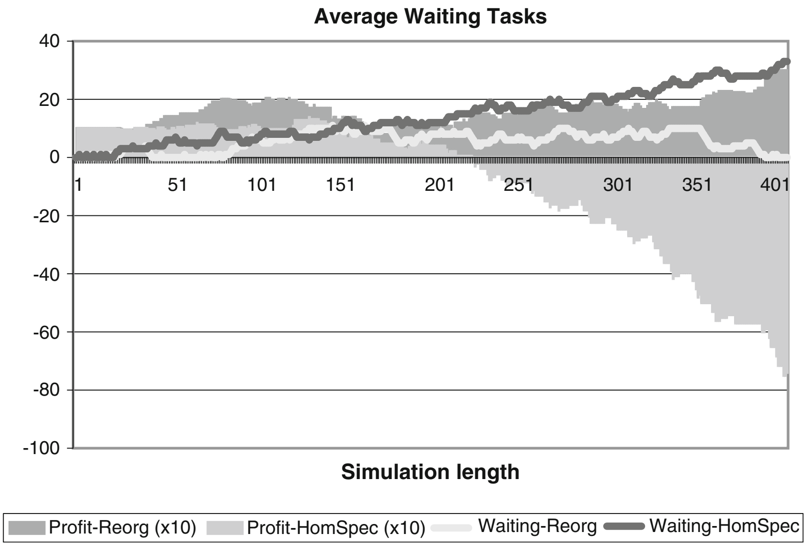



**Fig. 20.4** Reorganization effect

well in Homogeneous environments (workers capable of *paint-door* are too busy, other workers mostly idle). In such case, reorganization should be considered. We provided manager agents with the capability to check organisational performance and adapt accordingly. In the experiment, manager agents of Specialist organisations in a Homogeneous environment monitor the number of waiting tasks and perform a reorganization action when that number reaches a given threshold. Specifically, when the manager notices that there are ten or more waiting tasks, it will train one, if any, idle agent to perform the task *paint-door*. Formally, this is represented as: $E_m staff^-(O,a)$ for some agent $a$, followed by $E_m\, staff^+(O,p)$ where agent $p$ has capabilities $C_p paint\text{-}door$. Cost reorganization, that is, the cost of training one agent is fixed at $x$*3, where $x$ is the task profit. Figure 20.4 shows the difference of performance between organisation Hom-Spec and organisation

Hom-Spec-Reorg endowed with such reorganization capabilities, for task completion times $T_s \in [25, 50]$ and task frequency 25 %.

As it can be seen, reorganization results in fewer waiting tasks and therefore higher profit due to less penalties. Depending on the values for profit, penalties and reorganization costs, simulations can be used to determine an effective threshold.

## 20.6 Conclusions

In this chapter, we presented a theoretic model to describe organisations and their environments that enables the formal analysis of the fit between organisational design and environment characteristics. This enables the a priori comparison of designs and their consequences and therefore supports the decision making process on the choice of design. Reorganization is needed in order to enable systems to enforce or adapt to changes in the environment. This issue has been discussed by many researchers in both Organisational Theory as in Distributed Systems, resulting mostly in domain-oriented empiric solutions. The lack, in most cases, of a formal basis makes difficult the development of theories about reorganization, prevents the comparison or approaches and results, and makes difficult the adaptation of models to other domains or situations. The theoretical model presented in this chapter provides a formal, provable representation of organisations, with their environment, objectives and agents in a way that enables to analyze their partial contributions to the performance of the organisation in a changing environment. Moreover, the model includes enough realistic concepts that enable it to represent the more ‘pragmatic’ considerations faced by real organisations. The semantics of LAO, the model for organisational concepts and the reorganization process presented in this chapter are based on modal temporal logic. We further introduced a framework for the rapid development of organisational simulations. This framework provides a structured efficient way to deploy many different organisational and environment designs, by using ontologies describing organisation structures, environment characteristics and agent capabilities, and provides semi-automatic means to generate simulations from the ontology instances. Moreover, the simulation framework incorporates the cognitive characteristics of agents. We are currently finalizing the development of the framework and in the near future will be using it for the modelling and analysis of realistic organisations.

The simple scenario presented in Sect. 20.5 demonstrates the use of simulations to support the process of assessment of organisational performance. By varying the types and capabilities of the agents and the characteristics of the environment, is possible to analyze the fit of organisational structures and the potential of different reorganization decisions to the performance of the organisation. Currently, we are extending the evaluation of the model and the OOS tool to the analysis of more realistic organisations. In parallel, we are also engaged on the extension of LAO and the specification of a full axiomatization for LAO.

**Acknowledgements** The research presented in this chapter is funded by the Netherlands Organisation for Scientific Research (NWO), through Veni-grant 639.021.509. The author is grateful to Christiaan Tick and Daghan Acay for their contribution to the development of the OOS simulation framework.

## Further Reading

For readers interested in organisational design in general, (Burton et al. 2006) give a step-by-step approach whereas (Ostrom 2005) provides a framework for analysing diverse economic, political and social institutions. The use of agent-based models to investigate how internal policies and procedures affect the performance of organisations is demonstrated by (Carley 2002). Applying organisational design principle to multi-agent systems, Dignum and Paget (2013) give an overview of the state of the art, while Horling and Lesser (2004) provide a survey of current paradigms.

## References

Aldrich H, Pfeffer J (1976) Environments of organizations. Annu Rev Sociol 2:79–105
Artikis A, Pitt J (2001) A formal model of open agent societies. In: Proceedings of the fifth international conference on autonomous agents, Montreal, 28 May–01 June 2001. ACM Press, New York, pp 192–193
Barber KS, Martin CE (2001) Dynamic reorganization of decision-making groups. In: Proceedings of the fifth international conference on autonomous agents, Montreal, 28 May–01 June 2001. ACM Press, New York, pp 513–520
Burton R, DeSanctis G, Obel B (2006) Organizational design: a step by step approach. Cambridge University Press, Cambridge
Carley K (2002) Computational organisation science: a new frontier. PNAS 99(3):7257–7262
Carley K, Svoboda D (1996) Modeling organisational adaptation as a simulated annealing process. Sociol Method Res 25(1):138–168
Castelfranchi C (1995) Guarantees for autonomy in cognitive agent architecture. In: Wooldridge MJ, Jennings NR (eds) Intelligent agents, ECAI-94 workshop on agent theories, architectures and languages, Amsterdam, 8–9 Aug 1994. Springer, Berlin, pp 56–70
Cholvy L, Garion C, Saurel C (2005) Ability in a multi-agent context: a model in the situation calculus. In: Toni F, Torroni P (eds) CLIMA'05 proceedings of the 6th international conference on computational logic in multi-agent systems. Springer, Berlin, pp 23–36
Daft R, Murphy J, Willmott H (2010) Organization theory and design. South-Western Cengage Learning, Andover
Dastani M, Dignum V, Dignum F (2003) Role assignment in open agent societies. In: Proceedings of the second international joint conference on autonomous agents and multiagent systems, AAMAS 2003, 14–18 July 2003, Melbourne. ACM Press, New York, pp 489–496
Davidsson P (2002) Agent based social simulation: a computer science view. J Artif Soc Soc Simul 5(1). http://jasss.soc.surrey.ac.uk/5/1/7.html
Dignum V (2004) A model for organizational interaction: based on agents, founded in logic. PhD thesis, SIKS dissertation series 2004–1, Utrecht University

Dignum V (2010) Ontology support for agent-based simulation of organisations. Multiagent Grid Syst 6:191–208
Dignum V, Dignum F (2007) A logic for agent organization. In: Proceedings of the international workshop on formal aspects of MAS, Durham, 3–7 Sept 2007
Dignum V, Paget J (2013) Multi-agent organizations. In: Weiss G (ed) Multiagent systems, 2nd edn. MIT Press, Cambridge
Dignum V, Tick C (2007). Agent-based analysis of organisations: formalization and simulation. In: Proceedings of IEEE/WIC/ACM international conference on intelligent agent technology (IAT'07), Silicon Valley, 2–5 Nov 2007. IEEE, pp 244–247
Dignum V, Dignum F, Sonenberg L (2004) Towards dynamic organisation of agent societies. In: Vouros G (ed) Workshop on coordination in emergent agent societies, ECAI 2004. pp 70–78
Dignum V, Dignum F, Sonenberg L (2006) Exploring congruence between organisational structure and task performance: a simulation approach. In: Boissier O et al (eds) Coordination, organization, institutions and norms in multi-agent systems, ANIREM'05/OOOP'05, Utrecht, 25–26 July 2005, revised selected papers (Lecture notes in computer science, 3913). Springer, Berlin, pp 213–230
Duncan R (1979) What is the right organisational structure: decision tree analyis provides the answer. Organ Dyn 7(3):59–80
Fox-Wolfgramm S, Boal K, Hunt J (1998) Organizational adaptation to institutional change: a comparative study of first-order change in prospector and defender banks. Adm Sci Q 43 (1):87–126
Galbraith J (1977) Organization design. Addison-Wesley, Reading
Gazendam H, Simons JL (1998) An analysis of the concept of equilibrium in organization theory. In: Proceedings of the computational and mathematical organization theory workshop, Montreal, 25–26 Apr 1998
Grossi D, Dignum F, Dastani M, Royakkers L (2005) Foundations of organizational structures in multiagent systems. In: Dignum F et al (eds) Proceedings of the 4th international joint conference on autonomous agents and multiagent systems (AAMAS 2005), 25–29 July 2005, Utrecht. ACM Press, pp 690–697
Grossi D, Dignum F, Dignum V, Dastani M, Royakkers L (2007) Structural aspects of the evaluation of agent organisations. In: Noriega P et al (eds) Coordination, organizations, institutions, and norms in agent systems II, AAMAS 2006 and ECAI 2006 international workshops, COIN 2006, Revised selected papers (Lecture notes in computer science, 4386). Springer, Berlin, pp 3–18
Harrison R, Lin Z, Carroll G, Carley K (2007) Simulation modeling in organizational and management research. Acad Manage Rev 32(4):1229–1245
Horling B, Lesser V (2004) A survey of multi-agent organizational paradigms. Knowl Eng Rev 19 (4):281–316
Hübner JF, Sichman JS, Boissier O (2004) Using the moise + for a cooperative framework of MAS reorganisation. In: Bazzan A, Labidi S (eds) Advances in artificial intelligence – SBIA 2004, Proceedings of the 17th Brazilian symposium on artificial intelligence (Lecture notes in computer science, 3171). Springer, Berlin, pp 481–517
Kleinman D, Levchuk G, Hutchins S, Kemple W (2005) Scenario design for the empirical testing of organisational congruence (Technical report ADA440390). Department of informational science, Naval Postgraduate School, Monterey
Matson E, DeLoach SA (2004) Enabling intra-robotic capabilities adaptation using an organisation-based multiagent system. In: Proceedings of the 2004 I.E. international conference on robotics and automation (ICRA 2004)
Matson E, DeLoach SA (2005) Formal transition in agent organizations. In: Thompson C, Hexmoor H (eds) 2005 International conference on integration of knowledge intensive multiagent systems (KIMAS'05): modeling, exploration, and engineering, 18–21 Apr 2005, Waltham. IEEE, pp 235–240

Mintzberg H (1993) Structures in fives: designing effective organizations. Prentice Hall, Englewood Cliffs

Nissen M (2001) Agent-based supply chain integration. Inf Technol Manage 2(3):289–312

Ostrom E (2005) Understanding institutional diversity. Princeton Universiy Press, Princeton

Pörn I (1974) Some basic concepts of action. In: Stenlund S (ed) Logical theory and semantical analysis. Reidel, Dordrecht, pp 93–101

Roberts J (2004) The modern firm: organizational design for performance and growth. Oxford University Press, Oxford

Santos F, Jones A, Carmo J (1997) Action concepts for describing organised interaction. In: Proceedings of the 30th annual Hawaii international conference on system sciences (HICCS-30), 7–10 Jan 1997, Maui, vol 5. IEEE Computer Society Press, pp 373–382

Selznick P (1948) Foundations of the theory of organization. Am Sociol Rev 13(1):25–35

So Y, Durfee E (1998) Designing organisations for computational agents. In: Carley K, Prietula M, Gasser L (eds) Simulating organizations. Cambridge, MA: The MIT Press, pp 47–64

Tick C (2007) Organization structure analysis by simulation. Master's thesis, Utrecht University

van der Hoek W, Wooldridge M (2005) On the logic of cooperation and propositional control. Artif Intell 164(1–2):81–119

Weigand H, Dignum V (2004) I am autonomous, you are autonomous. In: Nickles M, Rovatsos M, Weiss G (eds) Agents and computational autonomy, potential, risks, and solutions, vol 2969, Lecture notes in computer science. Springer, Berlin, pp 227–236

Zambonelli F, Jennings N, Wooldridge M (2001) Organisational abstractions for the analysis and design of multi agent systems. In: Ciancarini P, Wooldridge M (eds) Agent-oriented software engineering, vol 1957, Lecture notes in computer science. Springer, Berlin, pp 235–251

# Chapter 21
# Distributed Computer Systems

**David Hales**

**Why Read This Chapter?** To understand how simulating social complexity might be used in the process of designing distributed computer systems.

**Abstract** Ideas derived from social simulation models can directly inform the design of distributed computer systems. This is particularly the case when systems are "open", in the sense of having no centralised control, where traditional design approaches struggle. In this chapter we indicate the key features of social simulation work that are valuable for distributed systems design. We also discuss the differences between social and biological models in this respect. We give examples of socially inspired systems from the currently active area of peer-to-peer systems and finally we discuss open areas for future research in the field.

## 21.1 Introduction

Massive and open distributed computer systems provide a major application area for ideas and techniques developed within social simulation and complex systems modelling. In the early years of the twenty-first century there has been an explosion in global networking infrastructure in the form of wired and wireless broadband connections to the internet encompassing both traditional general purpose computer systems, mobile devices and specialist appliances and services. The challenge is to utilise such diverse infrastructure to provide novel services that satisfy user needs reliably. Traditional methods of software design and testing are not always applicable to this challenge. Why is this? And what can social simulation and complexity perspectives bring to addressing the challenge? This chapter answers these questions by providing a general overview of some of the major benefits of approaching design from socially inspired perspectives in addition to examples of

D. Hales (✉)
University of Bologna, Bologna, Italy
e-mail: dave@davidhales.com

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_21,

applications in the area of peer-to-peer (P2P) systems and protocols. Finally we will speculate on possible future directions in the area.

This chapter is not an exhaustive survey of the area, for example, we have not discussed the application of social network analysis techniques to web graphs and social networks that are constructed within, or facilitated by, distributed software systems (Staab et al. 2005). Both these are active areas. Further, we have not discussed the active research area based on randomised "Gossiping" approaches, where information is diffused over networks through randomised copying of information between adjacent nodes (Wang et al. 2007).

## 21.2 What's Wrong with Traditional Design Approaches?

Traditional design approaches to systems and software often assume that systems are essentially "closed" – meaning they are under the control of some administrative authority that can control access, authenticate users and manage system resources such as issuing software components and updates. Consider the simplest situation in which we have a single computer system that is not connected to a network that is required to solve some task. Here, design follows a traditional process of analysis of requirements, specification of requirements then design and iterative refinement, until the system meets the specified requirements. User requirements are assumed to be discoverable and translatable into specifications at a level of detail that can inform a design. The designer has, generally, freedom to dictate how the system should achieve the required tasks via the coordination of various software components. This coordination generally follows an elaborate sequencing of events where the output from one component becomes the input to others. The design task is to get that sequence right.

In "open systems" it is assumed that there are multiple authorities. This means that components that comprise the system cannot be assumed to be under central control. An extreme example of this might be an open peer-to-peer (P2P) system in which each node in the system executes on a user machine under the control of a different user. Popular file-sharing systems operate in this way, allowing each user to run any variant of client software they choose. These kinds of systems function because the client software implements publicly available peer communication protocols, allowing the nodes to interconnect and provide functionality. However, due to the open nature of such systems, it is not possible for the designer, a priori, to control the sequence of processing in each node. Hence the designer needs to consider what kinds of protocols will produce *acceptable* system level behaviours under *plausible* assumptions of node behaviour. This requires a radically different design perspective in which systems need to be designed as self-repairing and self-organising systems in which behaviour emerges bottom-up rather than being centrally controlled.

One term for this approach, based on self-organisation and emergence, is so-called Self-Star (or Self-*) systems (Babaoglu et al. 2005). The term is a broad expression that aims to capture all kinds self-organising computer systems that continue to function acceptably under realistic conditions.

But what kinds of design approach can be employed? Currently there is no accepted general theory of self-organisation or emergence – rather there are some interesting models at different levels of abstraction that capture certain phenomena. Many such models have been produced within the biological sciences to explain complex self-organising biological phenomena. Biological systems, particularly co-evolving systems, appear to evidence many of the desirable properties required by self-* computer systems. Hence, several proposed self-* techniques have drawn on biological inspiration (Babaoglu et al. 2006).

## 21.3 Socially Versus Biological Inspiration

It is useful to ask in what way social organisation differs from the biological level. In this section we briefly consider this question with regard to desirable properties for information systems. An important aspect of human social systems (HSS) is their ability (like biological systems) to both preserve structures – with organisations and institutions persisting over time – and adapt to changing environments and needs. The evolution of HSS is not based on DNA, but rather on a complex interplay between behaviour, learning, and individual goals. Here we present some distinguishing aspects of HSS.

### *21.3.1 Rapid Change*

A feature of HSS is the speed at which re-organisations can occur. Revolutions in social organisation can take place within the lifetime of a single individual. Hence although HSS often show stable patterns over long periods, rapid change is possible. The ability to respond rapidly would appear to be a desirable property in rapidly changing information system environments; however, for engineering purposes, one must ensure that such fast changes (unlike revolutions!) can be both predicted and controlled.

### *21.3.2 Non-Darwinian Evolution*

HSS do not evolve in a Darwinian fashion. Cultural and social evolution is not mediated by random mutations and selection of some base replicator over vast time periods, but rather follows a kind of collective learning process. That is, the information storage media supporting the change by learning – and hence (as noted above), both the mechanisms for change and their time scale – are very different from those of Darwinian evolution. Individuals within HSS can learn both directly from their own parents (vertical transmission), from other members of the older generation (diagonal transmission), or from their peers (horizontal transmission). Hence, new cultural traits (behaviours, beliefs, skills) can be propagated quickly through a HSS. This can be contrasted with simple Darwinian transmission in which, typically, only vertical transmission of genetic information is possible. Although it is possible to characterise certain processes of cultural evolution based on the fitness of cultural replicators (Boyd and Richerson 1985) or memes (Dawkins 1976) it is important to realise such replicators are not physical – like DNA – but part of a socio-cognitive process – passing through human minds – and may follow many kinds of selective process (Lumsden and Wilson 1981). The problems of using the idea of biological evolution in the social sciences are discussed in more detail in Chap. 18 (Chattoe-Brown and Edmonds 2013) of this volume.

### *21.3.3 Stable Under Internal Conflict*

HSS exist because individuals need others to achieve their aims and goals. Production in all HSS is collective, involving some specialisation of roles. In large modern and post-modern HSS roles are highly specialised, requiring large and complex co-ordination and redistribution methods. However, although HSS may sometime appear well integrated, they also embody inherent conflicts and tensions between individual and group goals. What may be in the interests of one individual or group may be in direct opposition to another. Hence, HSS embody and mediate conflict on many levels.

This aspect is highly relevant to distributed and open information systems. A major shift from the closed monolithic design approach is the need to deal with and tolerate inevitable conflicts between sub-components of a system. For example, different users may have different goals that directly conflict. Some components may want to destroy the entire system. In open systems this behaviour cannot be eliminated and hence needs in some way to be tolerated.

### *21.3.4 Only Partial Views and Controversy*

Although HSS are composed of goal directed intelligent agents, there is little evidence that individuals or groups within them have a full view or understanding of the HSS. Each individual tends to have a partial view often resulting from specialisation within, and complexity of, the system. Such partial views, often dictated by immediate material goals, may have a normative (how things "should" be) character rather than a more scientific descriptive one (how things "are"). Consequently, the ideas that circulate within HSS concerning the HSS itself tend to take on an "ideological" form. Given this, social theories are rarely as consensual as those from biological sciences. Thus, social theories include a high degree of controversy, and they lack the generally accepted foundational structure found in our understanding of biology. However, from an information systems perspective, such controversy is not problematic: we do not care if a given social theory is true for HSS or not; we only care if the ideas and mechanisms in the theory can be usefully applied in information systems. This last point, of course, also holds for controversial theories from biology as well (e.g. Lamarckian evolution).

### 21.3.5 Trust and Socially Beneficial Norms

In trying to understand the stability of socially functional behaviour, much work within the social sciences has focused on the formation and fixation of "norms" of behaviour. Many researchers working with Multi-Agent Systems (MAS) have attempted to create artificial versions of norms to regulate MAS behaviours – although much of these have not been based on theories from HSS (although see Conte and Paolucci 2002). Certainly the establishment and stability of beneficial norms (such as not cheating one's neighbour) is a desirable property visible in all stable HSS (Hales 2002). This point (the existence and power of norms) is of course closely related to the previous point, which notes that norms can influence understanding and perception.

It is widely agreed that, in HSS, many observed behaviours do not follow the same pattern as would be expected from simple Darwinian evolution or individual "rational" behaviour – in the sense of maximising the chance of achieving individual goals. Behaviour is often more socially beneficial and co-operative or altruistic, generally directed towards the good of the group or organisation within which the individual is embedded. (We note the widespread appearance of altruistic behavior among many species of social mammals – such that, once again, we speak here of a difference in degree between HSS and other social animals.) Many theories and mechanisms have been proposed by social scientists for this kind of behaviour (Axelrod 1984), with many of these formalised as computer algorithms; furthermore, several of these have already been translated for use in information systems (Cohen 2003; Hales and Edmonds 2005).

### 21.3.6 Generalised Exchange and Economics

Almost all HSS evidence some kind of generalised exchange mechanisms (GEM) – i.e. some kind of money. The emergence of GEM allows for co-ordination through trade and markets. That, is, collective co-ordination can occur where individual entities (individuals or firms) behave to achieve their own goals. It is an open (and perhaps overly simplified) question whether certain norms are required to support GEM or, rather, most norms are created via economic behaviour within GEM (Edmonds and Hales 2005). Certainly, the formation and maintenance of GEM would be an essential feature of any self-organised economic behaviour within information systems – currently many information systems work by assuming an existing GEM a priori, i.e. they are parasitic on HSS supplying the trust and norms required. Such systems require trusted and centralised nodes before they can operate because they do not emerge such nodes in on-going interaction. However, given that GEM exist, a huge amount of economic theory, including new evolutionary economics and game theory, can be applied to information systems.

## 21.4 What Can Social Simulation Offer the Designer?

Social simulation work has the potential to offer a number of insights that can be applied to aid design of distributed computer systems. Social simulators have had no choice but to start from the assumption of open systems composed of autonomous agents – since most social systems embody these aspects. In addition, much social simulation work is concerned with key aspects of self-* systems such as:

- *Emergence and self-organisation*: understanding the micro-to-macro and the macro-to-micro link. Phenomena of interest often result from bottom-up processes that create emergent structures that then constrain or guide (top-down) the dynamics of the system.
- *Cooperation and trust*: getting disparate components to "hang together" even with bad guys around. In order for socially integrated cooperation to emerge it is generally necessary to employ distributed mechanisms to control selfish and free-riding behaviour. One mechanism for this is to use markets (see Chap. 23 in this volume, Rouchier 2013) but there are other methods.
- *Evolving robust network structure*: Constructing and maintaining functional topologies robustly. Distributed systems often form dynamic networks in which the maintenance of certain topological structures improves system level performance.
- *Constructing adaptive/evolutionary heuristics* rather than rational action models. Models of both software and user behaviour in distributed systems are based on implicit or explicit models. Traditional approaches in game theory and economics have assumed rational action but these are rarely applicable in distributed systems.

These key aspects have import into two broad areas of system design. Firstly, simulation models that produce desirable properties can be adapted into distributed system protocols that attempt to reproduce those properties. Secondly, models of agent behaviour, other than rational action approaches, can be borrowed as models of user behaviour in order to test existing and new protocols.

Currently, however, it is an open question as to how results obtained from social simulation models can be productively applied to the design of distributed information systems. There is currently no general method whereby desirable results from a social simulation model can be imported into a distributed system. It is certainly not currently the case that techniques and models can be simply "slotted into" distributed systems. Extensive experimentation and modification is required. Hence, in this chapter we give specific examples from P2P systems where such an approach has been applied.

## 21.5 What About Agent-Orientated Design Approaches?

Multi-Agent System (MAS) design approaches have previously been proposed (Wooldridge and Jennings 1995), which attempt to address some of the design issues raised by open systems. Those approaches start with a "blank sheet" design approach rather than looking for biological or social inspiration. The focus therefore has tended to be on logical foundations, proof, agent languages and communication protocols. For example, the BDI agent framework starts from the assumption that agents within a system follow a particular logical architecture based on "folk psychological" cognitive objects – such as beliefs or intentions (Rao and Georgeff 1991). However, such approaches have difficulty scaling to large societies with complex interactions particularly where the effects of emergence and self-organisation are important. A more recent approach within MAS work has been to look towards self-organising approaches using simulation to capture processes of emergence (Brueckner et al. 2006). In this work heavy use has been made of biological and socially inspired approaches.

## 21.6 Examples of Socially Inspired P2P Systems

Here we give very brief outlines of some P2P protocols that have been directly inspired by social simulation models. While P2P systems have not been the only distributed systems that benefited from social inspiration we have focused on this particular technology because it is currently, at the time of writing, a very active research area and increasingly widely deployed on the internet.

### *21.6.1 Reciprocity Based BitTorrent P2P System*

BitTorrent (Cohen 2003) is an open P2P file sharing system that draws directly from the social simulation work of Robert Axelod (1984) on the evolution of cooperation. The protocol is based on a form of the famous Tit-For-Tat (TFT) strategy popularised by Axelrods computer simulation tournaments. Strategies were compared by having agents play the canonical Prisoner's Dilemma (PD) game.

The PD game captures a social dilemma in the form of a minimal game in which two players each select a move from two alternatives (either to cooperate or defect) then each player receives a score (or pay-off). If both players cooperate then both get a reward payoff ($R$). If both defect they are punished, both obtaining payoff $P$. If one player selects defect and the other selects cooperate then the defector gets $T$ (the "temptation"), the other receives $S$ (the "sucker"). When these pay-offs, which are numbers representing some kind of desirable utility (for example, money), obey the following constraints: $T > R > P > S$ and $2R > T + S$ then we say the game represents a Prisoner's Dilemma. When both players cooperate this maximizes the collective good but when one player defects and another cooperates this represents a form of free riding with the defector gaining a higher score (T) at the expense of the co-operator (S).

Axelrod asked researchers to submit computer programs to a "tournament" where they repeatedly played the PD against each other accumulating payoffs. The result of the tournaments was that a simple strategy, TFT, did remarkably well against the majority of other submitted programs – although other strategies can also survive within the complex ecology that occurs when there is a population of competing strategies.

TFT operates in environments where the PD is played repeatedly with the same partners for a number of rounds. The basic strategy is simple: a player starts by cooperating then in subsequent rounds copies the move made in the previous round by its opponent. This means defectors are punished in the future: the strategy relies on future reciprocity. To put it another way, the "shadow" of future interactions motivates cooperative behaviour in the present. In many situations this simple strategy can outperform pure defection.

In the context of BitTorrent the basic mechanism is simple: files are split into small chunks (about 1 MB each) and downloaded by peers, initially, from a single hosting source. Peers then effectively "trade" chunks with each other using a TFT-like strategy – i.e. if two peers offer each other a required chunk then this equates to mutual cooperation. However if either does not reciprocate then this is analogous to a defect and the suckered peer will retaliate in future interactions.

The process is actually a little more subtle because each peer is constantly looking at the upload rate/download rate from each connected peer in time – so it does not work just by file chunk but by time unit within each file chunk. While a file is being downloaded between peers, each peer maintains a rolling average of the download rate from each of the peers it is connected to. It then tries to match its uploading rate accordingly. If a peer determines that another is not downloading fast enough then it may "choke" (stop uploading) to that other. Figure 21.1 shows a schematic diagram of the way the Bittorrent protocol structures population interactions.

Additionally, peers periodically try connecting to new peers randomly by uploading to them – testing for better rates. This means that if a peer does not upload data to other peers (a kind of defecting strategy) then it is punished by other peers in the future (by not sharing file chunks) – hence a TFT-like strategy based on punishment in future interactions is used.

Axelrod used the TFT result to justify sociological hypotheses such as understanding how fraternisation broke out between enemies across the trenches of WW1. Cohen has applied a modified form of TFT to produce a file sharing system

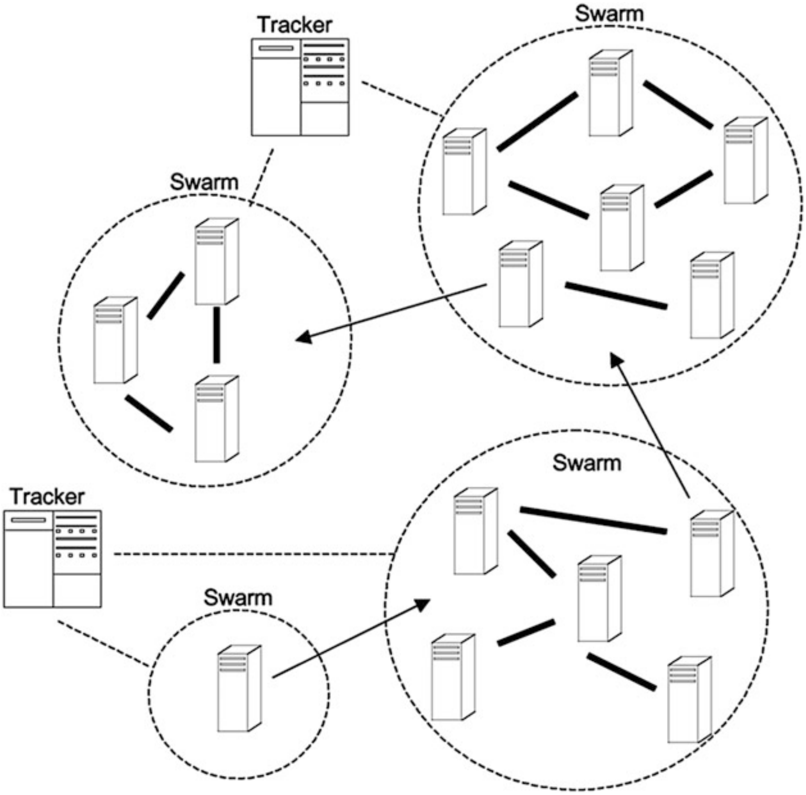



**Fig. 21.1** A schematic of a portion of a BitTorrent system. The trackers support swarms of peers each downloading the same file from each other. Thick lines indicate file data. They are constantly in flux due to the application of the TFT-like "choking" protocol. Trackers store references to each peer in each supported swarms. It is not unknown for trackers to support thousands of swarms and for swarms to contain hundreds of peers. *Arrows* show how peers might, through user intervention, move between swarms. Generally at least one peer in each swarm would be a "seeder" that holds a full copy of the file being shared

resistant to free riding. However, TFT has certain limitations, it requires future interactions with the same individuals and each has to keep records of the last move made by each opponent. Without fixed identities it is possible for hacked clients to cheat BitTorrent. Although it appears that widespread cheating has not actually spread in the population of clients. It is an open question as to why this might be (but see (Hales and Patarin 2006) for a hypothesis).

#### 21.6.1.1 Group Selection Based P2P Systems

Recent work, drawing on agent-based simulations of cooperative group formation based on "tags" (social labels or cues) and dynamic social networks suggests a mechanism that does not require reciprocal arrangements but can produce cooperation and specialisation between nodes in a P2P (Riolo et al. 2001; Hales and Edmonds 2005). It is based on the idea of cultural group selection and the well-known social psychological phenomena that people tend to favour those believed to be similar to themselves – even when this is based on seemingly arbitrary criteria (e.g. supporting the same football team). Despite the rather complex lineage, like TFT the mechanism is refreshingly simple. Individuals interact in cliques (subsets of the population). Periodically, if they find another individual who is getting higher utility than themselves they copy them – changing to their clique and adopting their strategy. Also, periodically, individuals form new cliques by joining with a randomly selected other.

Defectors can do well initially, suckering the co-operators in their clique – but ultimately all the co-operators leave the clique for pastures new – leaving the defectors all alone with nobody to free-ride on. Those copying a defector (who does well initially) will also copy their strategy, further reducing the free-riding potential in the clique. So a clique containing any free riders quickly dissolves but those containing only co-operators grow.

Given an open system of autonomous agents all cliques will eventually be invaded by a free rider who will exploit and dissolve the clique. However, so long as other new cooperative cliques are being created cooperation will persist in the overall population. In the context of social labels or "tags" cliques are defined as those individuals sharing particular labels (e.g. supporting the same football team). In the context of P2P systems the clique is defined as all the other peers each peer is connected to (its neighbourhood) and movement between cliques follows a process of network "re-wiring".

Through agent-based simulation, the formation and maintenance of high levels of cooperation in the single round PD and in a P2P file-sharing scenario has been demonstrated (Hales and Edmonds 2005). The mechanism appears to be highly scalable with zero scaling cost – i.e. it does not take longer to establish cooperation in bigger populations. Figure 21.2 shows the evolution of cooperative clusters within a simulated network of peer nodes. A similar approach was presented by Hales and Arteconi (2006) that produced small-world connected cooperative networks rather than disconnected components.

In addition to maintaining cooperation between nodes in P2P, the same group selection approach has been applied to other areas such as the coordination of robots in a simulated warehouse scenario and to support specialisation between nodes in a P2P job sharing system (Hales and Edmonds 2003; Hales 2006).

#### 21.6.1.2 Segregation Based P2P Systems

The model of segregation proposed by Thomas Schelling is well known within social simulation (Schelling 1969, 1971). The model demonstrates how a macro-structure of segregated clusters or regions robustly emerges from simple local behaviour rules. Schelling's original model consists of agents on a two dimensional

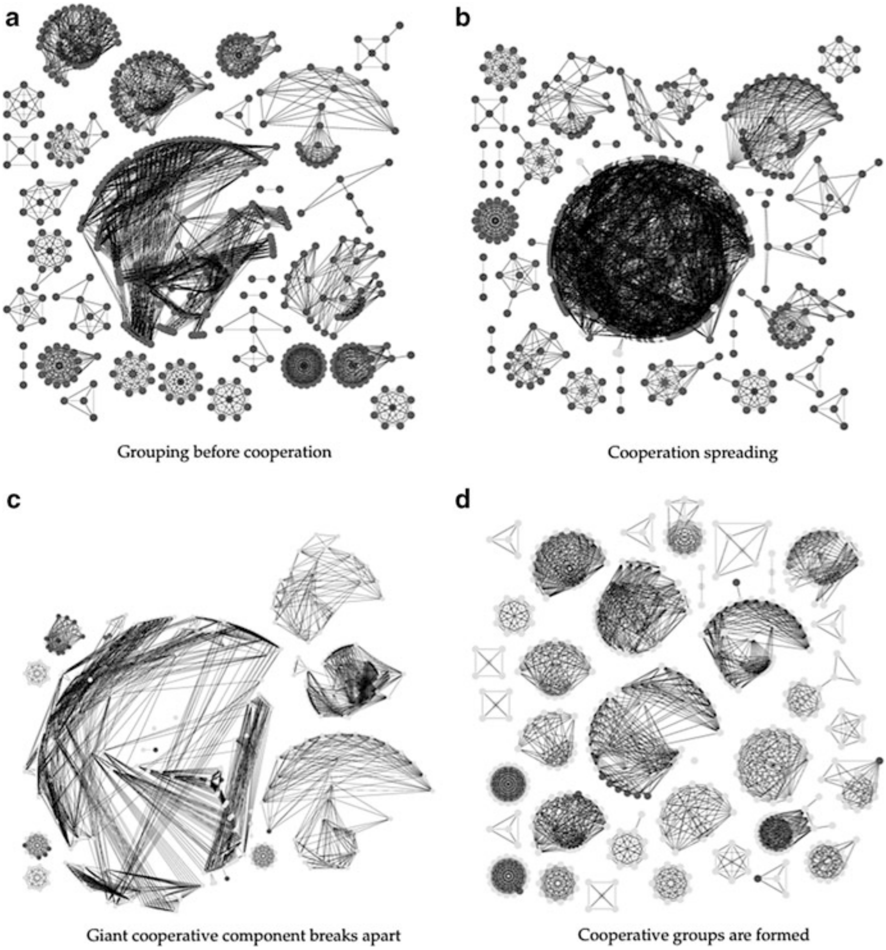



**Fig. 21.2** Evolution of the connection network between nodes playing the Prisoner's Dilemma. From an initially random topology composed of all nodes playing the defect strategy (*dark shaded nodes*), components quickly evolve, still containing all defect nodes (**a**). Then a large cooperative component emerges in which all nodes cooperate (**b**). Subsequently the large component begins to break apart as defect nodes invade the large cooperative component and make it less desirable for cooperative nodes (**c**). Finally an ecology of cooperative components dynamically persists as new components form and old components die (**d**). Note: the cooperative status of a node is indicated by a *light shade*

grid. Each grid location can hold a single agent or may be empty. Each agent maintains a periodically updated satisfaction function. Agents take one of two colours that are fixed. An agent is said to be satisfied if at least some proportion of adjacent agents on the grid have the same colour, otherwise the agent is said to be not satisfied. Unsatisfied agents move randomly in the grid to a free location. The main finding of the model is that even if the satisfaction proportion is very low, this

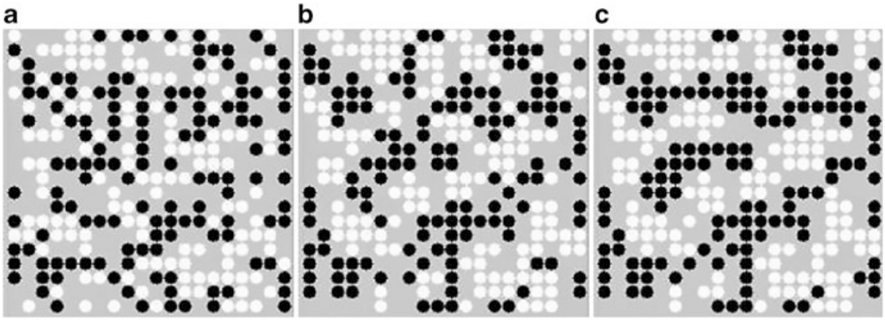



**Fig. 21.3** Example run of the Schelling segregation model. From an initially random distribution of agents (**a**) clusters emerge over successive iterations (**b**) and (**c**). In this example the satisfaction proportion is 0.5, meaning that agents are unsatisfied if less than 50 % of neighbours in the grid are a different colour (Taken from Edmonds and Hales 2005)

still leads to high levels of segregation by colour – i.e. large clusters of agents emerge with the same colour. Figure 21.3 shows an example run of the model in which clusters of similar colours emerge over time.

The results of the segregation model are robust even when nodes randomly leave and enter the system – the clusters are maintained. Also agents in the segregation model only require local information in order to decide on their actions. These properties are highly desirable for producing distributed information systems and therefore it is not surprising that designs based on the model have been proposed.

Sing and Haahr (2006) propose a general framework for applying a modified form of Schelling's model to topology adoption in a P2P network. They show how a simple protocol can be derived that maintains a "hub-based" backbone topology within unstructured networks. Hubs are nodes in a network that maintain many links to other nodes. By maintaining certain proportions of these within networks it is possible to improve system performance for certain tasks. For many tasks linking the hubs to form a backbone within the network can further increase performance. For example, the Gnutella[1] file-sharing network maintains high-bandwidth hubs called "super-peers" that speed file queries and data transfer between nodes. Figure 21.4 shows an example of a small network maintaining a hub backbone.

In the P2P model nodes represent agents and neighbours are represented by explicit lists of neighbour links (a so-called *view* in P2P terminology). Nodes adapt their links based on a satisfaction function. Hub nodes (in the minority) are only satisfied if they have at least some number of other hubs in their view. Normal nodes are only satisfied if they have at least one hub node in their view. Hence different node types use different satisfaction functions and exist in a network rather than lattice. It is demonstrated via simulation that the network maintains a stable and connected topology supporting a hub backbone under a number of conditions, including dynamic networks in which nodes enter and leave the network over time.

[1] http://www.gnutella.com

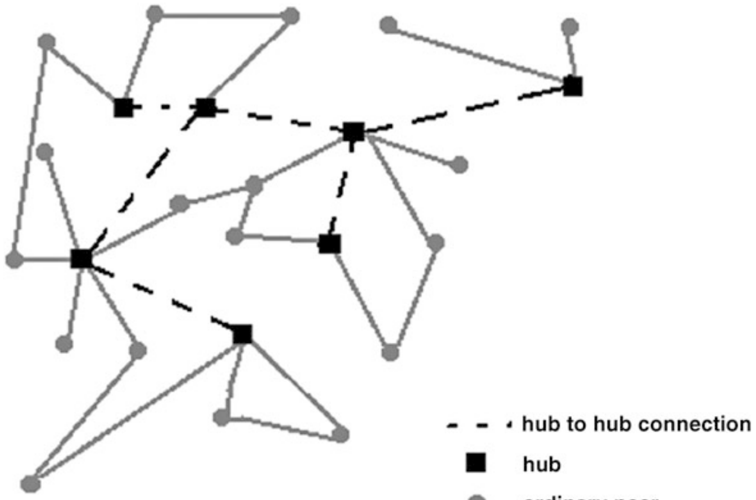



**Fig. 21.4** An example of a hub based peer-to-peer topology. Hubs are linked and perform specialist functions that improve system performance (Taken from Singh and Haahr 2006)

The approach presented by Sing and Haahr is given as a general approach (a template design pattern) that may be specialised to other P2P application areas rather than just self-organising hub topologies. For example, they apply the same pattern to decrease bandwidth bottlenecks and increase system performance of a P2P by clustering similar nodes based on bandwidth capacity (Singh and Haahr 2004).

## 21.7 Possible Future Research

In the following sections we give a brief outline of some promising possible areas related to socially inspired distributed systems research.

### *21.7.1 Design Patterns*

Social simulators and distributed systems researchers currently constitute very different communities with different backgrounds and goals. A major problem for moving knowledge between these disciplines is the different language, assumptions and outlets used by them. One promising approach for communicating techniques from social simulation to distributed systems designers is to develop so-called "design patterns" which provide general templates of application for given techniques. This approach has been influential within object-oriented programming and recently biologically inspired approaches have been cast as design patterns (Gamma et al. 1995; Babaoglu et al. 2006). Design patterns are not formal and need not be tied to a specific computer language but rather provide a consistent framework and nomenclature in which to describe techniques that solve recurrent problems the designer may encounter. At the time of writing few, if any, detailed attempts have been made to present techniques from social simulation within a consistent framework of design patterns.

#### 21.7.1.1 The Human in the Loop: Techno-social Systems

Most distributed and open systems function via human user behaviour being embedded within them. In order to understand and design such systems some model of user behaviour is required. This is particularly important when certain kinds of user intervention are required for the system to operate effectively. For example, for current file sharing systems (e.g. BitTorrent) to operate users are required to perform certain kinds of altruistic actions such as initially uploading new files and maintaining sharing of files after they have been downloaded (so called "seeding"). Web2.0 systems often require users to create, upload and maintain content (e.g. Wikipedia). It seems that classical notions of rational action are not appropriate models of user behaviour in these contexts. Hence, explicitly or implicitly such distributed systems require models of user behaviour which capture, at some level, realistic behaviour. Such systems can be viewed as techno-social systems – social systems that are highly technologically mediated.

One promising method for understanding and modelling such systems is to make use of the participatory modelling approach discussed in Chap. 10 (Barreteau et al. 2013). In such a system user behaviour is monitored within simulations of the technical infrastructure that mediates their interactions. Such an approach can generate empirically informed and experimentally derived behaviour models derived from situated social interactions. This is currently, at the time of writing, an underdeveloped research area.

Interestingly, from the perspective of distributed systems, if it is possible to model user behaviour at a sufficient level of detail based on experimental result then certain aspects of that behaviour could be incorporated into the technological infrastructure itself as protocols.

#### 21.7.1.2 Power, Leadership and Hierarchy

A major area of interest to social scientists is the concept of power – what kinds of process can lead to some individuals and groups becoming more powerful than others? Most explanations are tightly related to theories of inequality and economic relationships; hence this is a vast and complex area.

Here we give just a brief very speculative sketch of recent computational work, motivated by sociological questions, that could have significant import into understanding and engineering certain kinds of properties (e.g. in peer-to-peer systems), in which differential power relationships emerge and may, perhaps, be utilised in a functional way. See Chap. 25 in this volume (Geller and Moss 2013) for a detailed overview of modelling power and authority in social simulation.

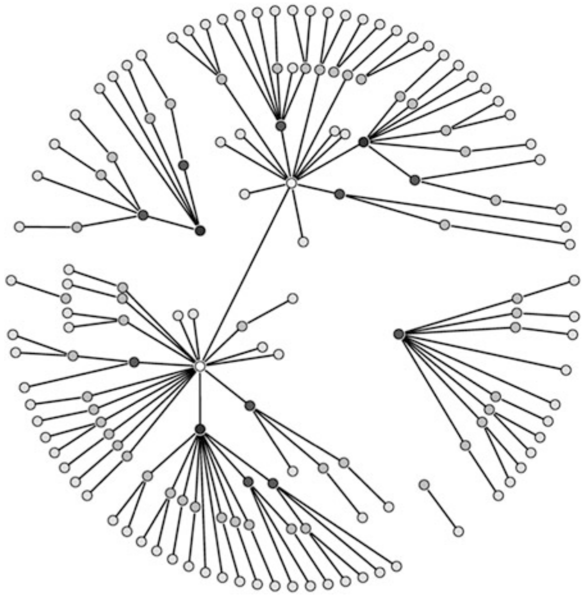


**Fig. 21.5** Forms of 'hierarchy', 'leadership' and unequal wealth distribution have been observed to emerge in simulated interaction networks (From Zimmerman et al. 2001). Nodes play PD-like games with neighbours and break connections based on a simple satisfaction rule. Hierarchies are produced in which some nodes are more connected and hence can affect the network dramatically by their individual actions – a form of 'topological power'

Interactions in human society are increasingly seen as being situated within formal and informal networks (Kirman and Vriend 2001). These interactions are often modelled using the abstraction of a game capturing interaction possibilities between linked agents (Zimmerman et al. 2001). When agents have the ability to change their networks based on past experience and some goals or predisposition, then, over time, networks evolve and change.

Interestingly, even if agents start with more-or-less equal endowments and freedom to act, and follow the same rules, vastly unequal outcomes can be produced. This can lead to a situation in which some nodes become objectively more powerful than other nodes through topological location (within the evolved network) and exploitative game interactions over time.

Zimmermann et al. (2001) found this in their simulations of agents playing a version of the Prisoner's Dilemma on an evolving network. Their motivation and interpretation is socio-economic: agents accumulate 'wealth' from the payoffs of playing games with neighbours and make or break connections to neighbours based on a simple satisfaction heuristic similar to a rule discussed in Kirman (1993).

Figure 21.5 shows an example of an emergent stable hierarchical network structure. Interestingly, it was found that, over time, some nodes accumulate large amounts of 'wealth' (through exploitative game behaviour) and other nodes become 'leaders' by being at the top of a hierarchy. These unequal topological and wealth distributions emerge from simple self-interested behaviour within the network. Essentially, leaders, through their own actions, can re-arrange the topology of the network significantly whereas those on the bottom of the hierarchy have little 'topological power'.

The idea of explicitly recognising the possibility of differential power between sub-units in self-* systems and harnessing this is an idea rarely discussed in engineering contexts but could offer new ways to solve difficult co-ordination problems.

Considering P2P applications, one can envisage certain kinds of task in which differential power would be required for efficient operation. Consider e.g. two nodes negotiating an exchange on behalf of their group or follower nodes. This might be more efficient than individual nodes having to negotiate with each other every time they wished to interact. Or consider a node reducing intra-group conflict by imposing a central plan of action.

We mention the notion of engineering emergent power structures, briefly and speculatively here, because we consider power to be an under-explored phenomenon within evolving information systems. Agents, units or nodes are often assumed to have equal power. It is rare for human societies to possess such egalitarian properties and perhaps many self-* like properties are facilitated by the application of unequal power relationships. We consider this a fascinating area for future work.

## 21.8 Summary

We have introduced the idea of social inspiration for distributed systems design and given some specific examples from P2P systems. We have argued that social simulation work can directly inform protocol designs. We have identified some of the current open issues and problem areas within this research space and pointed out promising areas for future research.

Increasingly, distributed systems designers are looking to self-organisation as a way to address their difficult design problems. Also, there has been an explosive growth in the use of such systems over the Internet, particularly with the high take-up of peer-to-peer systems. Also the idea of "social software" and the so-called "Web2.0" approach indicate that social processes are becoming increasingly central to system design. We believe that the wealth of models and methods produced with social simulation should have a major impact in this area over the coming decade.

## Further Reading

The interested reader could look at the recent series of the IEEE Self-Adaptive and Self-Organising systems (SASO) conference proceedings which started in 2007 and have been organised annually (http://www.saso-conference.org/). To get an idea of current work in social simulation a good place to start is the open access online Journal Artificial Societies and Social Simulation (JASSS), see http://jasss.soc.surrey.ac.uk/JASSS.html.

## References

Axelrod R (1984) The evolution of cooperation. Basic Books, New York
Babaoglu O et al (eds) (2005) Self-star properties in complex information systems: conceptual and practical foundations, vol 3460, Lecture notes in computer science. Springer, Berlin
Babaoglu O et al (2006) Design patterns from biology for distributed computing. ACM T Auton Adapt Syst 1(1):26–66
Barreteau O et al (2013) Participatory approaches. Chapter 10 in this volume
Boyd R, Richerson P (1985) Culture and the evolutionary process. University of Chicago Press, Chicago
Brueckner S, Di Marzo Serugendo G, Hales D, Zambonelli F (eds) (2006) Engineering self-organising systems. In: Proceedings of the 3rd workshop on engineering self-organising applications (EOSA'05) (Lecture notes in artificial intelligence, 3910). Springer, Berlin
Chattoe-Brown E, Edmonds B (2013) Evolutionary mechanisms. Chapter 18 in this volume
Cohen B (2003) Incentives build robustness in BitTorrent. In: Chuang J, Krishnan R (eds) Proceedings of the first workshop on the economics of peer-2-peer systems, 5–6 June 2003, Berkley. http://www2.sims.berkeley.edu/research/conferences/p2pecon/papers/s4-cohen.pdf
Conte R, Paolucci M (2002) Reputation in artificial societies: social beliefs for social order. Kluwer, Amsterdam
Dawkins R (1976) The selfish gene. Oxford University Press, Oxford
Edmonds B, Hales D (2005) Computational simulation as theoretical experiment. J Math Sociol 29 (3):209–232
Gamma E, Helm R, Johnson R, Vlissides J (1995) Design patterns: elements of reusable object-oriented software. Addison-Wesley, Reading
Geller A, Moss S (2013) Modeling power and authority: an emergentist view from Afghanistan. Chapter 25 in this volume
Hales D (2002) Group reputation supports beneficent norms. J Artif Soc Soc Simul 5(4). http://jasss.soc.surrey.ac.uk/5/4/4.html
Hales D (2006) Emergent group-level selection in a peer-to-peer network. Complexus 3:108–118
Hales D, Arteconi S (2006) SLACER: a self-organizing protocol for coordination in P2P networks. IEEE Intell Syst 21(2):29–35
Hales D, Edmonds B (2003) Evolving social rationality for MAS using "Tags". In: Rosenchein JS et al (eds) Proceedings of the 2nd international conference on autonomous agents and multi-agent systems, Melbourne, July 2003 (AAMAS 2003). ACM Press, New York, pp 497–503
Hales D, Edmonds B (2005) Applying a socially-inspired technique (tags) to improve cooperation in P2P networks. IEEE T Syst Man Cy A Syst Hum 35(3):385–395
Hales D, Patarin S (2006) How to cheat BitTorrent and why nobody does. In: Jost J, Reed-Tsochas F, Schuster P (eds) ECCS2006, Proceedings of the European conference on complex systems, towards a science of complex systems. http://www.cabdyn.ox.ac.uk/complexity_PDFs/ECCS06/Conference_Proceedings/PDF/p19.pdf
Kirman A (1993) Ants, rationality and recruitment. Q J Econ 108:137–156
Kirman AP, Vriend NJ (2001) Evolving market structure: an ACE model of price dispersion and loyalty. J Econ Dyn Control 25(3/4):459–502
Lumsden C, Wilson E (1981) Genes, mind and culture. Harvard University Press, London
Rao AS, Georgeff MP (1991) Modeling rational agents within a BDI-architecture. In: Fikes R, Sandewall E (eds) Proceedings of knowledge representation and reasoning (KR&R-91). Morgan Kaufmann, San Mateo, pp 473–484
Riolo R, Cohen M, Axelrod R (2001) Evolution of cooperation without reciprocity. Nature 414:441–443
Rouchier J (2013) Markets. Chapter 23 in this volume
Schelling T (1969) Models of segregation. Am Econ Rev 59:488–493
Schelling T (1971) Dynamic models of segregation. J Math Sociol 1(1):143–186

Singh A, Haahr M (2004) Topology adaptation in P2P networks using Schelling's model. In: Oh JC, Mosse D (eds) Proceedings of the workshop on games and emergent behaviours in distributed computing environments, co-located with PPSN VIII, Birmingham, 18–22 Sept 2004

Singh A, Haahr M (2006) Creating an adaptive network of hubs using Schelling's model. Commun ACM 49(3):69–73

Staab S et al (2005) Social networks applied. IEEE Intell Syst 20(1):80–93

Wang F-Y, Carley KM, Zeng D, Mao W (2007) Social computing: from social informatics to social intelligence. IEEE Intell Syst 22(2):79–83

Wooldridge M, Jennings NR (1995) Intelligent agents: theory and practice. Knowl Eng Rev 10(2):115–152

Zimmermann MG, Eguiluz VM, San Miguel M (2001) Cooperation, adaptation and the emergence of leadership. In: Kirman A, Zimmermann JB (eds) Economics with heterogeneous interacting agents. Springer, Berlin, pp 73–86

# Chapter 22
# Simulating Complexity of Animal Social Behaviour

**Charlotte Hemelrijk**

**Why Read This Chapter?** To get an overview of simulation models aimed at understanding animal social behaviour, such as travelling, foraging, dominance, or task division. The chapter also provides an analysis of the kinds of insight each simulation provides, how specific these insights are and whether they are testable.

**Abstract** Complex social phenomena occur not only among humans, but also throughout the animal kingdom, from bacteria and amoebae to non-human primates. At a lower complexity they concern phenomena such as the formation of groups and their coordination (during travelling, foraging, and nest choice) and at a higher complexity they deal with individuals that develop individual differences that affect the social structure of a group (such as its dominance hierarchy, dominance style, social relationships and task division). In this chapter, we survey models that give insight into the way in which such complex social phenomena may originate by self-organisation in groups of beetle larvae, in colonies of ants and bumblebees, in groups of fish, and groups of primates. We confine ourselves to simulations and models within the framework of complexity science. These models show that the interactions of an individual with others and with its environment lead to patterns at a group level that are emergent and are not coded in the individual (genetically or physiologically), such as the oblong shape of a fish school, specific swarming pattern in ants, the centrality of dominants in primates and the task division among bumble bees. The hypotheses provided by these models appear to be more parsimonious than usual in the number of adaptive traits and the degree of cognitive sophistication involved. With regard to the usefulness of these simulations, we discuss for each model what kind of insight it provides, whether it is biologically relevant, and if so, whether it is specific to the species and environment and to what extent it delivers testable hypotheses.

C. Hemelrijk (✉)
Professor of Self-organisation in social systems, Behavioural ecology and self-organisation, BESO, Centre for Ecological and Evolutionary Studies, University of Groningen, Centre for Life Sciences, room 72.05.54, Nijenborgh 7, 9747 AG GRONINGEN
e-mail: c.k.hemelrijk@rug.nl

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_22,

## 22.1 Introduction

Many complex social phenomena of human behaviour are also observed in animals. For instance, humans coordinate their movement while searching for the right restaurant. They also make and follow paths on lawns in between different university buildings in order to cross between them. This resembles the path marking and following behaviour of ants as they forage and select different food sources. Furthermore, a division of tasks is found in both human social organisations and in large colonies of social insects, such as honey bees. Human social relationships are diverse, and so are the social relationships of primates. In groups of humans and animals, competition results in stable relationships in which one individual consistently beats the other. Within a group, individuals can be ordered in a dominance hierarchy. Furthermore, dominant relationships are also found between groups and between classes of individuals, such as between the sexes. Further, societies may differ in their dominance style, such as whether they are egalitarian or despotic. In summary, despite the hugely inferior cognitive capacities of animals they show a number of complex social phenomena that resemble those of humans. From this the question arises whether or not these complex phenomena originate by self-organisation in the same way as both humans and animals. Therefore, it is important to survey complex social behaviour in animals in addition to that of humans.

We divide the chapter in subsections dealing with different (groups of) complex phenomena ordered in increasing complexity: From group formation (which is simple in animals) and coordination under various circumstances (such as during travelling, foraging, selection of shelter and nest site) we continue to the social organisation of the group (its dominance hierarchy, dominance style, dominance classes, social relationships, personality style and task division).

We mainly confine ourselves to individual-based models that are spatially explicit. Individuals in the model are steered by behavioural rules that are fixed or that are based on parameters that change (compare self-organisation with(out) structural changes, Pfeifer and Scheier 1999); they react to their local environment only. The environment may contain only other individuals, or may also contain food. Food may be continuously abundant, or it may be depleted and may re-grow or not after being eaten.

In general, we discuss for each model whether it has led to a new perspective in the study of the phenomenon; whether this perspective is more ‘parsimonious’ than previous explanations in terms of cognitive sophistication and the number of specific behavioural adaptations; whether the new explanation concerns a general principle or is specific to a certain species or environment; and whether it produces hypotheses that can easily be tested in real animals. In the evaluation we also provide literature for further reading and indicate important areas for future modelling.

## 22.2 Group Formation and Coordination

Everywhere in nature groups are formed. They are formed among individuals of all kinds of species in all stages of their lives and in a few cases groups contain several species. Groups may differ in size, longevity, the degree of heterogeneity and also in the way they are formed. We distinguish two processes of group formation, namely social and environmental processes. Social processes concern forms of social attraction, and environmental processes relate to attraction to food resources, shelter and the like. Below we discuss models for each process separately and for their combination.

### *22.2.1 Social Attraction*

Attraction is often mediated through visual and chemical cues. Visual attraction is important in many species of fish and birds. Chemical attraction (through pheromones) occurs among single celled individuals (such as bacteria and amoebae), ants and beetle larvae.

One of the simplest aggregation processes has been studied in certain experiments and models of cluster formation of beetle larvae (Camazine et al. 2001). In their natural habitat, these beetle larvae live in oak trees and grouping helps them to overcome the production of toxic resin by the host tree. They emit a pheromone and move towards higher concentrations of this chemical. In the experimental setup, larvae are kept in a Petri dish. In the model, the behavioural rules of following a gradient are represented in individuals that roam in an environment that resembles a Petri dish. It appears that both in the model and in the experiment the speed of cluster growth and the final number and distribution of clusters depend on the initial distribution and density of the larvae. By means of both model and experimental data it has been shown that cluster growth can be explained by a positive feedback. A larger group emits more pheromone and therefore attracts more members. Consequently, its size increases in turn emitting more pheromone, etcetera. This process is faster at a higher density, because individuals meet each other more frequently. Thus clusters appear sooner. Furthermore, growth is faster at higher densities, because more individuals are available for growing.

The location of the clusters has been studied. After starting from a random distribution of individuals, a single cluster remains in the centre of the Petri dish. This location is due to the higher frequency with which individuals visit the centre than the periphery. Starting from a peripheral cluster, there is an attraction to both the peripheral cluster and the centre. Thus, there is a kind of competition between clusters to attract additional individuals. The final distribution of clusters may consist of one cluster (at the centre, the periphery or at an intermediate location), or of two clusters (peripheral and in the centre). The final distribution depends on self-organisation and on three factors, namely, the initial density of individuals, initial distribution of clusters, and whether there is social attraction. Although the model is devised for the specific experimental setup and species, the mechanisms of group formation and growth shown in the model are found in many animal species. For instance, similar patterns are observed in models and experiments of cockroaches.

Cockroaches aggregate by sensing the relative humidity. They tend to move towards lower humidity. Here larger groups grow faster, because individuals have a higher tendency to stop and form a group if they collide with a larger number of others. Besides, they rest longer if more individuals are resting close by. They use the relative humidity as a cue to estimate the number of individuals. A lower relative humidity correlates with a higher number of individuals close by (Jeanson et al. 2005). Despite the different underlying process of social attraction (pheromonal, visual or based on relative humidity) pattern formation is similar to that of beetle larvae.

### *22.2.2 Foraging*

A model that deals with group formation solely through environmental causes concerns the splitting and merging of groups as it is found in the so-called fission-fusion system of primates, in particular that of spider monkeys (Ramos-Fernández et al. 2006). It relates the pattern of group formation to various distributions of food, because particularly in these fission-fusion societies, the food distribution may have an effect on grouping. In the model, the food distribution is based on a distribution of resources in forests that follows an inverse power law as described by Enquist and co-authors (Enquist and Niklas 2001; Enquist et al. 1999). The foragers maximise their food intake by moving to larger food patches that are closer by (they minimise the factor distance divided by patch size). Further, they do not visit patches that they have visited before. Individuals have no social behaviour and they meet in groups only because they accidentally visit the same food patch. For a distribution with both large and small trees (patches) in roughly equal numbers the model leads to a frequency distribution of subgroup sizes that resembles that of spider monkeys. Results hold, however, only if foragers have complete knowledge of the environment in terms of patch size and distance. The resemblance is not found if individuals only know (a random) part of the environment. Furthermore, if they have complete knowledge, individuals meet with certain partners more often than expected if the choice of the food patch is a random choice. In this way, certain networks of encounters develop. Another serious shortcoming of the model is that food can only be depleted; there is no food renewal and individuals do not return to former food patches. The main conclusion is that the ecological environment influences grouping: Ecologies that differ in the variation of their patch sizes (high, medium or low variance) also differ in the distribution of subgroup sizes. This model delivers mainly a proof of principle.

### *22.2.3 Combination of Social and Environmental Processes*

The relation between ecology and sub-group formation is also shown in a simulation study of orang-utans in the Ketambe forest in Indonesia (te Boekhorst and Hogeweg 1994a). The predictions for and resemblance to empirical data that are delivered in this model (te Boekhorst and Hogeweg 1994a) are more specific than those in the model of spider monkeys discussed above (Ramos-Fernández et al. 2006). Here, realistic patterns of grouping arise from simple foraging rules in interaction with the structure of the forest and in the presence of two social rules. The environment and population are built to resemble the Ketambe area in the composition of the community and the size of the habitat. Two main categories of food trees are distinguished, figs (very large and non-seasonal) and fruit trees (small and seasonal). There are 480,000 trees of which 1,200 bear fruit at the same time, once a year for ten consecutive days unless depleted earlier by the individuals. The crop size and spatial distribution of trees are specified: fig trees are clustered and fruit trees distributed randomly. Furthermore there are randomly distributed sources of protein. These are renewed immediately after being eaten. In the fruiting season extra individuals migrate into the area. With regard to the behavioural rules, individuals first search for a fig tree. If this is not found, while moving approximately in the direction of others, because other individuals may indicate the presence of food on a distant fig tree, individuals look for fruits close by. Upon entering a tree, the individual feeds until satiated or the tree is emptied. Next the individual rests close by and starts all over again later on. A further social rule causes adult males to avoid each other.

The resulting grouping patterns in the model resemble those of real orang-utans in many respects: Temporary aggregations are found in the enormous fig trees. This happens because in these big trees individuals may feed until satiated and then leave separately. However, when feeding in the much smaller fruit trees, food is insufficient and therefore individuals move to the next tree. This causes them to travel together. Thus, travelling bands develop mainly in the fruit season. In this season parties are larger than when fruit is scarce. Groupings emerge as a consequence of simple foraging rules in interaction with the forest structure. Thus, the model makes clear that differences between forest structures will lead to different grouping patterns in apes. This is a parsimonious explanation, because it is not necessary to think in terms of costs and benefits of sociality to explain these grouping patterns, rather they arise as a side effect of feeding behaviour. The empirical data analysis and ideas for the model of orang-utans were inspired by findings in another model on grouping in chimpanzees.

This model of chimpanzees concerns both foraging and social attraction and is meant to explain group formation, in particular the fission-fusion society (te Boekhorst and Hogeweg 1994b). It offers new explanations for the relatively solitary life of females and the numerous and large subgroups of males. Subgroups of males had so far been supposed to form in order to join in defending the community. To explain such cooperation, males were believed to be genetically

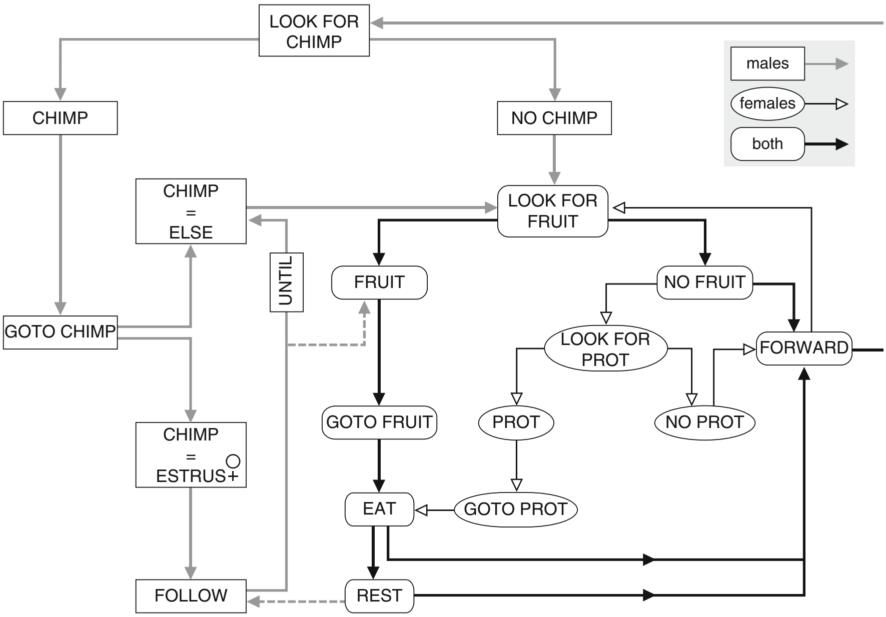



**Fig. 22.1** Behavioural rules of artificial chimpanzees (te Boekhorst and Hogeweg 1994b)

related to each other. Further, the solitary life of females was attributed to competition for food. However, the model shows that their patterns of grouping may arise in an entirely different way: they may result from a difference between the sexes in diet and in the priority of foraging versus reproduction. The model resembles a certain community of chimpanzees in Gombe in its community composition (number of males and females, 11 and 14 respectively, and the number of females that are synchronously in oestrus) and its habitat size (4 km$^2$), its number of trees, the number of trees that bear fruit at the same time, the length of their fruit bearing, their crop size, and the speed of their depletion. The number of protein resources is modelled more arbitrarily, approximating such forests. With regard to their behavioural rules, individuals of both sexes look for fruit and when satiated have a rest close to the tree (Fig. 22.1). If not satiated they move towards the next fruit tree. They continue to do this until they are satiated. If there is no fruit on a specific tree, females (but not males) look for proteins before searching another fruit tree. Males have one priority over finding food: finding females. Whenever they see a chimpanzee, they approach it to investigate whether it is a female in oestrus. If this is the case, males in the model follow her until she is no longer in oestrus.

In the model, patterns of grouping resemble those of real chimpanzees. Male groups emerge because all males have the same diet which differs from that of females and because they will follow the same female in oestrus together. Furthermore, in the model, as in real chimpanzees, males appear to travel over longer distances per day than females. They do so because they are travelling in larger groups and this leads to faster depletion of food sources and therefore they have to visit a larger number of trees in order to become satiated. This interconnection between the number of trees visited, group size and distance covered led to the hypothesis for orang-utans that different kinds of groupings (aggregation and travel bands) originate in different trees (fig and fruit). Note that these results were robust for a large range of different ecological variables and different compositions of the community. Here, a difference between the sexes in their diet and their priorities for sex and food appeared essential. With regard to the chimpanzee community, the authors conclude that, to explain its fission-fusion structure, the genetically based theory that kin-related males are jointly defending the community is not needed. In fact, in subsequent DNA studies no special genetic relatedness was found among cooperating male chimpanzees in Kibale (Goldberg and Wrangham 1997). Instead, chimpanzee-like party structures may emerge by self-organisation if chimpanzees search for food and for mates in a forest. Besides, the model can be used to explain the frequent bi-sexual groups observed in bonobos as being caused by their prolonged period of oestrus. Whereas these models have been controversial among primatologists for a long time, their usefulness is slowly becoming accepted (Aureli et al. 2008).

### *22.2.4 Group Coordination and Foraging*

Groups of social insects, for instance ants, are remarkably efficient in foraging. They collectively choose a food source that is closer rather than one (of the same quality) that is further away (Deneubourg and Goss 1989). Their choice is made without comparing the distance to different food sources. Experiments and models show that ants use trail pheromones as a mechanism of collective 'decision' making: Individuals mark their path with pheromone and at crossings follow the path that is more strongly marked. As they return to the nest sooner when the food source is close by, they obviously imprint the shorter path more often with pheromones. This results in a positive feedback: as the shorter path receives stronger markings, it also receives more ants, etcetera. Thus, the interaction between ants and their environment results in the adaptive and efficient exploitation of food sources. The 'preference' for a food source nearby rather than one further away is a side-effect of pheromonal marking. This marking also helps a single ant to find its way and may initially have been an adaptation to cause the ant to return to its nest. It is actually more than just that, because its intensity is adapted to the quality of the food source. The process of marking and following may also lead to mistakes. For instance, if a path is initially developed to a food source of low quality, and later to a food source of high quality is introduced elsewhere, due to 'historical constraint' the ants may remain fixated on the source of low quality even if it is further from the nest.

Army ants live in colonies of 200,000 individuals; they are virtually blind; they travel in a swarm with about 100,000 workers back and forth to collect food.

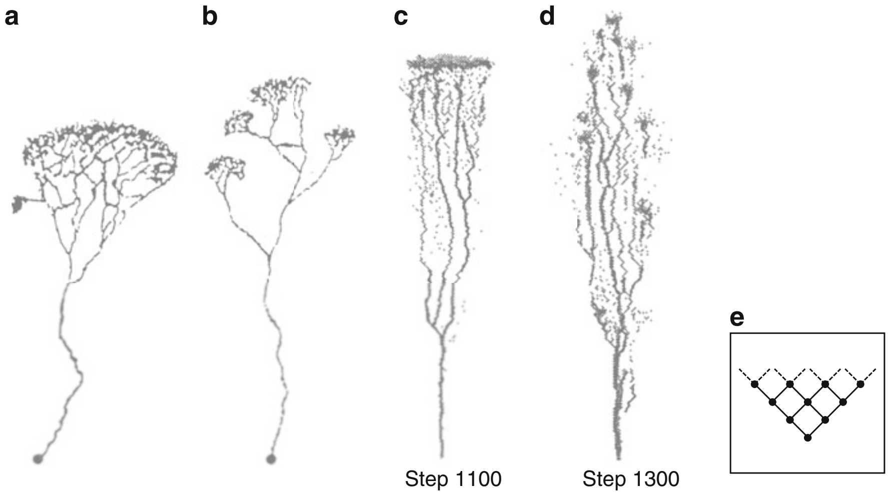



**Fig. 22.2** Foraging patterns of two species of army ants, Eciton burchelli and Eciton rapax, empirical data and models. Empirical data: Eciton burchelli (**a**) and Eciton rapax (**b**); models: (**c**) a few food clumps, (**d**) frequent occurrence of single food items, (**e**) network of nodes (After Deneubourg and Goss 1989)

Different species of army ants display highly structured patterns of swarming that may be species specific (Fig. 22.2a, b). For example, *Eciton burchelli* has a more dispersed swarm than *E. rapax*. Such species-specific differences in swarming are usually regarded as a separate adaptation, which is assumed to be based on corresponding differences in the underlying behavioural tendencies of coordination. Deneubourg and co-authors (Deneubourg and Goss 1989; Deneubourg et al. 1998), however, give a simpler explanation. In a model, they have shown that such markedly different swarming patterns may arise from a single rule-system of laying and following pheromone trails when ants are raiding food sources with different spatial distributions (Fig. 22.2c, d). The authors have shown this in a model in which ‘artificial ants’ move in a network of points (Fig. 22.2e), and mark their path with pheromones. When choosing between left and right, they prefer the more strongly marked direction. By introducing different distributions of food in the model (either uniformly distributed single insects or sparsely distributed colonies of insects), different swarming patterns arise from the interaction between the flow of ants heading away from the nest to collect food and the spatial distribution of the foragers returning with the food. These different swarm types are remarkably similar to those of the two species of army ants mentioned above (for empirical confirmation see Franks et al. 1991). Therefore, these different swarm forms reflect the variation in diet of the different species. Thus, the explanation of the model is more parsimonious than if we assume the different swarm forms to arise from a specific adaptation in rules of swarming. In summary, this model teaches us the effects of the environment on swarm coordination.

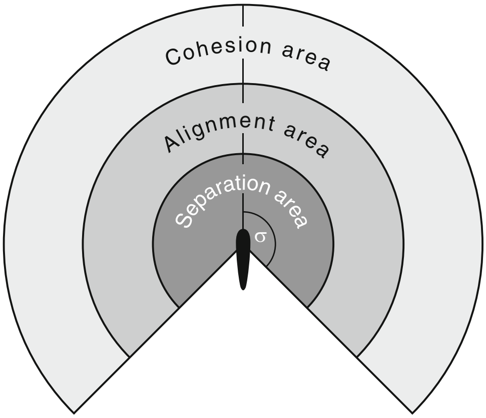



**Fig. 22.3** Behavioural areas of avoidance (separation), alignment and attraction (cohesion) with a dead ('blind') angle at the back (From Hemelrijk and Hildenbrandt 2008). The separation angle is indicated as $\sigma$

With regard to its evolution, natural selection shapes the necessary traits for the successful marking and following of trails depending on the size of the food source and other environmental characteristics (for an evolutionary model of this, see Solé et al. 2001).

### 22.2.5 Group Coordination in a Homogeneous Environment

Even in environments (such as the open sea, savannah and sky) that are virtually uniform without environmental structure, remarkable coordination is observed in the swarms of many animal species, e.g. of insects, fish, birds and ungulates. Coordination appears flexible even in swarms of a very large size (for instance, of up to ten million individuals in certain species of fish). Swarming behaviour has been modelled in several ways. The most simplistic representations of emergent phenomena have used partial differential equations. Slightly more complex behaviour has been obtained using particle-based models derived from statistical mechanics in physics (Vicsek et al. 1995). These models have been used to explain the phase transition of unordered to ordered swarms in locusts (Buhl et al. 2006). Yet the biologically most relevant results come from models wherein individuals coordinate with their local neighbours by following only three rules based on zones of perception (Fig. 22.3): They avoid neighbours that are close by (separation), align to others up to an intermediate distance (alignment) and approach those further away (cohesion). These models have been applied to describe herds of ungulates (Gueron et al. 1996), schools of fish (Couzin et al. 2002; Hemelrijk and Kunz 2005; Huth and Wissel 1992, 1994; Kunz and Hemelrijk 2003; for a review see Parrish and Viscido 2005), and swarms of birds (Hildenbrandt et al. 2010; Reynolds 1987; Hemelrijk & Hildenbrandt 2011; Hemelrijk & Hildenbrandt 2012).

Through these models we obtain insight into a number of important biological aspects of swarming, which have mainly been related to schools of fish.

Firstly, we get insight into the *coordination* of schools. Schools coordinate with remarkable flexibility even into the absence of a *leader* and without a directional preference. The direction of their movement is merely the consequence of the location and heading of others in the school. With regard to the question whether there is a leader in a swarm, such a leader fish is supposed to be located at the front (Bumann and Krause 1993). However, neither in such models of fish schools nor in real swarms individuals appear to have a fixed location. Instead frontal locations are continuously switched in the model on average every 2 s, in real fish every 1.4 s (Gnatapogon elongates, Huth and Wissel 1994). Thus, there cannot be consistent leaders.

Furthermore, if a number of individuals have a directional preference (for instance, for certain food sources or breeding locations), but most of them do not, those with such a preference will automatically lead the school. Huse and co-authors (Huse et al. 2002) showed that even if the percentage of individuals that has a certain preferred direction is very small (though above 7 %), this may influence the direction of the entire school.

If individuals prefer a different direction, for instance, because they aim to go to different food locations, the school may react differently depending on the degree to which two preferred directions differ (Couzin et al. 2005): If the directions differ little, the group will follow the average direction between the two ('a compromise'). If the angle between both directions is large, the group will either follow the direction that is preferred by the majority or in the absence of a 'convincing majority' it will randomly choose one of the two. In these examples of swarming the models help us to understand the processes that determine the direction in which a school is heading.

Secondly, we obtain insight into the *segregation* of individuals that differ in a certain trait. In a school, individuals may be segregated, for example, according to size. Usually this is attributed to an intentional or genetic preference for being near individuals of the same size or body form. Models show, however, that this segregation may also arise directly as a side-effect of differences in body size without any preference (e.g. see Couzin et al. 2002; Hemelrijk 2005; Kunz and Hemelrijk 2003). This may, for instance, arise because larger individuals due to their larger body have a larger range at which they avoid others who are too close. Thus, by avoiding smaller individuals more often than the reverse, large individuals may end up at the periphery leaving the small ones in the centre (as has been found in water insects (Romey 1995)).

Thirdly, natural schools of fish show a number of traits that are believed to be helpful in the *protection against predators*: Their shape is oblong and their density is highest at the front. Bumann et al. (1997) argue that an *oblong form* and *high frontal density* protect against predation: the oblong shape reduces the size of the frontal area, where predators are supposed to attack and high frontal density protects individuals against approaching predators. Hemelrijk and co-authors (Hemelrijk and Hildenbrandt 2008; Hemelrijk and Kunz 2005) have noted that it is unlikely that individual fish actively organise themselves so as to create these two patterns. Therefore, the authors studied in a model whether these patterns might arise by self-organisation as a side-effect of their coordinated movements. This indeed appeared to be the case and their emergence appeared to be robust (independent of school size and speed). These patterns come about because, during travelling, individuals avoid collisions mostly by slowing down (as they do not perceive others to be directly behind them, in the so-called blind angle, Fig. 22.3). Consequently, they leave a gap between their former forward neighbours and subsequently these former neighbours move inwards to be closer together. Thus, the school becomes oblong.

Furthermore, when individuals fall back, a loose tail builds up and this automatically leaves the highest density at the front. In the model it appears that larger schools are relatively more oblong because they are denser and so more individuals fall back to avoid collision. Faster schools appear to be less oblong. This arises because fast individuals have greater difficulty to turn, thus, the path of the school is straighter, the school is more aligned (polarised) and, therefore, fewer individuals fall back. Consequently, the core of a faster school is denser and the tail is looser than they are in slower schools. Recent tests in real fish (mullets) confirm the specific relationships between group shape, and its size, density, and polarisation as found in the model (Hemelrijk et al. 2010). Although this indicates that the shape in real schools develops in a similar way, it is still necessary to investigate shape and frontal density in more species and to study the effects of different speeds on these traits.

Fourth, in real animals predation and attacks on swarms result in a spectacular range of behavioural *patterns of evasion by schooling prey*. These patterns are supposed to confuse the predator. They have been labelled 'tight ball', 'bend, 'hourglass', 'fountain effect', 'vacuole', 'split', 'join', and 'herd'. They have been described for schools of several prey species and predators (Axelsen et al. 2001; Lee 2006; Nottestad and Axelsen 1999; Parrish 1993; Pitcher and Wyche 1983). Most of these patterns may be obtained in a model by simple behavioural rules of prey and predator (e.g., see Inada and Kawachi 2002). Many of the different patterns of evasion result in self-organisation in models of schooling which are built in such a way that upon detecting the predator, individuals compromise between their tendency to avoid the predator, and to coordinate with their group members. Though these models do not exactly fit real data, they give us insight into how specific collective evasion patterns may arise.

## 22.3 Social Organisation

Although groups may be beneficial for their members in so far as they provide protection against predation, they also result in competition for food, mates and space. If individuals meet for the first time, such competitive interactions may initially have a random outcome. Over time, however, a dominance hierarchy develops, whereby certain individuals are consistently victorious over others and are said to have a higher dominance value than others (Drews 1993). Individuals may perceive the dominance value of the other from the body posture of the other (in primates) or from their pheromone composition (in insects).

With regard to the question which individual becomes dominant, there are two extremely opposing views: dominance as a fixed trait by inherited predisposition and dominance by chance and self-organisation. While some argue for the importance of predisposition of dominance by its (genetic) inheritance (Ellis 1991), others reject this for the following reasons: Experimental results show that the dominance of an individual depends on the order of its introduction in a group (Bernstein and Gordon 1980). Dominance changes with experience, because the effects of victory and defeat in conflicts are self-reinforcing, the so-called winner-loser effect. This implies that winning a fight increases the probability of victory in the next fight and losing a fight increases the probability of defeat the next time. This effect has been established empirically in many animal species and is accompanied by psychological and physiological changes, such as hormonal fluctuations (Bonabeau et al. 1996a; see Chase et al. 1994; Hemelrijk 2000; for a recent review, see Hsu et al. 2006; Mazur 1985).

The self-reinforcing effects of fighting are the core of a model called *DomWorld*, which concerns individuals that group and compete (Hemelrijk 1996, 1999b; Hemelrijk and Wantia 2005). It leads to a number of emergent phenomena that have relevance for many species, in particular for primates and more specifically macaques.

### 22.3.1 The Basic DomWorld Model

The model *DomWorld* consists of a homogeneous world in which individuals are grouping and competing. It does not specify what they compete about. Grouping is implemented by making individuals react to others in different manners depending on the distance to the other (Fig. 22.4). An individual is attracted to others far away (within MaxView), it continues to follow its direction if it perceives others at an intermediate distance (within NearView), and it decides whether or not to perform a competitive dominance interaction if it encounters others close by (within PerSpace) (Hemelrijk 2000). After winning a dominance interaction, it chases away the other and after losing a fight, it flees from it.

Dominance interactions are implemented after the DoDom interactions by Hogeweg (Hogeweg and Hesper 1985), which were extended to reflect differences in the intensity of aggression between species and between the sexes and with the decision whether or not to attack depending on risks involved (Hemelrijk 1998, 1999b). Each individual has a dominance value which indicates the individual's capacity to win. At the beginning of a competitive interaction both individuals display their dominance value and observe that of the other. The outcome of the fight depends on the relative dominance of both partners and on chance.

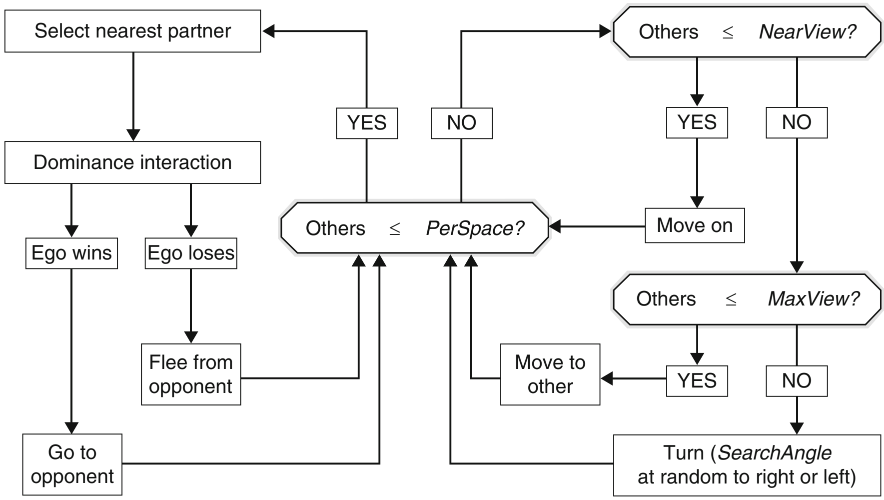



**Fig. 22.4** Flowchart of the behavioural rules of individuals in DomWorld. PerSpace, NearView and MaxView indicate close, intermediate and long distances, respectively

The probability of winning is higher the higher the dominance value of an individual in relation to that of the other. Initially, the dominance values are the same for all individuals. Thus, during the first encounter, chance decides who wins. After winning, the dominance of the winner increases and that of the loser decreases. Consequently, the winner has a greater chance to win again (and vice versa) which reflects the self-reinforcing effects of the victories (and defeats) in conflicts of real animals.

We allow for rank reversals; when, unexpectedly, a lower-ranking individual defeats a higher-ranking opponent, this outcome has a greater impact on the dominance values of both opponents, which change with a greater amount than when, as we would expect, the same individual conquers a lower-ranking opponent (conform to detailed behavioral studies on bumble bees by van Honk and Hogeweg 1981). Furthermore, in their decision whether or not to attack, we made individuals sensitive to the risks, i.e. the 'will' to undertake an aggressive interaction (instead of remaining non-aggressively close by) increases with the chance to defeat the opponent, which depends on the relative dominance ranks of both opponents (Hemelrijk 1998).

We also represented the intensity of aggression in which primate societies differ (in some species individuals bite, in other species they merely approach, threaten and slap) as a scaling factor (called StepDom) that weighs the changes in dominance value after a fight more heavily if the fight was intense (such as biting) than if the fight was mild (involving threats and slaps, or merely approaches and retreats) (Hemelrijk 1999b). In several models, we distinguished two types of individuals representing males and females (Hemelrijk et al. 2003). We gave males a higher initial dominance value and a higher intensity of aggression (reflecting their larger

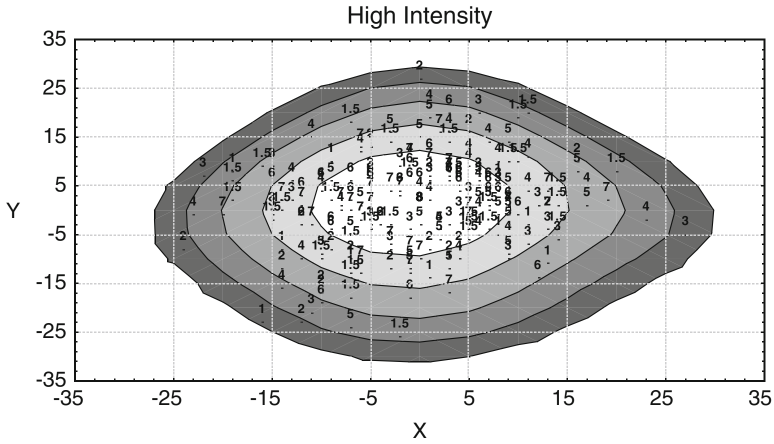



**Fig. 22.5** Spatial-social structure. *Darker shading* indicates areas with individuals of decreasing dominance rank

body size, stronger muscular structure and larger canines than those of females). In fights between both types of individuals, the intensity of the fight was determined by the initiator (the attacker).

### *22.3.2 Spatial Structure*

The major advantage of group life is supposedly to be protection against predators. Central positions are supposed to be safest, because here individuals are shielded off by other group-members from predators approaching from the outside. Therefore, according to the well-known 'selfish herd' theory of Hamilton (1971), individuals have evolved a preference for a position in the centre, the so-called 'centripetal instinct'. If competition for this location is won by dominants, dominants will end up in the centre. This is thought to be the main reason why in many animal species dominants are seen to occupy the centre. However, in DomWorld this spatial structure emerges, even though such a preference is lacking and there is no 'centripetal instinct' nor threat of predation (Hemelrijk 2000).

The spatial configuration, with dominant individuals in the centre and subordinates at the periphery of the group (Fig. 22.5), emerges in the model due to a feedback between the dominance hierarchy and the spatial location of the individuals of different rank. During the development of the hierarchy, some individuals become permanent losers. Such low-ranking individuals must end up at the periphery by being constantly chased away. This automatically leaves dominants in the centre. Also, in real animals, such a spatial structure may occur although its members have no centripetal instinct nor experience a threat of predation. For instance, in the elegant experiments with fish by Krause (1993), central dominants were observed, although no centre-oriented locomotion appeared (Krause and Tegeder 1994). Furthermore, this spatial structure has been described in hammerhead sharks in spite of the absence of any predatory threat (Klimley 1985). Thus, the model provides a new way of understanding spatial structure.

### *22.3.3 Dominance Style: Egalitarian and Despotic Societies*

High dominance ranking is supposed to be associated with benefits such as priority of access to mates, food and safe locations. If benefits are strongly biased towards higher-ranking individuals, the society is called 'despotic', whereas if access to resources is more equally distributed, it is called 'egalitarian'. These terms have been used to classify social systems of many animal species (such as insects, birds and primates). Egalitarian and despotic species of primates, such as macaques, appear to differ in many other traits too, such as in group density, their intensity and frequency of aggression and in their frequency and patterns of affiliation (grooming). Usually these differences are explained from the perspective of optimisation of single traits by natural selection. However, Thierry (1990b) suggests that in macaques the many behavioural differences can be traced back to two inherited differences, namely degree of nepotism (i.e. cooperation among kin) and intensity of aggression. Note that despotic macaques display aggression of a higher intensity, i.e. they bite more often, whereas egalitarian macaques display aggression that is milder, they only threaten and slap.

The model *DomWorld* presents an even simpler hypothesis (Hemelrijk 1999b), namely that a mere difference in intensity of aggression produces both types of societies. By increasing the value of one parameter, namely that of intensity of aggression, the artificial society switches from a typically egalitarian society to a despotic one. For instance, compared to egalitarian artificial societies, despotic ones are more loosely grouped, showing a higher frequency of attack, their behaviour is more rank-related, aggression is more asymmetric, spatial centrality of dominants is clearer and female dominance over males is greater. All these differences between despotic and egalitarian societies arise via feedback between the development of the hierarchy and spatial structure and this happens only at a high intensity of aggression. The steep hierarchy develops as a consequence of the high aggression intensity, because each outcome has a stronger impact than at a low intensity and it is strengthened further via a mutual feedback between the hierarchy and the spatial structure with dominants in the center and subordinates at the periphery (Hemelrijk 1999b, 2000).

Pronounced rank-development causes low-ranking individuals to be continuously chased away by others and thus the group spreads out (1 in Fig. 22.6). Consequently, the frequency of attack diminishes among the individuals (2 in Fig. 22.6) and therefore the hierarchy stabilizes (3 in Fig. 22.6). While low-ranking individuals flee from everyone, this automatically leaves dominants in the center, and thus a spatial-social structure develops (Fig. 22.5). Since individuals of similar

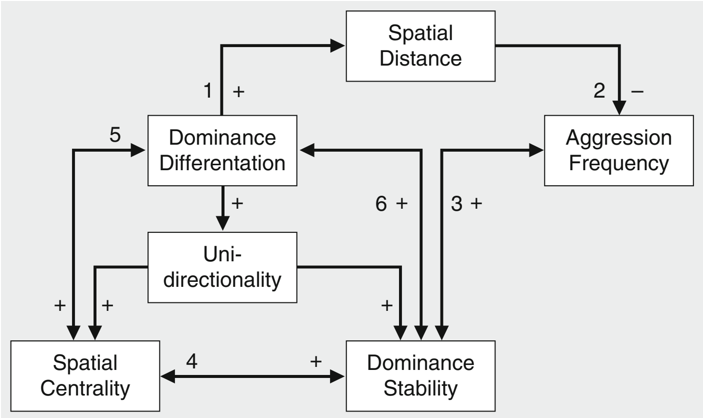



**Fig. 22.6** Interconnection between variables causing spatial-social structuring at a high aggression intensity

dominance are treated by others in more or less the same way, similar individuals remain close together; therefore, they interact mainly with others of similar rank; thus, if a rank reversal between two opponents occurs, it is only a minor one because opponents are often similar in dominance. In this way the spatial structure stabilizes the hierarchy and it maintains the hierarchical differentiation (4 and 5 in Fig. 22.6). Also, the hierarchical differentiation and the hierarchical stability mutually strengthen each other (6 in Fig. 22.6).

In short, the model (Hemelrijk 1999b) makes it clear that changing a single parameter representing the intensity of aggression, may cause a switch from an egalitarian to a despotic society. Since all the differences resemble those found between egalitarian and despotic societies of macaques, this implies that in real macaques these differences may also be entirely due to a single trait, intensity of aggression. Apart from intense aggression, such as biting, however, a high frequency of aggression can also cause this switch (Hemelrijk 1999a). A higher frequency of aggression also leads to a steeper hierarchy. This in turn, results in a clearer spatial structure which again strengthens the development of the hierarchy and this has the cascade of consequences as described for aggression intensity.

One of the shortcomings of the model may be considered as the lack of individual recognition among group members. In a subsequent model, this was corrected by having each individual keep a record of the dominance value of each group member. This value was updated depending on its experiences gained with other group members (Hemelrijk 1996, 2000). With regard to the development of spatial structure, hierarchy and dominance style remained similar, but patterns were weaker than in the case of a direct perception of dominance without individual recognition. Weaker patterns arise due to the contradictory experiences that different individuals have with specific others. This will impede a hierarchical development and thus weaken the accompanying consequences. Even though this may be more realistic for certain modelling purposes, it is more useful to have clearer patterns in a simpler model. Such a caricature is more helpful for building upon understanding and developing new ideas.

Dominance style is usually considered to be species specific, but Preuschoft and colleagues (Preuschoft et al. 1998) raised the question whether competitive regimes (egalitarian versus despotic) should not rather be considered as sex specific. In their study of the competitive regime of both sexes of Barbary macaques, they found an unexpected sex difference: males behave in an egalitarian way whereas females are despotic. It is unexpected that the sex with the larger body size and fiercer aggression evolved a more egalitarian dominance style. Therefore, it seems to be a separate adaptation. However, the same difference in dominance style between the sexes was also found in DomWorld: males appear to be more egalitarian than females. The unexpectedly stronger egalitarianism of males in the model is due to yet another difference between the sexes, the higher initial dominance of males compared to females (which relates to differences in body size). Consequently, single events of victory and defeat have less impact on their overall power or dominance. Therefore, they lead to less hierarchical differentiation than among females, who are much smaller and weaker and on whom each victory and defeat therefore has more impact. The greater the sexual difference in initial dominance between the sexes, the more egalitarian the males behave among themselves compared to the behaviour of the females among themselves. The conclusion of this study is that the degree of sexual dimorphism may influence the competitive regime of each sex, in the model and in real primates. Further empirical studies are needed to investigate whether the degree of sexual dimorphism is directly proportional to the steepness of the gradient of the hierarchy of females compared to males.

With regard to its evolution, dominance style is supposed to be a consequence of different degrees of competition within and between groups (van Schaik 1989). According to this theory, when competition within groups is high and that between groups is low, despotic societies evolve and if it is reversed, and competition between groups is high, egalitarian groups emerge. In line with this, DomWorld already shows that high competition within a group leads to a despotic society (Hemelrijk 1999b). However, in a subsequent study, competition between groups appears to favour despotic rather than egalitarian groups (Wantia 2007). To study effects of competition between groups, the DomWorld model was extended to several groups (i.e. GroupWorld, Wantia 2007). Here, as in real primates, usually individuals of high rank participate in encounters between groups (e.g. see, Cooper 2004).

The model generates a number of unexpected results. Firstly, among groups of the same dominance style, competition between groups does not affect dominance style, since it happens at a very low frequency compared to competition within groups. However, in competition between groups of different dominance style, remarkable results came to light. Unexpectedly, under most conditions groups with a more despotic dominance style were victorious over others with a more egalitarian style. This arose due to the greater power of individuals of the highest rank of the despotic group compared to that of the egalitarian group. In the model, this is a consequence of the steeper hierarchy in despotic groups compared to that in egalitarian groups. In reality this effect may be even stronger, because higher-ranking individuals may also obtain relatively more resources in a despotic group than in an egalitarian one. The outcome of fights between groups depends, however, on the details of the fights between groups and the composition of the group. When participants of inter-group fights fought in dyads or in coalitions of equal size, the despotic group out-competed the egalitarian one. If, however, individuals of egalitarian groups, for one reason or another, fought in coalitions of a larger size or if their coalitions included more males than those of the despotic groups, the egalitarian group had a chance to win. Thus, the main conclusion of the study is that it depends on a number of factors simultaneously, which dominance style will be favoured. Therefore, this model suggests that group composition and details of what happens in fights between groups should be studied in order to increase our understanding of the origination of dominance style.

### *22.3.4 Distribution of Affiliation, Grooming*

Grooming (to clean the pelage of dirt and parasites) is supposed to be altruistic and therefore it should be rewarded in return (Trivers 1971). Using the theory of social exchange, Seyfarth (1977) argues that female primates try to exchange grooming for receipt of support in fights. For this, they direct the supposedly altruistic grooming behaviour more towards higher- than towards lower-ranking partners. As a consequence of this supposed competition for grooming partners of high rank, and by being defeated by dominant competitors, females will end up grooming close-ranking partners most frequently and are groomed themselves most often by females ranking just below them. According to him, this explains the following grooming patterns that are apparent among female primates: (a) high ranking individuals receive more grooming than others and (b) most grooming takes place between individuals that are adjacent in rank.

DomWorld presents a simpler alternative for the explanation by Seyfarth (Hemelrijk 2002b) in which a mental mechanism of social exchange for future social benefits is absent, for newer updates see Puga-Gonzalez et al. (2009), Hemelrijk and Puga-Gonzalez (2012). This is important, because it is doubtful whether grooming involves any real costs at all (see Wilkinson 1988).

The core of the argument is that individuals more often groom those whom they encounter more frequently. In the case of spatial structure with dominants in the centre and individuals in closer proximity to those that are closer in dominance (as described for primates (e.g. see Itani 1954)), these patterns will follow automatically. Individuals more often groom those that are nearby in rank (as found by Seyfarth 1977). Further, because dominants are more often in the centre, they are frequently surrounded on all sides by others. Subordinates, however, are at the edge of the group, and therefore have nobody on one side. Consequently, dominants simply meet others more frequently than subordinates do (as is shown in

DomWorld Hemelrijk 2000). Therefore, dominants are more often involved in grooming than subordinates (as found by Seyfarth).

In support of the model, grooming patterns in despotic macaques appear to be more dominance oriented, and in egalitarian species grooming is more often directed at anyone (Thierry et al. 1990). To establish the relevance of this model-based hypothesis for real primates, it should be tested further in more species and groups, whether or not the patterns of grooming among individuals of similar rank and of the receipt of grooming by individuals of higher rank occur especially in groups with centrally located dominants, and less so in those with a weak spatial structure.

### *22.3.5 Dominance Relationships Between the Sexes*

Most primate species live in bi-sexual groups. Apart from the order of Lemuriformes, males are usually dominant. However, variation occurs and even if males are larger than females, females are sometimes dominant over a number of them (Smuts 1987). To study whether this may arise through chance and the self-reinforcing effects of dominance interactions, the sexes are modelled in DomWorld. When, for the sake of simplicity, the sexes in DomWorld are distinguished only in terms of an inferior fighting capacity of females as compared to that of males, then, surprisingly, males appear to become less dominant over females at a high intensity of aggression than at a low intensity (Hemelrijk 1999b; Hemelrijk et al. 2003). This is due to the stronger hierarchical differentiation, which causes the hierarchy to develop in a more pronounced way. This implies that the hierarchy is also more strongly developed for each sex separately. Since this differentiation increases the overlap in dominance between the sexes, it causes more females to be dominant over males than in the case of a hierarchy that has been less developed (Fig. 22.7a, b).

Similarly to females, adolescent males of despotic macaques have more ease than those of egalitarian species in outranking adult females (Thierry 1990a). Thierry explains this as a consequence of the stronger cooperation to suppress males among related females of despotic macaques than of egalitarian ones, and van Schaik (1989) emphasizes the greater benefits associated with coalitions for females of despotic species than egalitarian ones. However, DomWorld explains greater female dominance, as we have seen, simply as a side-effect of more pronounced hierarchical differentiation.

Differences in overlap of ranks between the sexes may affect sexual behavior. Males of certain species, such as Bonnet macaques, have difficulty in mating with females to which they are subordinate (Rosenblum and Nadler 1971). Therefore, following the model, we expect that despotic females (because they outrank more males) have fewer males to mate with than egalitarian ones. In line with this, in macaques despotic females are observed to mate with fewer partners and almost exclusively with males of the highest ranks (Caldecott 1986); this observation is

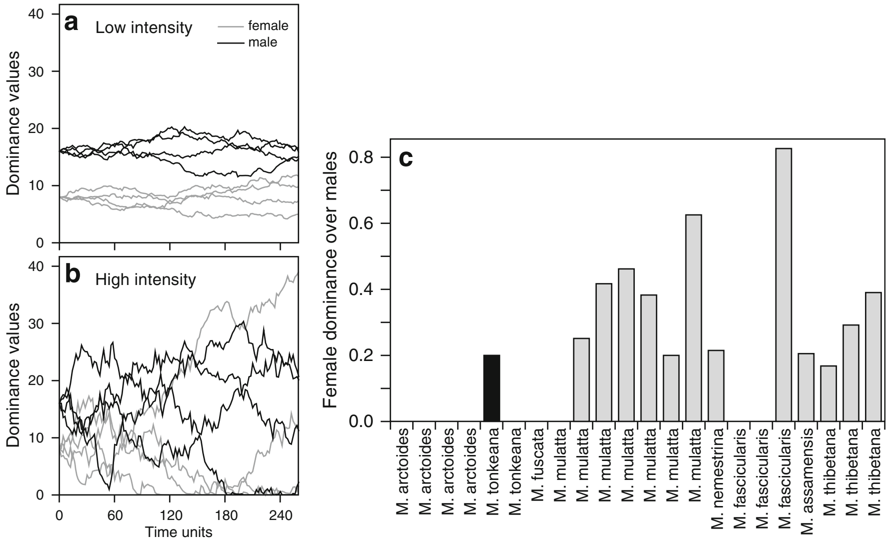



**Fig. 22.7** Female dominance over males or degree of rank overlap: (**a**, **b**) female dominance over males over time and aggression intensity (Hemelrijk 1999b), (**c**) degree of rank overlap of females with males (i.e. female dominance over males) in groups of egalitarian macaques (in black) and in groups of despotic macaques (in grey) (Hemelrijk et al. 2008)

attributed to the evolution of a more pronounced female preference in despotic than in egalitarian species of macaques. The explanation derived from the model, however, is simpler. In a subsequent study, in support of this hypothesis, the relative dominance position of both sexes in egalitarian and despotic macaque species indeed appeared to differ in macaques as expected: females of despotic species were dominant over a significantly higher percentage of the males in their group than females of egalitarian species (Fig. 22.7c).

In a similar way as high aggression intensity, a high frequency of aggression in the model also results in more female dominance. A higher frequency of aggression in the model can be obtained by increasing the SearchAngle over which individuals search for others (Fig. 22.4). Due to a greater SearchAngle they return to the others sooner, the group becomes denser and thus the frequency of aggression is higher. A difference in group density may be of relevance to the difference in female dominance observed between common chimpanzees and bonobos (also known as pygmy chimpanzees (Stanford 1998)). Despite their similar sexual dimorphism, female dominance in pygmy chimpanzees is higher than in common chimpanzees. This is usually attributed to more intensive coalition formation among pygmy females against males. However, in line with DomWorld, we may also explain it as a side-effect of the difference in density between both species (Hemelrijk 2002a, 2003). Density is high in groups of pygmy chimpanzees. Due to the higher density there is a higher frequency of aggression and according to DomWorld this may result in more female dominance over males. This hypothesis should be tested by

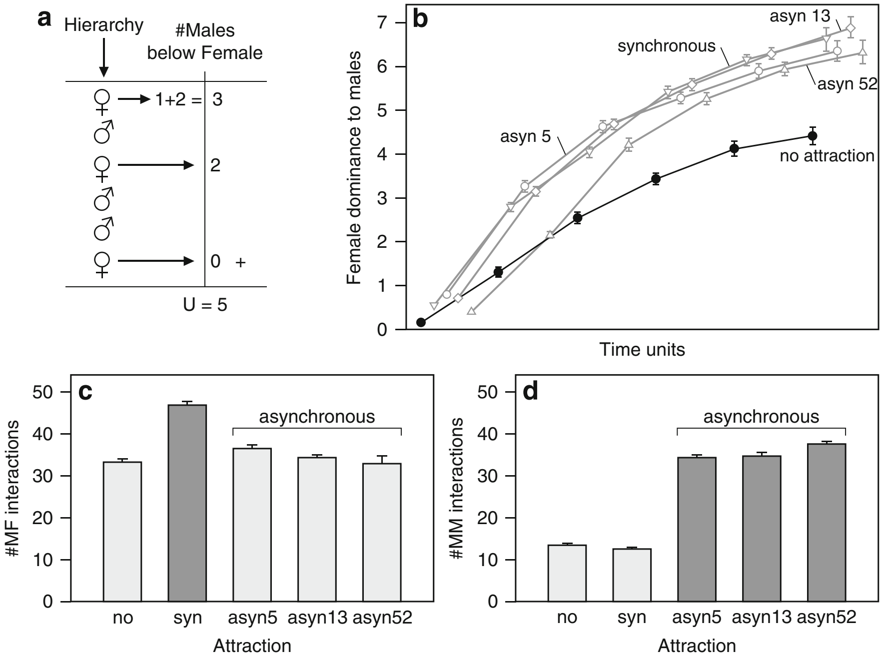



**Fig. 22.8** Female dominance over males and sexual attraction. (**a**) Measurement of female dominance; (**b**) female dominance over males over time without attraction (control) and when attracted to females that cycle synchronously and asynchronously; (**c**) number of interactions between sexes with(out) attraction during (**a**) synchronously cycling; (**d**) number of interactions among males with(out) attraction during (**a**) synchronously cycling. *Asyn* asynchronous cycles, *syn* synchronous cycling; 5, 13, 52 are duration of estrus period

comparing different groups of bonobos and by studying the relationship between female dominance and frequency of aggression.

Sexual attraction in real animals is usually thought to be accompanied by strategies of exchange. For instance, chimpanzee males are described as exchanging sex for food with females (Goodall 1986; Stanford 1996; Tutin 1980). Yet, in spite of detailed statistical studies, we have found no evidence that males obtain more copulations with, or more offspring from those females with whom they share their food more often (Hemelrijk et al. 1992, 1999, 2001; Meier et al. 2000). Male tolerance of females seems to increase during the females' period of estrus even without noticeable benefits. Thus, we need another explanation for male tolerance of females. DomWorld provides us with such an alternative hypothesis. 'Sexual attraction' of males to females is implemented in such a way that males have a greater inclination to approach females than individuals of their own sex. In the model (and in the preceding models and empirical studies of Fig. 22.7), we measure the relative dominance position of females compared to males by counting the number of males ranking below each female and calculating this figure (Mann–Whitney U-value, Fig. 22.8a). It appears that this value of relative female dominance to males increases with sexual attraction as an automatic consequence of the more frequent encounters between the sexes (Fig. 22.8b, synchronous). This result is in line with the observation that female dominance in chimpanzees increases when males are sexually attracted to the females (Yerkes 1940). The question of whether female dominance over males also increases during sexual attraction in other species should be studied in the future.

Whereas the examples mentioned above concern species in which females are synchronously sexually attractive (tumescent), in other species they cycle asynchronously. In the model, however, female dominance over males is relatively similar regardless of whether they are attractive synchronously or asynchronously (Fig. 22.8b, synchronous, asyn). The process leading to increased female dominance differs, however, for the two conditions. If single females are attractive in turn, many males cluster close to a single female. Consequently, in contrast to synchronous tumescence, the frequency of interaction between the sexes remains similar to that when females are not attractive to males, but the frequency of male-male interactions is increased markedly (Fig. 22.8c, d). Due to the higher frequency of interactions among males, the differentiation of the male hierarchy is stronger than without attraction and this causes certain males to become subordinate to some females (Fig. 22.8b, asyn).

Furthermore, the adult sex ratio (or percentage of males) in the group influences the relative dominance of females compared to that of males (Hemelrijk et al. 2008). Female dominance appears to be higher when there are more males in the group (Fig. 22.9a). This arises from a shift in the relative number of intra- and intersexual interactions. A higher proportion of males causes both sexes to interact more often with males. Due to the males' higher intensity of aggression, this causes a greater differentiation of the dominance values of both females and males. Consequently, at a high intensity of aggression, the hierarchy of females overlaps more with that of males and thus, the dominance position of females is higher in relation to males than if there are fewer males in the group. Subsequent analysis of these patterns in real primates has confirmed that female dominance increases with a higher percentage of males in the group (Fig. 22.9b). It appeared that in line with the preceding modelling results, in groups of a despotic species (rhesus macaques) a higher percentage of males appeared to be correlated with greater female dominance over them, whereas such a correlation was absent among groups of an egalitarian species (stump-tailed macaques).

### *22.3.6 Strategies of Attack*

When real animals are brought together for the first time, they perform dominance interactions only during a limited period. This has been empirically established in several animal species, e.g. chickens (Guhl 1968) and primates (Kummer 1974). The interpretation is that individuals fight to reduce the ambiguity of their relationships (Pagel and Dawkins 1997); once these are clear, energy should be saved. On the other hand, it has also been suggested that individuals should

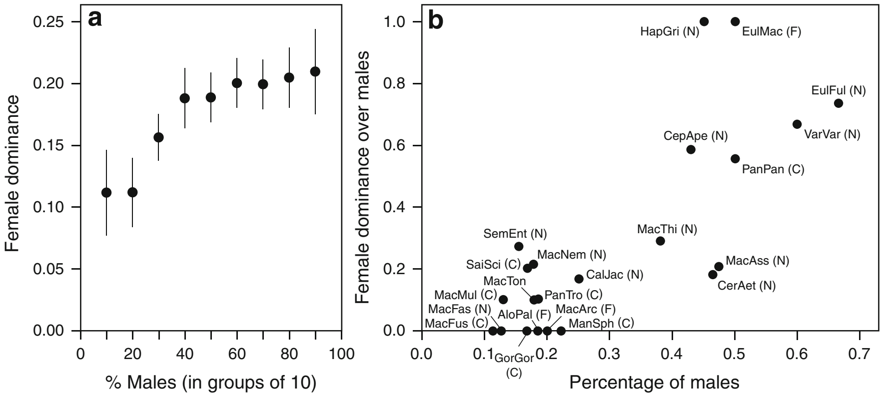



**Fig. 22.9** Percentage of males in the group and female dominance over males. (**a**) Model (average and SE), (**b**) real data of primates. 6-letter codes indicate species. Environmental conditions: *N* natural, *F* free-ranging, *C* captive condition

continuously strive for a higher ranking and therefore always attack, unless an opponent is clearly believed to be superior (e.g. see Datta and Beauchamp 1991).

In the DomWorld model, we compare these popular ethological views with each other and a control strategy, in which individuals invariably attack others upon meeting them (this is called the 'Obligate' attack strategy). Here, the 'Ambiguity-Reducing' strategy is a symmetrical rule in which individuals are more likely to attack opponents that are closer in rank to themselves. In the so-called 'Risk Sensitive' strategy, the probability of an attack is higher when the opponent is of a lower rank (Hemelrijk 1998). Remarkably, it appears that, with time, the frequency of aggression decreases in all three attack strategies, at least when groups are cohesive and the intensity of aggression is sufficiently high (Fig. 22.10).

This decrease of aggression is a direct consequence of the 'Ambiguity-Reducing' strategy, but unexpectedly it also develops in the other two. Because of the high intensity of aggression each interaction has a strong impact on the rank of both partners, and thus, a steep hierarchy develops. This automatically implies that some individuals become permanent losers and that, by fleeing repeatedly, they move further and further away from others (bottom row in Fig. 22.10). The increased distance among individuals in turn results in a decrease of the frequency of encounters and hence aggression. This provides a test for the real world: it has to be examined whether the development of the dominance hierarchy is accompanied not only by a reduction of aggression but also by an increase in inter-individual distances.

## 22.3.7 Personality Types

Two of these attack strategies, i.e. 'risk-sensitive' and 'obligate' attack (Hemelrijk 2000) resemble the attack strategies of individuals of different personality types,

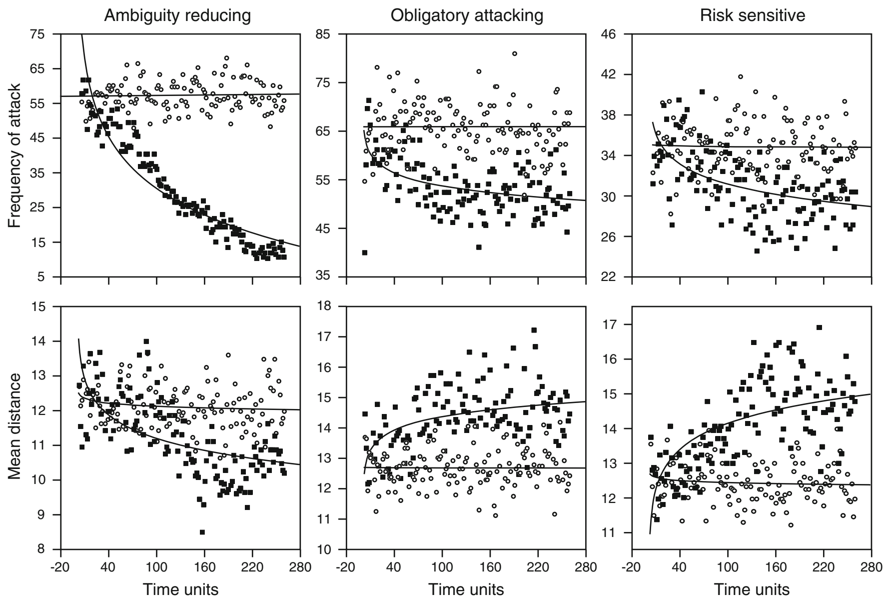



**Fig. 22.10** Development of frequency of aggressive interactions *(top half)* and mean distance *(lower half)* among individuals for different attack strategies and intensities of aggression (logarithmic line fitting). *Open circles* represent StepDom of 0.1,*closed blocks* of 1.0

namely those of the cautious and the bold personality, respectively (Koolhaas 2001). When groups in the model consist of both types (mixed groups), the differentiation of dominance values appears to be greater among individuals that are attacking obligatorily than risk-sensitively due to their higher frequency of attack (Hemelrijk 2000). Consequently, obligate attackers rise very high and descend very low in the hierarchy (resulting in a bimodal distribution of dominance values, fat lines in Fig. 22.11a), whereas risk-sensitive attack leads to less variation, a unimodal distribution of values (fat lines in Fig. 22.11a), and therefore to more intermediate dominance positions. Further, among risk-sensitive individuals, the average dominance is slightly higher than among those that always attack (obligatorily). This is due to higher 'intelligence' of the risk-sensitive attack-strategy, because these individuals will attack, especially when the risk of losing the fight is minimal.

This resembles the distribution of dominance in mixed groups of a bird species, the great tits (Verbeek et al. 1999). Here, bold individuals become very high up in the dominance hierarchy or descend very low, whereas cautious individuals have intermediate ranks that on average are above those of bold individuals. Differences in high and low rank were explained by different stages in which the individuals were. With regard to the moulting of their feathers and to a difference in tendency to attack from a familiar territory or an unfamiliar one and to a difference in speed of recovery from defeats. DomWorld shows that there is no need to add causes based on different stages of moulting, or on familiarity to a territory or to different speed of recovery. Thus, the model produces a far simpler explanation for the distribution of dominance

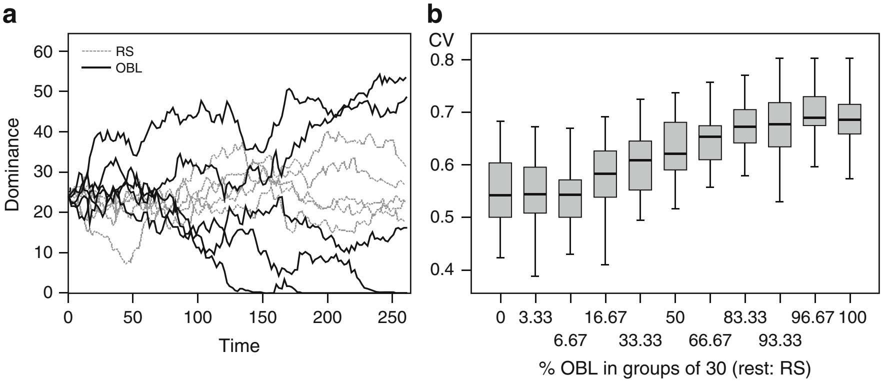



**Fig. 22.11** Dominance distribution and personality types: (**a**) example of hierarchical development of mixed group (*fat lines*: obligate attackers, *dotted lines*: risk-sensitive individuals). N = 10, 5 of each type. (**b**) hierarchical differentiation in mixed groups with different ratios of obligately attacking (OBL) and risk-sensitive (RS) individuals. *CV* Mean coefficient of variation of DomValues. *Box* = S.E., *whiskers* = S.D

values in these groups of great tits. To verify these results empirically, differences in risk sensitivity of both types of personality need to be confirmed empirically.

Secondly, the model provides us with an alternative explanation for the associations between dominance behavior and personality style in great tits found by Dingemanse and de Goede (2004). This association appears to differ among individuals that own a territory and those that do not; whereas among territory owners bolds were dominant over cautious ones, the reverse held for those without a territory. To explain this the authors use a context-specific argument in which they need an additional trait, namely speed of recovery from defeat (Carere et al. 2001). They argue that particularly among those individuals without a territory, bolds have more difficulty in recovering from defeat than cautious ones and that therefore, they become low in rank, whereas territory-owners do not suffer this setback and, therefore, they become higher in rank.

Alternatively, a simpler explanation, in line with our model, may apply. Here, we start from existing dominance relationships and suppose that these exert a decisive influence on the question who will obtain a territory (instead of the other way around). We assume that, because territories are limited in numbers, the higher-ranking individuals (say the top half of them) will acquire them, whereas individuals in the lower part of the hierarchy are unable to get one. Due to the bi-modal distribution of dominance values among the bold birds, and the uni-modal distribution of the cautious ones, the most extreme dominance positions in the colony will be occupied by bold-ones, and the cautious are located in the middle of the hierarchy. Thus, among individuals in the top half of the hierarchy (the territory-owners) the bolds will rank above the cautious, whereas in the bottom half of the hierarchy, namely among the individuals without a territory, the reverse is true (Fig. 22.11a). For this explanation to be proven correct, we must verify whether territory owners belong to the upper half of the dominance hierarchy or not.

Thus, DomWorld produces new explanations for dominance relationships of these 'personality styles' in great tits.

An important question regarding personality is how different types, bold and cautious may co-exist, and why one type did not take over. Although there are a number of explanations for various species, none of them applies to primates. Since in primates group survival and individual survival depend on competition within and between groups, Wantia and Hemelrijk (Wantia 2007) have studied the two personality types in these contexts. They have found that risk-sensitive individuals out-competed obligate-attackers in fights within groups, but that in fights between groups the obligate attackers did better: the higher the percentage of individuals that attacked obligatorily in fights between groups, the greater the chance of the group winning (Fig. 22.11b). The better performance within groups of risk-sensitive individuals was due to their more cautious and deliberate strategy: to attack when the chance of winning was high. Greater success by obligate attackers in fights between groups was a consequence of the higher dominance value of the highest ranking individuals in groups with more obligate attackers. This is due to the steeper hierarchy as a consequence of the higher frequency of aggression in groups with more individuals that carry out obligatory attacks. Thus, whereas risk-sensitive individuals out-compete obligate-attackers in conflicts within groups, the reverse happens in conflicts between groups. Since competition within and between groups is essential for primate societies (van Schaik and van Hooff 1983), and the success of both attack strategies depends on these contexts, we may imagine that a similar differential performance may contribute to the co-existence of bold and cautious primates.

### *22.3.8 Distribution of Tasks*

It seems a miracle that a colony of social insects consisting of tens of thousands of individuals is able to cope with the huge socio-economic demands of foraging, building, cleaning nests and nursing the brood. It appears that group-members are somehow able to divide the work efficiently among them. Such a division of labour is flexible, i.e. the ratio of workers performing different tasks varies according to changes in the needs and circumstances of the colony. There are several ways in which this task-division may arise. Different mechanisms may operate in different species (for a review, see Beshers and Fewell 2001). Task-division may be based on a genetic difference in predisposition (e.g. Moritz et al. 1996; Robinson 1998; Robinson and Page 1988), or a response-threshold to perform certain tasks (Bonabeau et al. 1996b). Such a threshold may be combined with a self-reinforcing learning process (Gautrais et al. 2002; Theraulaz et al. 1998). Thus, after performing, the threshold is lessened (by learning) and after a long period of no involvement in the task, it increases by forgetting.

The execution of tasks may also be a consequence of dominance values, as is shown in a model of bumblebees (Bombus terrestris) developed by Hogeweg and

Hesper (1983, 1985). This was based on an earlier experimental study (van Honk and Hogeweg 1981) that showed that during the growth of the colony, workers develop into two types, the low-ranking, so-called 'common', and the high-ranking, so-called 'elite', workers. The activities carried out by the two types differ noticeably: Whereas the 'common' workers mainly forage and take rest, the 'elite' workers are more active, feed the brood, interact with each other and with the queen, and sometimes lay eggs. In their study of the minimal conditions needed for the formation of the two types of workers (on the assumption that all workers are identical when hatching), Hogeweg and Hesper (1983, 1985) used an individual-based model based on biological data concerning the time of the development of eggs, larvae, pupae, etc. Space in the model is divided into two parts, peripheral (where inactive common workers doze for part of the time) and central (where the brood is, and all interactions take place). The rules of the artificial, adult bumblebees operate 'locally' in so far as their behaviour is triggered by what they encounter. What they encounter in the model is chosen randomly from what is available in the space in which the bumblebee finds itself. For instance, if an adult bumblebee meets a larva, it feeds it, if it meets a pupa of the proper age, it starts building a cell in which a new egg can be laid, etc. All workers start with the same dominance value after hatching, with only the queen gifted with a much higher dominance rank. When an adult meets another, a dominance interaction takes place, the outcome of which (victory or defeat) is self-reinforcing. Dominance values of the artificial bumblebees influence almost all their behavioural activities (for instance, individuals of low rank are more likely to forage).

This model automatically, and unexpectedly, generates two stable classes, those of 'commons' (low-ranking) and 'elites' (high-ranking) with their typical conduct. This differentiation only occurs if the nest is divided into a centre and a periphery (as in real nests).

The flexibility of the distribution of tasks among individuals manifests when we take half the work force out of the model. In line with observations in similar experiments with real bumblebees, this reduction in work force causes the remaining ones to take over the work. In the model this arises because the decreased number of workers reduces the frequency of encounters among them and increases encounters between them and the brood (which has not been reduced in number). An increased rate of encounters with brood induces workers to collect food more frequently. Therefore, workers are absent from the nest more often and, consequently, they meet other workers less frequently.

In real bumblebees the queen switches from producing sterile female offspring to fertile offspring (males and females) at the end of the season. Note that whereas females are produced by fertilised eggs, males are produced from unfertilised eggs. Usually females will develop into sterile workers, but if they are fed extra 'queenfood' during larval development, they will develop into queens. The switch seems to take place at an optimal moment, because it occurs at the time when the colony is at its largest and can take care of the largest number of offspring. Oster and Wilson (1978) point out that it is difficult to think of a possible external signal that could trigger such a switch, because it takes 3 weeks to raise fertile offspring and during these 3 weeks there must still be enough food.

Hogeweg and Hesper (1983) discovered that no such external signal is needed in their bumble bee model, but that the switch originates automatically as if scheduled by a socially regulated 'clock'; it arises from the interaction between colony growth and stress development of the queen as follows. During the development of the colony, the queen produces a certain pheromone that inhibits the extra feeding behaviour of larvae by 'elite' workers (that lead to queens) and worker-ovipositions (i.e. unfertilised male eggs). Just before she is killed, she can no longer suppress the 'elite' workers from feeding larvae to become queens and from laying drone-eggs, because the colony has grown too large. Consequently, individual workers meet the queen less often and are less subjected to the dominance of the queen, so they start to lay unfertilised (drone) eggs. Furthermore, the stress on the queen increases whenever she has to perform dominance interactions with workers during her egg laying. When the stress on the queen has reached a certain threshold value she switches to producing male eggs (drones).

Because generative offspring are also sometimes found in small nests, van der Blom (1986) challenges the notion that dominance interactions induce stress in the queen and thus lead to this switch. However, in the model, it is not the number of workers that causes the switch: Hogeweg and Hesper (1985) have shown that in small nests the switch also appears to occur at the same time in the season as in large nests. For this they studied the bumblebee model for a reduced speed of growth. They found that the switch occurs at a similar moment due to the following complicated feedback process. If the colony grows faster, the heavy duties of caring for the brood leave the workers little time to interact with each other and the queen. Consequently, the dominance hierarchy among workers only develops weakly. Therefore, if workers interact with the queen they do not pose much of a threat to her and as a result, the queen is not severely stressed and the colony can grow very large. In contrast, in a slowly growing colony, the small number of brood gives the workers little work and leaves them time to interact with each other. Consequently, their dominance relationships are clearly differentiated. Furthermore, by often interacting with the queen, they become on average higher-ranking themselves. In the end, the queen receives as much stress from frequent interactions with a few high-ranking workers in a slowly growing nest as from few interactions with very many low-ranking workers in a fast growing nest. Consequently, the switch to reproductive offspring takes place at about the same moment in both nests in the model.

## 22.4 Evaluation

The power of these kinds of models is the generation of phenomena that are emergent. These emergent phenomena lead to explanations that are more parsimonious than usual, because the patterns emerge from the interaction among individuals and their environment rather than from the cognition of an individual. These explanations can be tested empirically.

Considering the kind of questions posed in the models discussed above, it becomes clear that most of them can only be studied using certain kinds of agent-based models. Other kinds of models, such as partial differential equations based on density functions or even individual-based models of fluids and gases cannot incorporate the complexity of the rules and/or emergent effects.

The behavioural rules of the agents in the agent-based models that were treated here were in all cases biologically inspired. In some cases, behavioural rules were based on precise experimental results specific to a certain species, such as in the case of group formation (Camazine et al. 2001; Jeanson et al. 2005). Usually, however, parameters were tuned only loosely to the real system (e.g. the angle of vision in fish is set at about 270° and that of mammals at 120°). Sometimes, mathematical equations were used that were developed for a different species, such as in the case of dominance interactions. Here, the equations were initially derived for the self-reinforcing effects in bumblebees (van Honk and Hogeweg 1981) and subsequently extended and applied to explain social systems in primates (Hemelrijk 1999b). In all cases the behavioural rules are supposed to capture the essentials of those of real animals.

If, however, macro patterns are obtained that resemble patterns observed in the real system, this is still no proof of the correctness of the rules in reflecting behavioural rules of real animals, because different sets of rules may lead to the same macro-pattern. The relevance of the rules and other aspects of the model could be studied further by investigating additional hypotheses and comparing further results of the model to natural systems.

Agent-based models appear to be particularly useful in showing the consequences of interactions among individuals for social structure and the reverse, i.e. from social structure to social interactions. In this way, they teach us about the integration of traits. This is particularly apparent when the models include a representation of space (time is always represented). The representation of time and space is of general importance, because all real individuals live in time and space and thus it is unrealistic to ignore these aspects. It appears that by representing the interaction of individuals with their social and physical environment in time and space, patterns result in such a way that explanations become unusually parsimonious in terms of their specific adaptive traits and the cognition needed (in line with findings of situated cognition Pfeifer and Scheier 1999).

These explanations give a new insight into the phenomenon of social complexity that may either be very general (e.g. conceptual model) or specific to the type of animal and the kind of behaviour under investigation. Usually, it leads to hypotheses that are easily tested in the real system. These hypotheses make unexpected connections between variables, such as between the production of sterile and reproductively-active offspring with colony dynamics in bumblebees. Testing can be done by natural observation, for example, by studying the effects of different group sizes on group shape in case of fish. Furthermore, testing can be done by experimental interventions, such as by putting together individuals of both sexes in different group compositions to study effects on relative dominance of females to males.

While ignoring models of certain topic areas where no emergent effects appeared (so-called output-oriented models), we have here surveyed a large number of the most important models of social complexity. Of course, it is impossible to discuss them all. Note that we have not treated models concerning biological evolution since only a few are related to social systems (such as Kunz et al. 2006; Oboshi et al. 2003; Ulbrich et al. 1996). This is due to the fact that evolution usually happens on a far larger time scale than phenomena of social complexity. However, see Chap. 18 in this handbook (Chattoe-Brown and Edmonds 2013) on how concepts from biological evolution have influenced recent approaches in programming simulation models.

Furthermore, we have not treated a number of other complex phenomena, such as synchronisation of lighting behaviour of insects, temperature regulation of nests and the building of nests, because in these cases no agent-based models have been used, or the behaviour is insufficiently social. These topics are, however, covered by Camazine and co-authors (Camazine et al. 2001).

## 22.5 Future Work

With regard to areas that may be important for future work, I suggest (1) models of self-organisation that help us understand specific social systems better by being closely tuned to empirical systems, such as fish schools, insect colonies or primate groups, (2) models that present social behaviour and its ontogenetical development in greater detail and (3) evolutionary models that include also spatial effects.

**Acknowledgments** I would like to thank Andreas Flache and Rineke Verbrugge for comments on an earlier draft. I thank Daan Reid for correcting the English.

## Further Reading

For further reading I recommend the book on self-organisation in biological systems by Camazine and co-authors (Camazine et al. 2001). This forms an extensive introduction to almost all topics treated above and more (with the exception of task division and social systems of primates). Furthermore, new extensions of a number of the models are treated by Sumpter (2006). For teaching I suggest the well-written book by Resnick (1994). This book has been used in secondary schools and teaches to think in terms of complexity and self-organisation in general. For more advanced readers, I recommend "Self-organisation and Evolution of Social Systems" (Hemelrijk 2005). This is an edited book and contains recent articles on modelling of social systems, including those of humans.

## References

Aureli F et al (2008) Fission-fusion dynamics: new research frameworks. Curr Anthropol 49:627–654
Axelsen BE, Anker-Nilssen T, Fossum P, Kvamme C, Nottestad L (2001) Pretty patterns but a simple strategy: predator–prey interactions between juvenile herring and Atlantic puffins observed with multibeam sonar. Can J Zool – Rev Canadienne De Zoologie 79:1586–1596
Bernstein IS, Gordon TP (1980) The social component of dominance relationships in rhesus monkeys (Macaca mulatta). Anim Behav 28:1033–1039
Beshers SN, Fewell JF (2001) Models of division of labour in social insects. Annu Rev Entomol 46:413–440
Blom Jvd (1986) Reproductive dominance within the colony of Bombus terrestris. Behaviour 97:37–49
Bonabeau E, Theraulaz G, Deneubourg J-L (1996a) Mathematical models of self-organizing hierarchies in animal societies. Bull Math Biol 58:661–717
Bonabeau E, Theraulaz G, Deneubourg J-L (1996b) Quantitative study of the fixed threshold model for the regulation of division of labour in insect societies. Proc R Soc Lond B 263:1565–1569
Buhl J et al (2006) From disorder to order in marching locusts. Science 312:1402–1406
Bumann D, Krause J (1993) Front individuals lead in shoals of 3-spined sticklebacks (Gasterosteus-Aculeatus) and Juvenile Roach (Rutilus-Rutilus). Behaviour 125:189–198
Bumann D, Krause J, Rubenstein D (1997) Mortality risk of spatial positions in animal groups: the danger of being in the front. Behaviour 134:1063–1076
Caldecott JO (1986) Mating patterns, societies and ecogeography of macaques. Anim Behav 34:208–220
Camazine S et al (2001) Self-organization in biological systems. Princeton University Press, Princeton/Oxford, Princeton studies in complexity
Carere C, Welink D, Drent PJ, Koolhaas JM, Groothuis TGG (2001) Effect of social defeat in a territorial bird (Parus major) selected for different coping styles. Physiol Behav 73:427–433
Chase ID, Bartelomeo C, Dugatkin LA (1994) Aggressive interactions and inter-contest interval: how long do winners keep winning? Anim Behav 48:393–400
Chattoe-Brown E, Edmonds B (2013) Evolutionary mechanisms. Chapter 18 in this volume
Cooper MA (2004) Inter-group relationships. In: Thierry B, Singh M, Kaumanns W (eds) Macaque societies: a model for the study of social organization. Cambridge University Press, Cambridge
Couzin ID, Krause J, James R, Ruxton GD, Franks NR (2002) Collective memory and spatial sorting in animal groups. J Theor Biol 218:1–11
Couzin ID, Krause J, Franks NR, Levin SA (2005) Effective leadership and decision-making in animal groups on the move. Nature 433:513–516
Datta SB, Beauchamp G (1991) Effects of group demography on dominance relationships among female primates. 1. Mother-daughter and sister-sister relations. Am Nat 138:201–226
Deneubourg JL, Goss S (1989) Collective patterns and decision-making. Ethol Ecol Evol 1:295–311
Deneubourg JL, Goss S, Franks N, Pasteels JM (1998) The blind leading the blind: modelling chemically mediated army ant raid patterns. J Insect Behav 2:719–725
Dingemanse NJ, De Goede P (2004) The relation between dominance and exploratory behavior is context-dependent in wild great tits. Behav Ecol 15:1023–1030
Drews C (1993) The concept and definition of dominance in animal behaviour. Behaviour 125:283–313
Ellis L (1991) A biosocial theory of social stratification derived from the concepts of pro/antisociality and r/K selection. Polit Life Sci 10:5–44
Enquist BJ, Niklas KJ (2001) Invariant scaling relations across tree-dominated communities. Nature 410:655–660

Enquist BJ, West GB, Charnov EL, Brown JH (1999) Allometric scaling of production and life-history variation in vascular plants. Nature 401:907–911

Franks NR, Gomez N, Goss S, Deneubourg J-L (1991) The blind leading the blind in army ant raiding patterns: testing a model of self-organization (Hymenopter: Formicidae). J Insect Behav 4:583–607

Gautrais J, Theraulaz G, Deneubourg JL, Anderson C (2002) Emergent polyethism as a consequence of increased colony size in insect societies. J Theor Biol 215:363–373

Goldberg TL, Wrangham RW (1997) Genetic correlates of social behaviour in wild chimpanzees: evidence from mitochondrial DNA. Anim Behav 54:559–570

Goodall J (1986) The chimpanzees of Gombe: patterns of behaviour. Belknapp Press of Harvard University Press, Cambridge MA/London

Gueron S, Levin SA, Rubenstein DI (1996) The dynamics of herds: from individuals to aggregations. J Theor Biol 182:85–98

Guhl AM (1968) Social inertia and social stability in chickens. Anim Behav 16:219–232

Hamilton WD (1971) Geometry for the selfish herd. J Theor Biol 31:295–311

Hemelrijk CK (1996) Dominance interactions, spatial dynamics and emergent reciprocity in a virtual world. In: Maes P, Mataric MJ, Meyer J-A, Pollack J, Wilson SW (eds) Proceedings of the fourth international conference on simulation of adaptive behavior, vol 4. The MIT Press, Cambridge, MA, pp 545–552

Hemelrijk CK (1998) Risk sensitive and ambiguity reducing dominance interactions in a virtual laboratory. In: Pfeifer R, Blumberg B, Meyer J-A, Wilson SW (eds) From animals to animats 5, Proceedings of the fifth international conference on simulation on adaptive behavior, Zurich, 17–21 Aug 1998. MIT Press, Cambridge, MA, pp 255–262

Hemelrijk CK (1999) An individual-oriented model on the emergence of despotic and egalitarian societies. Proc Royal Soc Lond B Biol Sci 266:361–369

Hemelrijk CK (1999a) Effects of cohesiveness on intersexual dominance relationships and spatial structure among group-living virtual entities. In: Floreano D, Nicoud J-D, Mondada F (eds) Advances in artificial life. Fifth European conference on artificial life (Lecture notes in artificial intelligence, 1674). Springer, Berlin, pp 524–534

Hemelrijk CK (1999b) An individual-oriented model on the emergence of despotic and egalitarian societies. Proc R Soc Lond B Biol Sci 266:361–369

Hemelrijk CK (2000) Towards the integration of social dominance and spatial structure. Anim Behav 59:1035–1048

Hemelrijk CK (2002a) Self-organising properties of primate social behaviour: a hypothesis on intersexual rank-overlap in chimpanzees and bonobos. In: Soligo C, Anzenberger G, Martin RD (eds) Primatology and anthropology: into the third millennium. Wiley, New York, pp 91–94

Hemelrijk CK (2002b) Understanding social behaviour with the help of complexity science. Ethology 108:655–671

Hemelrijk CK (2003) Female co-dominance in a virtual world: ecological, cognitive, social and sexual causes. Behaviour 140:1247–1273

Hemelrijk CK (ed) (2005) Self-organisation and evolution of social systems. Cambridge University Press, Cambridge

Hemelrijk CK, Hildenbrandt H (2008) Self-organized shape and frontal density of fish schools. Ethology 114:245–254

Hemelrijk CK, Kunz H (2005) Density distribution and size sorting in fish schools: an individual-based model. Behav Ecol 16:178–187

Hemelrijk CK, Wantia J (2005) Individual variation by self-organisation: a model. Neurosci Biobehav Rev 29:125–136

Hemelrijk CK, van Laere GJ, van Hooff J (1992) Sexual exchange relationships in captive chimpanzees? Behav Ecol Sociobiol 30:269–275

Hemelrijk CK, Meier CM, Martin RD (1999) 'Friendship' for fitness in chimpanzees? Anim Behav 58:1223–1229

Hemelrijk CK, Meier CM, Martin RD (2001) Social positive behaviour for reproductive benefits in primates? A response to comments by Stopka et al. (2001). Anim Behav 61:F22–F24

Hemelrijk CK, Wantia J, Daetwyler M (2003) Female co-dominance in a virtual world: ecological, cognitive, social and sexual causes. Behaviour 140:1247–1273

Hemelrijk CK, Wantia J, Isler K (2008) Female dominance over males in primates: self-organisation and sexual dimorphism. PLoS One 3, e2678

Hemelrijk CK, Hildenbrandt H, Reinders J, Stamhuis EJ (2010) Emergence of oblong school shape: models and empirical data of fish. Ethology 116:1099–1112

Hemelrijk CK, Hildenbrandt H (2011) Some causes of the variable shape of flocks of birds. PLoS One 6:e22479. doi:10.1371/journal.pone.0022479

Hemelrijk CK, Hildenbrandt H (2012) Schools of fish and flocks of birds: their shape and internal structure by self-organization. Interface Focus 2:726–737. doi:10.1098/rsfs.2012.0025

Hemelrijk CK, Puga-Gonzalez I (2012) An individual-oriented model on the emergence of support in fights, its reciprocation and exchange. PLoS One 7:e37271. doi:10.1371/journal.pone.0037271

Hildenbrandt H, Carere CC, Hemelrijk CK (2010) Self-organized aerial displays of thousands of starlings: a model. Behav Ecol 21(6):1349–1359

Hogeweg P, Hesper B (1983) The ontogeny of interaction structure in bumble bee colonies: a MIRROR model. Behav Ecol Sociobiol 12:271–283

Hogeweg P, Hesper B (1985) Socioinformatic processes: MIRROR modelling methodology. J Theor Biol 113:311–330

Hsu Y, Earley RL, Wolf LL (2006) Modulation of aggressive behaviour by fighting experience: mechanisms and contest outcomes. Biol Rev 81:33–74

Huse G, Railsback S, Ferno A (2002) Modelling changes in migration pattern in herring: collective behaviour and numerical domination. J Fish Biol 60:571–582

Huth A, Wissel C (1992) The simulation of the movement of fish schools. J Theor Biol 156:365–385

Huth A, Wissel C (1994) The simulation of fish schools in comparison with experimental data. Ecol Model 75(76):135–145

Inada Y, Kawachi K (2002) Order and flexibility in the motion of fish schools. J Theor Biol 214:371–387

Itani J (1954) The monkeys of Mt Takasaki. Kobunsha, Tokyo

Jeanson R et al (2005) Self-organized aggregation in cockroaches. Anim Behav 69:169–180

Klimley AP (1985) Schooling in Sphyrna lewini, a species with low risk of predation: a non-egalitarian state. Zeitschrift fürTierpsychology 70:279–319

Koolhaas JM (2001) How and why coping systems vary among individuals. In: Broom DM (ed) Coping with challenge: welfare including humans. Dahlem University Press, Berlin

Koolhaas JM, de Boer SF, Buwalda B, van der Vegt BJ, Carere C, Groothuis AGG (2001) How and why coping systems vary among individuals. In: Broom DM (ed) Coping with challenge. Welfare including humans. Dahlem University Press, Berlin

Krause J (1993) The effect of 'Schreckstoff' on the shoaling behaviour of the minnow: a test of Hamilton's selfish herd theory. Anim Behav 45:1019–1024

Krause J, Tegeder RW (1994) The mechanism of aggregation behaviour in fish shoals: individuals minimize approach time to neighbours. Anim Behav 48:353–359

Kummer H (1974) Rules of dyad and group formation among captive baboons (theropithecus gelada). S. Karger, Basel

Kunz H, Hemelrijk CK (2003) Artificial fish schools: collective effects of school size, body size, and body form. Artif Life 9:237–253

Kunz H, Züblin T, Hemelrijk CK (2006) On prey grouping and predator confusion in artificial fish schools. In: Rocha LM et al (eds) Proceedings of the tenth international conference of artificial life. MIT Press, Cambridge, MA, pp 365–371

Lee SH (2006) Predator's attack-induced phase-like transition in prey flock. Phys Lett A 357:270–274

Mazur A (1985) A biosocial model of status in face-to face primate groups. Soc Forces 64(2):377–402

Meier C, Hemelrijk CK, Martin RD (2000) Paternity determination, genetic characterization and social correlates in a captive group of chimpanzees (Pan troglodytes). Primates 41:175–183

Moritz RFA, Kryger P, Allsopp MH (1996) Competition for royalty in bees. Nature 384(6604), 31

Nottestad L, Axelsen BE (1999) Herring schooling manoeuvres in response to killer whale attacks. Can J Zool – Rev Canadienne de Zoologie 77:1540–1546

Oboshi T, Kato S, Mutoh A, Itoh H (2003) Collective or scattering: evolving schooling behaviors to escape from predator. In: Standish RK, Bedau M, Abbass HA (eds) Artificial life VIII. MIT Press, Cambridge, MA, pp 386–389

Oster GF, Wilson EO (1978) Caste and ecology in social insects. Princeton University Press, Princeton

Pagel M, Dawkins MS (1997) Peck orders and group size in laying hens: 'future contracts' for non-aggression. Behav Processes 40:13–25

Parrish JK (1993) Comparison of the hunting behavior of 4 piscine predators attacking schooling prey. Ethology 95:233–246

Parrish JK, Viscido SV (2005) Traffic rules of fish schools: a review of agent-based approaches. In: Hemelrijk CK (ed) Self-organisation and the evolution of social behaviour. Cambridge University Press, Cambridge, pp 50–80

Pfeifer R, Scheier C (1999) Understanding intelligence. MIT Press, Cambridge, MA

Pitcher TJ, Wyche CJ (1983) Predator avoidance behaviour of sand-eel schools: why schools seldom split? In: Noakes DLG, Lindquist BG, Helfman GS, Ward JA (eds) Predators and prey in fishes. Dr. W. Junk Publications, Dordrecht, pp 193–204

Preuschoft S, Paul A, Kuester J (1998) Dominance styles of female and male Barbary maceques (Macaca sylvanus). Behaviour 135:731–755

Puga-Gonzalez I, Hildenbrandt H, Hemelrijk CK (2009) Emergent patterns of social affiliation in primates, a model. PLoS Comput Biol 5:e1000630. doi:10.1371/journal.pcbi.1000630

Ramos-Fernández G, Boyer D, Gómez VP (2006) A complex social structure with fission-fusion properties can emerge from a simple foraging model. Behav Ecol Sociobiol 60(4):536–549

Resnick M (1994) Turtles, termites and traffic jams: explorations in massively parallel microworlds. MIT Press, Cambridge, MA

Reynolds CW (1987) Flocks, herds and schools: a distributed behavioral model. Comput Graph 21:25–36

Robinson GE (1998) From society to genes with the honey bee. Am Scientist 86:456–462

Robinson GE, Page RE (1988) Genetic determination of guarding and undertaking in honey-bee colonies. Nature 333:356–358

Romey WL (1995) Position preferences within groups: do whirligigs select positions which balance feeding opportunities with predator avoidance? Behav Ecol Sociobiol 37:195–200

Rosenblum LA, Nadler RD (1971) The ontogeny of sexual behavior in male bonnet macaques. In: Ford DH (ed) Influence of hormones on the nervous system. Karger, Basel, pp 388–400

Seyfarth RM (1977) A model of social grooming among adult female monkeys. J Theor Biol 65:671–698

Smuts BB (1987) Gender, aggression and influence. In: Smuts BB, Cheney DL, Seyfarth RM, Wrangham RW, Struhsaker TT (eds) Primate societies. Chicago University Press, Chicago, pp 400–412

Solé RV, Bonabeau E, Delgado J, Fernádez P, Marín J (2001) Pattern formation and optimization in army ant raids. Artif Life 6:219–226

Stanford CB (1996) The hunting ecology of wild chimpanzees: implications for the evolutionary ecology of Pliocene hominids. Am Anthropol 98:96–113

Stanford CB (1998) The social behaviour of chimpanzees and bonobos. Curr Anthropol 39:399–420

Sumpter DJT (2006) The principles of collective animal behaviour. Philos Trans R Soc B Biol Sci 361:5–22 te Boekhorst IJA, Hogeweg P (1994a) Effects of tree size on travel band formation in orang-utans: data-analysis suggested by a model. In: Brooks RA, Maes P (eds) Artificial life, vol IV. The MIT Press, Cambridge, MA, pp 119–129

te Boekhorst IJA, Hogeweg P (1994b) Self structuring in artificial 'CHIMPS' offers new hypotheses for male grouping in chimpanzees. Behaviour 130:229–252

Theraulaz G, Bonabeau E, Deneubourg J-L (1998) Threshold reinforcement and the regulation of division of labour in social insects. Proc R Soc Lond B 265:327–333

Thierry B (1990a) Feedback loop between kinship and dominance: the macaque model. J Theor Biol 145:511–521

Thierry B (1990b) The state of equilibrium among agonistic behavior patterns in a group of Japanese macaques (Macaca fuscata). C R Acad Sci Paris 310:35–40

Thierry B, Gauthier C, Peignot P (1990) Social grooming in Tonkean macaques (Macaca tonkeana). Int J Primatol 11:357–375

Trivers RL (1971) The evolution of reciprocal altruism. Q Rev Biol 46:35–57

Tutin CEG (1980) Reproductive behaviour of wild chimpanzees in the Gombe National Park. J Reprod Fertil Suppl 28:43–57

Ulbrich K, Henschel J R, Jeltsch F, Wissel C (1996) Modelling individual variability in a social spider colony (Stegodyphus dumicola: Eresidae) in relation to food abundance and its allocation. Revue suisse de Zoologie (Proceedings of the XIIIth international congress of arachnology, Geneva, 3–8 Sept 1995), hors série, 1, pp 661–670

van Honk C, Hogeweg P (1981) The ontogeny of the social structure in a captive Bombus terrestris colony. Behav Ecol Sociobiol 9:111–119

van Schaik CP (1989) The ecology of social relationships amongst female primates. In: Standen V, Foley GRA (eds) Comparative socioecology, the behavioural ecology of humans and other mammals. Blackwell, Oxford, pp 195–218

van Schaik CP, van Hooff JARAM (1983) On the ultimate causes of primate social systems. Behaviour 85:91–117

Verbeek MEM, de Goede P, Drent PJ, Wiepkema PR (1999) Individual behavioural characteristics and dominance in aviary groups of great tits. Behaviour 136:23–48

Vicsek T, Czirók A, Ben-Jacob E, Cohen I, Shochet O (1995) Novel type of phase transition in a system of self-driven particles. Phys Rev Lett 75:1226–1229

Wantia J (2007) Self-organised dominance relationships: a model and data of primates. In: Wiskunde en Natuurwetenschappen. Rijksuniversiteit Groningen, Groningen, p 115

Wilkinson GS (1988) Reciprocal altruism in bats and other mammals. Ethol Sociobiol 9:85–100

Yerkes RM (1940) Social behavior of chimpanzees: dominance between mates in relation to sexual status. J Comp Psychol 30:147–186

# Chapter 23
# Agent-Based Simulation as a Useful Tool for the Study of Markets

**Juliette Rouchier**

**Why Read This Chapter?** To understand the various elements that might be needed in a simulation of a market, including: some options for implementing learning by agents in a market simulation, the role and kinds of market indicators, and kinds of buyer-seller interaction (bargaining, loyalty-based, and reputation-based). The chapter will give you an overview of the complexity of markets, including multi-good markets and complicated/decentralised supply chains. It will help you understand financial markets (especially in contrast to markets for tangible goods) and the double-auction mechanism. Finally, it will give some indications of how such models either have informed or might inform the design of markets.

**Abstract** This chapter describes a number of agent-based market models. They can be seen as belonging to different trends in that different types of markets are presented (goods markets, with or without stocks, or financial markets with diverse price mechanisms, or even markets with or without money), but they also represent different aims that can be achieved with the simulation tool. For example, they show that it is possible to develop precise interaction processes to include loyalty among actors; try to mimic as well as possible the behaviour of real humans. After they have been recorded in experiments; or try to integrate psychological data to show a diffusion process. All these market models share a deep interest in what is fundamental in agent-based simulation, such as the role of interaction, inter-individual influence, and learning, which induces a change in the representation that agents have of their environment.

J. Rouchier (✉)
GREQAM, Marseille, France
e-mail: juliette.rouchier@univ-amu.fr

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_23,

## 23.1 Introduction

In recent years, there has been a growing recognition that an analysis of economic systems – which includes markets – as being complex could lead to a better understanding of the participating individuals' actions. In particular, one element that could be incorporated in such analyses is the heterogeneity of agents and of their rationality. For example, the existence of multiple prices on a market for the same product, sold at the same moment and at the same place, cannot be captured in an equilibrium model, whereas it appears in real life and can be reproduced easily in an agent-based model (Axtell 2005).

This issue is at the centre of much debate among economists. In classical economy, agents are considered as rational, having a perfect knowledge of their environment, and hence are homogeneous. This view of agents stayed unchallenged for a while. Friedman (1953) argued, for example, that non-rational agents would be driven out of the market by rational agents, who would trade against them and simply earn higher profits. However, in the 1960's, the view on rationality evolved. Even Becker (1962) suspected that agents could be irrational and yet produce the same results as rational agents (i.e. the negative slope of market demand curves). However, the author who definitely changed the view on economic agents is Simon, who stated that any individual could be seen as a "bounded rational" agent, which means that it has an imperfect knowledge and has limited computing abilities (Simon 1955). In most markets, agents do not have perfect knowledge of the behaviour and preferences of other agents, which makes them unable to compute an optimal choice. If they do have perfect knowledge, they will require unlimited computational capacity in order to calculate their optimal choices.[1]

Indeed, for some contemporary authors, the understanding that one can get of real market dynamics is more accurate if the assumption of a representative agent or of the homogeneity of agents is dropped at the same time as the perfect knowledge assumption (Kirman 2001). The way bounded rationality is approached can be very formal and tentatively predictive (Kahneman and Tversky 1979), but some authors go further by stating that the notion of rationality has to be abandoned. The correct view is that agents possess rules of behaviours that they select thanks to diverse mechanisms, which are most of the time based on past evaluation of actions (Kirman 2001). Some authors also compare the different results one can get from a representative agent approach and a bounded rationality approach for agents, and try to integrate both simplicity and complexity of these two points of view (Hommes 2007).

[1] For example, an issue that anyone representing learning (not only on markets) has to face is the exploration-exploitation dilemma. When an action gives a reward that is considered as "good", the agent performing it has to decide either to continue with this action – and hence possibly miss other, more rewarding actions – or to search for alternative actions, which implies indeterminate results. Leloup (2002), using the multi-armed bandit (Rothschild 1974) to represent this dilemma, showed that a non-optimal learning procedure could lead to a better outcome than an optimal – but non computable – procedure.

To complete the view that agents have bounded rationality in a market and that they evolve through time, it is necessary to consider one more aspect of their belonging to a complex system: their interactions. The seminal works on markets with interacting heterogeneous agents date back to the beginning of the 1990's (e.g. Palmer et al. 1993; Arthur et al. 1997b); many of them were collected in the post-proceedings of a conference at the Santa Fe Institute (Arthur et al. 1997a). Since then, a large part of research in agent-based simulation concerns market situations, as can for example be seen in the cycles of WEHIA/ESHIA conferences and ESSA conferences. The former ones gather many physicists who apply dynamic systems techniques to representing heterogeneous interacting agents to deal with economics issues, and they tend to buy the "agent-based computational economics" (ACE) approach of simulated markets promoted by Leigh Tesfatsion[2] (Phan 2003). In the latter ones, not only economy, but also sociology, renewable resources management, and computer science participate and often try to generate subtle representations of cognition and institutions and a strong view of agents as computerized independent entities to deal with broad social issues and are thus being closer to multi-agent social simulation (MAS). We will refer to both terms (ACE or MAS) separately.

Not only the techniques can be different when studying markets with distributed agents, but also the aims can be. Some try to infer theoretical results about rationality and collective actions (Vriend 2000) or about market processes (Weisbuch et al. 2000). Others want to create algorithms to represent human rationality on markets, and try to assess the value of these algorithms by comparing the simulated behaviour with actions of real humans in order to understand the latter (Hommes and Lux 2008; Duffy 2001; Arthur 1994). Eventually some explorations about the impact of diverse rationalities in a market context enable the identification of possible worlds using a sort of artificial society approach (Rouchier et al. 2001).

Being part of a handbook, this chapter should provide tools to be able to build and use agent-based simulation techniques to create artificial markets and analyze results. However, a "know-how" description of the building of artificial market is so dependent on the type of issue that is addressed, that it was decided here to rather establish a classification of the type of markets and modelling techniques that can be found in agent-based simulation research. We are interested in representations of markets that focus on a micro-representation of decision processes, with limited information or limited computational abilities for agents and which take place in a precise communication framework. This form of representation forces authors to focus on new notions such as interaction or information gathering, and this induces new issues such as the bases for loyalty or the exploration-exploitation dilemma. The use of stylized facts is fundamental in this branch, where modellers try to mimic some elements of the world in their research, focusing either on representations of reasoning that are inferred from observed micro behaviours, or

[2] http://www.econ.iastate.edu/tesfatsi/ace.htm.

trying to mimic global behaviour through learning algorithms, thereby to stay closer to common view orthodox economics.

Contrary to this view, which does not distinguish among individual rationalities and assumes aggregate, centralized knowledge and decision-making, researchers involved in the use of multi-agent simulation usually try to understand the local point of view of agents, and its influence on global indicators. Since the agent is then seen as unable to have complete knowledge, it has to accumulate data about its environment and treat those data according to its aims. The study of markets is interesting when it comes to this type of analysis because markets display a much more simple set of possible actions and motivations than many other social settings. Income and reproduction of activity are direct motivations, which imply choices in the short and the long term, prices and quantities are what have to be chosen (as well as sometimes acquaintances, in the case of bilateral bargaining) and information is limited to offers and demands, as well as responses to these two types of proposals.

This chapter presents the main fields of application for multi-agent simulation dealing with markets, to show how researchers have focused on different aspects of this institution and to conclude on the great interest of agent-based simulation when trying to understand the very dynamics of these social environments.

In the next section, we will describe the main notions that are covered by the term "market" in agent-based literature, and also the main ways to represent rationality and learning that can be found. Subsequently, the main topics of market studies will be described in three parts. In Sect 23.3, agents are on a market and actually meet others individually, having private interactions with each other. Choices that have to be modelled are about matching business partners, as well as buying or selling decisions. In all other sections, agents are facing an aggregate market and they have to make decisions based on global data, sometimes associated to networks. In part 4 agents are either consumers or producers in a large market. In part 5 we will deal with auctions, financial markets and market design.

## 23.2 Market and Agents' Reasoning

Although the study of market(s) has been predominantly carried on by economists, ethnographers and sociologists have also been active in this domain (Geertz et al. 1979; White 1988), and the field is now being developed through field studies and analysis. The main difference for those two approaches is that economists generally build models that are rather abstract and formal, whereas ethnologists and sociologists describe actual markets after observing them and generally produce models that are based on classifications of large amount of data. The notion of market itself has a double meaning, even more important with the increasing use of internet: it is at the same time the institution that enables individuals to coordinate their actions through the fixing of a price or, alternatively, a physical place where buyers meet sellers. There is no easy decision to choose the relevant scale to study when dealing with market, neither is the limit of observation that is needed in the supply chain easy to set. In agent-based simulation, markets are represented as closed societies with a specified set of agents (market being open or closed to new entry) that are possibly interconnected.

Simulations can be based on very specific case studies in order to describe as accurately as possible the behaviour of real actors, but they can also be mainly theoretic in an attempt to generate expected theoretical results. In all cases, a simulated market cannot be implemented as described in neoclassical economic theory since agents need to be independently specified and interact directly with one another during the simulation. For example, to develop a model where individual agents have to make a decision, a demand curve (that gives for any price of a product, the number of agents that are ready to buy or sell at this price) cannot be imposed on the model but has to be derived from the determinants of agent behaviour. One approach is to distribute reservation prices (the maximum price for buying or minimum for selling) among agents which can then be used to aggregate demand and offer curves.

The main elements to define a market model are given in the next section. We will then describe a few approaches to rationality and learning for agents in markets that depend on the type of market and goods that are represented. A similar analysis, more oriented towards consumers' behaviour can be found in (Jager 2007).

### *23.2.1 Main Elements to Build an Artificial Market*

Several dimensions are important on a market, and each description of an element is easy to relate to a dimension in the building of an artificial system with agents (simply put: the market institution and the agents' rules of behaviour). Axtell (2005) proposes a very abstract view of decentralized exchange on an agent-based market, where he gives no explanation of the bargaining process that organizes the exchange, but shows the existence of a computationally calculable equilibrium to increase all agents' utility. Here, the aim is to find out actual processes that can be used by modellers to represent markets.

The first distinction that can be made when developing a model is to know if one is building the representation of a *speculative market* or a *goods market*. What I call a speculative market, typically a financial one, is such that agents who have a commodity can keep it, sell it or buy it. They have to anticipate prices and wait to perform their actions so as to make the highest profit. Seminal works on agent-based simulation were related to speculative markets, which display interesting regularities in their stylized facts. A large body of literature has developed on this topic, which is also due to the fact that data to calibrate models are more easily available than for non-speculative markets. On a goods market, agents have only one role, to sell or buy a certain number of units of one or more products, and they usually have a reservation price to limit the prices they can accept. The products can be perishable (with an intrinsic value that decreases over time) or durable so that stock management is an issue.

Then, both types of market can be organized either through auctions (with diverse protocols: double-auction, ascending, descending, with posted prices or continuous announcements) or via pair-wise interactions which imply face-to-face negotiation (with many different protocols, such as “take-it-or-leave-it”, one shot negotiation, a series of offers and counter-offers, the possibility for buyers to explore several sellers or not). In the design of the protocol, it can also be important to know if money exists in the system or if agents exchange one good for another directly.

Agents’ cognition must entail some choice algorithm. Agent-based simulation is almost always used to design agents with bounded rationality because agents have limited computational abilities or limited access to information. They have tasks to perform, within a limited framework, and have to make decisions based on the context they can perceive. Most of the time, they are given a function, equivalent to a utility function in economics, that associates a value to each action, enabling the agents to classify the profit it gets and hence to compare actions. First, an agent must have constraints in its actions, in order to be able to make arbitration between all possible options:

– To each commodity is associated *a reserve price:* if a buyer (resp. seller) goes on a market, there is a maximum (resp. minimum) price it is willing to pay for the good.
– The importance of obtaining a commodity can be indicated by the price of entry to the market. Agents have a greater incentive to buy and get a 0 profit, rather than not buying. The constraint for selling can be the same.
– In some papers, the price is not represented in the system, and the acquisition of a product is limited by a utility function, where the agent acquires the product only if it makes enough profit.
– In the case of negotiation, time constraints are usually put on buyers who can visit a limited number of sellers, having hence a limit on their search for a commodity.
– There can be a discount factor: at each period, the risk of seeing the market close is constant and hence agents never know if they will be able to trade at the next period.

The type of decisions that agents have to perform on a market:

- *For buyers*: how many units to buy, who to visit, how to decide to stay in a queue depending on its length, which price to propose/accept, which product to accept, and more fundamentally to participate or not.
- *For sellers*: how to deal with the queue of buyers (first-come-first-served or with a preference for some buyers), which offer to make or accept, and in the case of repeated markets how many units to buy for the next time step, in the case of a market where quality is involved: which type of product to propose or produce for the next time step.

Due to the increasing complexity when adding another type of decision to a model, it is rare that all of these decisions will be made in one single model. For example, although interactions could potentially be represented in a continuous way, I know of no model where it is the case: all choices and meetings are made and messages sent at discrete time steps.

### *23.2.2 Agents' Learning*

As said before, in most markets that are studied with agent-based models, the central element is that agents are heterogeneous in knowledge as well as in need. This situation can be decided from initialization or can emerge during the course of the simulation, while agents learn. Another element that is rarely given at initialization but is acquired by agents while learning is information about other agents' characteristics. In most cases, this learning takes place as a result of action, at the same time as the acquisition of an object or the acquisition of money.

The way learning is organized is often linked to a performance of actions, with a selection of actions that "satisfice" or give the best performance. On a market, it is often assumed that agents are interested in getting the highest payoff for their individual actions: the performance is either the profit that agents get from their sells or the utility they get from consuming the product. In most models, learning agents have a set of pre-defined actions they can take and they have to select the one they like the best, following a probabilistic choice.

One of the simplest learning models is reinforcement learning (Erev et al. 1998; Bendor et al. 2001; Macy and Flache 2002; see also Chap. 17 (Macy et al. 2013) in this volume), where agents attribute a probability of choice for each possible action that follows a logit function. The algorithm includes a forgetting parameter and a relative weight attributed to exploitation (going to actions known as having high value) and exploration (the random choice part in the system). The exploration parameter can be fixed during the simulation, where the level of randomness has to be chosen, or can vary during the simulation (decrease) so that there is a lot of exploration at the beginning of the simulation whereas as time passes agents focus on the "best" actions. This issue of which level of exploration and exploitation to put in a learning algorithm is of current concern in the literature about markets (Moulet and Rouchier 2008).

Another model of rationality for agents is based on a representation of strategies of other agents, fictitious play: a distribution of past actions is built by each agent and they can then infer the most probable set of actions of others, and hence choose their optimal behaviour (Boylan and El-Gamal 1993). The EWA model has been proposed by Camerer (Camerer and Ho 1999) to gather both characteristics of these models: the agent not only learns what profit it got for each action, but also computes notional gains for each possible action, and attributes the resulting notional profit to each of those possible actions.

A slightly more complex representation of knowledge commonly found in the literature is the classifier-system where each decision is made by considering a context, a choice and the past profit made in this precise context by making this special choice (Moulet and Rouchier 2007). This type of algorithm is very similar to what (Izquierdo et al. 2004) call case-based learning, but it does not seem to be applied to market situations. In general the number of possible actions is fixed from the beginning, but the classifier system can be associated to a genetic algorithm that generates new rules over the time (Kopel and Dawid 1998). Genetic Algorithm learning is also a quite usual way to represent learning, where the information that is used by agents to estimate the profit of each rule can be based on actual past actions (Vriend 2000) or also on the imaginary profit of all possible actions considering the actions of others (Hommes and Lux 2008). The presence of other agents can also be relevant information when agents use imitation or social learning.

Brenner (2006) undertook an extensive review of common learning processes.[3] In this paper an interesting element arises: the way "satisficing" rationality can be developed by having agents not look for the best action but for one which enables them to get a "good enough" profit; the notion of "good enough", called "aspiration level", can then evolve during the simulation (Cyert and March 1963).

An alternative to learning algorithms that are only based on profit is to consider that agents have a social utility, which need not have the same dimensionality as profit. Indeed, it has been demonstrated that translating the social utility in costs or profits that can be added to monetary profit gives a dynamics of learning and behaviour that is radically different from a situation where agents reason in two dimensions, social and monetary, with a lexicographic ordering (Rouchier et al. 2001). A way to implement social utility without lexicographic ordering is to include in the utility the appreciation of similarity of members of the network, such as in consumer choice models (Delre et al. 2007).

In most models, action and information gathering are made in one action, and circulation of information as such is not really modelled. The reason is certainly because it would take modellers too far from the neoclassical economic approach to market, where the only information is the observation of transactions, sometimes of intermediate prices as in auctions or, sometimes, bargaining. One model of a market by Moss (2002) represents communication among agents before exchange takes place and Rouchier and Hales (2003) (whose model evolved into the one of Rouchier (2013)) also allocate one period out of three every time step for agents to look for information.

[3] The main objection to Brenner's exposition is the lack of homogeneity of notation, which makes the algorithms difficult to compare and maybe to implement.

### *23.2.3 Indicators and Method*

Several types of modelling can be found in papers about markets, just like in any other application domain of simulation. Some prefer to work at a purely abstract level, while others try to fit as well as possible data that they extract from observation and experience. Whatever the approach, indicators that are often observed in markets are prices, efficiency (the total profit that is extracted from agents compared to the maximum profit that could be extracted), and relative power of different agents. The notion of convergence is central to the modelling of markets, since most research refers to economics and has to compare results to economic static equilibrium. In some cases, what is observed is interaction patterns, which can be represented as the random part of the agents' choice when interacting, e.g. the number of different sellers that a buyer meets in a number of steps. In bargaining models and in general exchange models, the (non-)existence of an exchange is also something that is observed. Sometimes, the cognition of agents themselves is observed: their belief about the others' preferences, or even the distribution of propensities to choose sellers.

Among all the indicators that can be observed in the model, these last ones cannot be observed in real life, and hence cannot be compared to human systems. In a lot of models, agents' behaviour is compared to that of humans in order to establish the validity of the cognitive model. The data are very rarely extracted from real life situation (although it sometimes happens), but are mainly constructed via experiments. Experimental economists control all information circulation and record all actions of agents. It is thus possible to compare in a very precise and quantitative way the global behaviour of the group and individual behaviour, on one side with artificial agents and on the other side with human agents. Real life situation can also be seen as the mix of human and of artificial agents, such as in financial online markets.

Other researchers do not want to match data too precisely. As Vriend (2007) says, agent-based models, like any other models, are abstract settings that have to be interpreted as such. The comparison between real and abstract data should go through a first step which is the building of stylized facts that are already a summary of human behaviours, where only the most striking elements are integrated. Vriend is much more interested in the reaction of his abstract model to changes in parameters, and in its self-consistency. It could be said that by construction, a model can only capture a small part of human cognition, which is full of long-term experiments and memories, and should not be compared to quantitative data without caution.

Eventually, some researchers want their model to influence real life and try to use the results they find to give advices on the way to build markets. Different ways to fit models with real life will be found in each example of a model – be it to fit precisely, to fit stylized facts, or to be an abstract study of the effect of some rules in a social setting.

## 23.3 Buyer-Seller Interactions

In the literature about agent-based markets, a great attention has been given to the analysis of local interactions, hardly ever studied in classical economics, apart from rare exceptions (Rubinstein and Wolinsky 1985). An aim is to reproduce as well as possible the features of real markets. It is indeed to be noticed in real observation that buyers display regularity in the choice of sellers with whom to interact and that this regularity emerges in time with experience – this attempt to reproduce patterns of interaction is the only way to understand, rather than just describe, how individuals coordinate (Weisbuch et al. 2000). Authors study local bargaining processes with great care, as well as repetition of interactions over several days, where choices are not only based on prices but also on the fact that buyers need to buy and sellers need to sell. The basic feature of these models is pair-wise interactions on markets with several sellers and buyers where prices are not fixed and result from negotiation only. The number of visits buyers can undertake, the way sellers manage queues, or the number of steps it takes to negotiate a price are different in all these systems that focus only on small parts of the whole set of stylized facts that are observable on such markets. Some usual aspects are described in this section and the subsequent choices for modelling the organization of pair-wise interactions are given.

### *23.3.1 Bargaining Processes*

Brenner (2002) studies agents' learning in a bilateral bargaining market, focusing on the convergence of prices and the dynamics of bargaining. There is one commodity in the market, and buyers and sellers meet at every time step to exchange it. Each buyer can choose one seller for each step, with the selection based on the price being acceptable. The sellers respond to buyers waiting in their queue in order of arrival by proposing a price. Buyers have to decide who to visit; sellers have to decide on the first price to propose and the number of subsequent proposals if the buyer rejects the offer, bargaining being costly for both agents. All decisions are made following *reinforcement learning* based on past experience. Hence, all choices are based on the satisfaction that is associated with each past action and on a rigidity variable. A buyer will continue to choose a seller as long as he is satisfied. His probability to change depends on his expectations with another agent. A seller also calculates the probability to change behaviour depending on his belief about what he would gain by performing another choice.

The rigidity parameter, which is the opposite of noise in the system, has a great impact on results. If rigidity is high, buyers keep visiting the same seller. The cost of bargaining also is important: if it is relatively high, sellers learn to offer the price they know to be acceptable to buyers, and they do not bargain after a few rounds. In this system, since the relations are so individual, the convergence of the price overall is not very fast and can be highly variable for a long time, although converging in the end. The model is extremely sensitive to all parameters that define aspiration levels for agents.

Brenner's paper is of the class that compares simulation to theoretical results. Here, sub-game equilibria are used to compare the possible outcomes and their efficiency to the generated prices and behaviours. There is no reference to any real world data. However, it is interesting that both micro behaviour (the number of bargaining steps) and macro data (prices) are of importance, justifying an agent-based analysis.

Influenced by this paper, but referring to real data, Moulet and Rouchier (2008) reported a bargaining model based on two sets of data: qualitative, from a field study in the wholesale market of Marseilles (Rouchier and Mazaud 2004), and quantitative, giving all proposals, intermediate and final prices for a series of transactions in the same market (Kirman et al. 2005). Like the previous model, the market gathers buyers and sellers who meet at every time step. However, buyers can visit several sellers in one market opening. The choice for a buyer has several dimensions: to decide which seller to visit, to decide to accept an offer or to reject it, to propose a counter-offer or leave, and which value to counter-offer. Sellers must choose the first price to offer, to accept buyer's counter-offers or not, and the value of the second offer they can make. In this model, decisions evolve following classifier system learning, where each agent evaluates a list of possible options following his past experience. The results that are produced are compared with indicators derived from real-world data: values of offers and counter-offers of the agents that vary depending of the kind of product that is purchased and *ex post* bargaining power of sellers (which is the difference between the first offer and the price of transaction compared to the difference between the counter-offer and the price of transaction).

In the simulations, the values that are obtained fit the data quite well in that the observed bargaining sequences and agents' behaviours are reproduced. The two main parameters are the number of sellers that agents can meet (from one to four) and the speed of learning of sellers. The relative importance of learning for the agents can be seen as situating them in a negotiation for in-season goods and a negotiation for out-of-season goods. The model produces results similar to those of out-of-season goods when agents have to learn repeatedly, when there is no known market price but a lot of heterogeneity in the buyers' readiness to pay. In the case of in-season goods, the market price is more settled, and agents do not explore the possible values of goods as much, relying instead on their experience. Between the different in-season goods, the main difference could be the number of visits buyers make, but this number tends to reduce after a learning period, when buyers have selected their preferred seller. This aspect of the model – the growing loyalty of agents – is not the centre of the research and was represented mainly with the aim of matching the actual behaviours of the market actors. Other papers, described in the following section, are more focused on this issue.

Another direction for the study of bargaining processes is related to the creation of robots or artificially adaptive agents (AAA) to participate in electronic commerce (Oliver 1997). Such models focus on complicated negotiations in that they integrate several dimensions of trade in the deal: price, quantity and delivery time. The main argument for the value of the algorithm that is proposed in the paper is that the agents learn to negotiate at least "as well as humans", which means that as many negotiations lead to an agreement as in human bargaining situations so that profit is extracted from both sides of the bargaining. The bargaining consists of several steps, where a customer reacts to the first offer by comparing its profit to a threshold, and the offer is accepted if it is higher than the threshold and rejected with a counter-offer otherwise. Clearly, such models capture satisficing and bounded rationality rather than profit maximization. The bargaining can then carry on with several successive offers being made by customer and seller. Strategies for accepting and counter-offering evolve through a Genetic Algorithm. Five different games are used to test the learning, in a population of 20 agents with 3 rounds of bargaining at most and each agent is given 20 chromosomes for decision-making. It is then proven that AAA perform better than random, that agents are able to learn general strategies that can be used against different bargaining partners, and eventually that AAA perform as well as humans (depending on the game, sometimes better and sometimes worse, maybe depending on affective values for humans) in terms of number of agreements that are reached. This is an interesting result to consider when one wants to introduce artificial agents into electronic markets, since one wants to be able to reach as many agreements as possible.

### *23.3.2 Loyalty*

Loyalty is present in quite a few models of markets where agents interact repeatedly. It is popular to deal with this topic with agents, mainly because it is related to two main advances of agent-based modelling: heterogeneity and interactions. There exist two representations of this loyalty in the literature: either fixed loyalties, assumed in order to understand its impact (Rouchier 2013), or emerging loyalties, as the result of endogenous interactions. Vriend refers to "endogenous interactions" when he uses individual learning to generate an evolution of interactions among agents (Vriend 2007). The idea is that agents learn to select which actions to perform as well as which agent to interact with; it is clear that this can lead to the apparition of loyalty, and that it can take different regular patterns.

One main field where this loyalty issue has been important is the study of perishable goods markets (fruits and vegetables and fish). The participants of these markets are very dependent on the regularity – which implies predictability – of their interactions. The main reason is that buyers need to purchase goods almost every day: they have very little ability to stock and they must provide their customers with all possible goods (a retailer can become unattractive to his customers just because of the lack of one commodity). In case of shortage, they need to have good relations with a seller to make sure the commodity will be available to them. Conversely, Rouchier (2013) shows in a model that the presence of loyal agents in a perishable goods market is necessary for the sellers to predict the right number of goods to provide every day. In this artificial market, two types of buyers interact with sellers: those that look for the cheapest prices ("opportunistic") and those that are faithful and try to get the product rather than to get it cheap ("loyal"). To be able to be opportunistic, agents first gather information about prices, and then decide on the seller they want to meet to make the best transaction. The more opportunistic agents are present in the market, the more garbage is produced and shortage occurs. Although there is some randomization of needs for the buyers, the presence of loyal agents makes the sellers estimate their stocks in the best way. This result holds for different learning algorithms (step by step learning, simple reinforcement learning and classifier systems). Considering that the field study on the fruits and vegetables market of Marseilles, France, showed that most of the agents are loyal (according to the definition of the model: first loyal and then try to find all the goods in a minimum of visits to sellers), this result can give a functional explanation of their action choices.

In a slightly different context, Rouchier et al. (2001) have represented the shape of emerging patterns of relations that could be created by two types of rationality with agents. The situation is a market-like situation, where offers are dependent on the situation in the preceding step, since the commodity is a renewable resource. Agents are herdsmen and farmers, with the latter selling access rights to their land. If none or if too many herdsmen are using the same land, it will get depleted, and hence the offer will be reduced. Two types of micro behaviour are defined: either the herdsmen choose the cheapest farmers, or they choose the ones that offered them access most often. In the first case, the simulations resulted in depletion of the resource, with congestion of demand for the cheapest farmers. The links that were created were highly stable (once an agent found the cheapest it would not change), but on the other hand agents could not readapt when there was a shock in the resource quantity (a drought) because everyone would converge to the same farms. With the second rationality, agents had a representation of a "good" farmer, which was only based on individual experience, and hence they would be heterogeneous. They would also have several "good" farmers to visit in case one was not available. This made them much more flexible in their choice, avoiding depletion of the resource, so everyone was better off. The macro situation, although the process is different, also shows that a loyal micro-behaviour is a help to repartition of goods were there can be shortages. In this setting the loyal micro-behaviour also enables a more equal repartition of gain among farmers as well as herdsmen.

Kirman and Vriend (2001) explored the emergence of loyalty in an artificial market based on a field study in the fish market of Marseille. The aim is to see loyalty emerge, and in addition to see which emergent behaviour sellers display. They use classifier systems to represent learning, where their agents can have a lot of different actions, some of which are, a priori, not good for their profit. Through the exploration of good and bad possible actions, they select those that bring the highest profit in the past. Some buyers learn to be loyal, and those that learn this get higher profit than others in the long run (it is actually a co-evolution where sellers learn to offer lower prices to those that are loyal). The buyers are then differentiated: their reservation price is heterogeneous (for example to represent that they do not sell their fish to the same population, some are in rich neighbourhood, some in poor ones). Sellers on the market learn to discriminate, and they offer higher prices to those that have higher reservation prices. Eventually, some of the sellers get themselves specialized since only low prices buyers can visit them. Using a very basic learning where agents are not rational but learn by doing, the results are very satisfying because they reproduce stylized facts of the fish market.

A third model represents the same market but refers more to quantitative data of this market (Weisbuch et al. 2000). The data represents sales that took place over more than 3 years and concern 237,162 individuals. From these it is possible to observe that most buyers who are faithful to a seller buy a lot of quantities every month. The model was built in two parts: one which is simple enough to generate analytical results and a second that displays more realistic hypotheses. In the analytical model, agents use the logit function to select their action (basic reinforcement learning), which means that their choice depends on a $\beta$ value, between 0 and $\infty$, which decrease gives a higher propensity to randomly test all sellers and which increase induces a higher propensity to look for the best past interaction. Agents can either imitate others or only base their choice on their own learning. The results can be found using the mean field approach, coming from physics. It is shown that there are radically different behaviours – either totally loyal or totally "shop around agents" depending non-linearly on $\beta$.

The model becomes more complex with sellers being able to sell at two different prices, high and low. What can happen in this system is that a buyer becomes loyal to a seller when the price is low, and remains loyal even after the price has switched to high. The only important thing is that, as seen before, the product is actually provided. One indicator that is used to synthesize diverse information of the model is "order", which is defined as the number of loyal agents. The more regular the agents, the more ordered the society. Although the results, once interpreted, are very coherent with real data in terms of possible states of the market, it is a bit difficult to understand precisely some concepts of the paper because it refers mainly to physics indicators that are translated into social indicators, but this translation is not always straightforward.

### *23.3.3 Reputation of Sellers*

Pinyol et al. (2007) have developed a market model as a benchmark for a reputation-based learning algorithm for agents in a social system. The model integrates quality and judgment of a relationship. Reputation is used in the group to enable agents to gather enough information in a context when it is scarce. The market that is used is a rather simple institution, where buyers have to select one seller at each time step to buy one unit of a commodity. The quality of the commodity is different for each seller. For a buyer, the acquisition of a commodity of lower quality will give less utility than the acquisition of a commodity of high quality. Sellers have a limited quantity of units, which they can sell at any period (the commodity is non-perishable) and they disappear from the system when they have sold everything. The most important information for buyers is the quality of the commodity that each seller offers. However, when the number of sellers is large, this information cannot be acquired efficiently if the buyer has to meet a seller to learn of the quality of his commodity. This is why information circulates among buyers, who communicate once every time step. A buyer who meets a seller forms an image of this seller; a buyer who gets information about a seller has access to a reputation of this seller. When giving information to another buyer, an agent can decide to give the direct knowledge it has (the image it formed of a seller) or the reputation it has already received (which is more neutral since it is not its own evaluation). Reputation can also circulate about the buyers, and in that case concerns the validity of the knowledge they give about sellers. When a buyer is not satisfied with the information given by another buyer, it can also retaliate and cheat when this very agent asks him a question.

Pinyol et al. (2007) describe in detail the choices that agents make when asking a seller for a commodity, asking another buyer for information about a seller or a buyer, answering a question and the lying process.

The simulated market contains a large number of sellers (100 for 25 buyers). Simulation runs are defined by (a) the type of information that is used by agents (only image or image and reputation) and (b) the number of bad-quality sellers in the system (99, 95, 90, and 50). The addition of reputation to the system makes the difference between a normal learning mechanism where buyers select their favourite seller and a learning mechanism where agents aggregate information of different quality to (maybe) increase their performance. The results show that globally the agents indeed learn more efficiently when using reputation, in that the average quality that is bought is higher. The quantity of information that circulates is much higher and this enables buyers to increase their utility. This social control mechanism is especially important when quality is really scarce (1 % of good sellers). This result is all the more interesting since this is a very rare case of a simulated market where communication among buyers is represented, although this behaviour is commonly observed in real life situations. For a more in-depth discussion of reputation see Chap. 15 in this handbook (Giardini et al. 2013).

## 23.4 Consumers, Producers and Chains

Another way to look at the notion of goods markets is to consider large markets, where individual interactions are not important for the agents who do not record the characteristics of the ones they meet, but only the fact that they can or cannot perform an exchange. A large market can indeed include numerous goods, that are distributed among different other agents and not necessarily easy to access. Another interest in large market is to study endogenous preferences for goods, and imagine their evolution depending on the type of good and some cognitive characteristics of agents. Eventually some authors are interested in the coordination process within the supply chain itself, where the issue is about the amount of information that each agent has to use to anticipate the needs of distant, end consumers.

### *23.4.1 Multi-goods Economy*

As examples of a market with several goods we will discuss two very abstract models where agents have to produce one commodity and consume others, which they can acquire only through exchanges with other agents. The first model was built to produce speculative behaviours in agents, which means acquiring a product that has no value for consumption but only a value for exchange (Duffy 2001); the second model's aim is to witness the emergence of commonly used equivalence value for the goods, which is interpreted as relative prices (Gintis 2006). Both models are interesting for their pure description of an abstract economy with minimalist but sufficient assumptions to induce economic exchanges. In the works cited here, the methodology used in realising the models is slightly different: one is purely abstract whereas the other tries to refer to experimental results and mimic human players behaviours.

In his paper, John Duffy (2001) designs a model that was originally proposed by Kiyotaki and Wright (1989) to induce some agents to store a good that is not their designated consumption good and is more costly to store than their own produced commodity, because they think it easier to exchange with others. There are three different goods in the economy, agents need to consume one unit of good to increase their utility and produce one unit of good each time they have consumed one. There are also three types of agents: agent type 1 needs good 1 and produces good 2; agent type 2 consumes good 2 and produces good 3; agent type 3 consumes good 3 and produces good 1 (in short, agent type $i$ consumes good $i$ and produces good $i + 1$ modulo 3). Hence agents have to exchange when they want to consume and not all agents can be satisfied by just one exchange. Indeed, if two agents exchange their own production goods, one can be satisfied but the other would get a useless good, which is neither its own production good nor its consumption good. In this economy, only bilateral trading exists and it takes place after a random pairing of agents. This involves that some agents must keep a good for at least one time step after production before getting their consumption good.

In this economy, speculation is defined as the storage of the good $i + 2$, since the agent does exchange to get this good which it has not produced, only because of the chances to use it as an exchange good at the next time step. The economy is made non-symmetric by having different costs for the storage of goods, here $0 < c_1 < c_2 < c_3$. The original Kiyotaki and Wright model is all about calculating, given the storage costs, the discount factor (the probability that the economy stops at the end of a time step) and the utility of consumption, whether agents decide to get the most expensive good to store or keep their production good. In an economy with perfect information, the expected profit for each type of agent depends on the proportion of agents of type $i$ holding good $i + 1$ (their production good), which is 1 – proportion of agents of type $i$ holding good $i + 2$ (their “speculative” good).

With this model John Duffy tries to investigate how agents could learn which choice to make when they are able to acquire a good they do not consume. Especially agents of type 1 are those that should hesitate, since good 3 is the most expensive. A lot of models have been built on this topic (Basci 1999) already, but what Duffy wants to produce is a setting that is close to laboratory experiments he has been leading (Duffy and Ochs 1999) in order to be able to judge if his agents are behaving in a way which is coherent with that of human actors. So, from a theoretical setting he builds experiments and simulations, and compares all the results this technique produces. In this paper he therefore proposes an algorithm that is close to his intuition of what individuals should do (he has also asked questions to people involved in his experiments), and he then tries to mimic the results of his experiments, at a global level and a local level. He also proposes some original settings where he mixes human agents with artificial agents to test at the same time his algorithm and how much he can make humans change behaviour depending on the stimuli they get. He is satisfied with his results, where his model of learning enables to reproduce human behaviour correctly. Unfortunately, the reproduction of his model is not so straightforward (Rouchier 2003), but all in all, this description of a very basic economy with few goods and where agents learn in an intuitively plausible way is a very interesting example of market for agent-based modellers.

The paper by Gintis (2006) presents similarities, although the aim and the central question are different. The economy that is presented can be seen in a very general way but is only implemented in one setting, which is described here. In the economy there are three goods and 300 agents. Each agent can produce one good and needs to consume both goods it cannot produce; hence it is forced to exchange with other agents. At the beginning of each period, an agent only holds the good it produces (in a quantity that it can choose and which is costless) and can meet two agents, each producing one of the good he needs to acquire. Each agent has a representation of “prices”, which is here defined as the equivalence quantity between two goods. There is no common knowledge of prices and each agent has its own representation. When an agent meets another agent who can provide him with the needed good, he offers to trade, by sending as a message its representation of relative “prices”. The exchange takes place at this exchange rate if it is acceptable to both agents and the exchanged quantities are the highest quantity that both can exchange. Agents cannot choose who they meet, they just pick randomly from the other producers’ groups. After exchanging, they can consume, which gives them utility, and defines a performance for each individual. Learning in this system is an event that takes place every 20 periods, where the 5 % of least performing agents (who get the lowest utility) copy the price representation of the highest performing agents.

What is observed in the system is the efficiency of the represented market, meaning the sum of all profits, compared to a setting where prices would be public. When prices are public, all exchanges can take place since all agents agree right away on the equivalence that is proposed and there is no refusal in exchange. In the long-term, the system converges to the highest possible efficiency, so although the agents have private prices, these prices get to be close enough to have regular exchanges. This result in itself is not very surprising in terms of simulation (considering the process at stake), but is interesting in economics since it gives, at least, a process to attain to a common knowledge which is often presupposed to exist.

### *23.4.2 Adoption by Consumers*

The study of the behaviour of large numbers of consumers facing the introduction of new products on a market is a topic that is very interesting to approach with agent-based simulation, since it allows, once more, looking for the influence of heterogeneity of individuals and of networks in the evolution of global results. Wander Jager is a prominent figure in this area of research, positioned between psychology and marketing. In a paper with Marco Janssen, he presents the basic model (Janssen and Jager 2001). The idea behind the study of the acquisition of a new product in a group is that agents have a preference that is based on two main parameters: the individual preference for the consumption of the product and the interest that the agent has to consume the same product as his acquaintances. Hence, a utility function depends on these two parameters, and this will influence an agent's decision to buy a new product. Agents are heterogeneous in such a system, and the representation of "early adopters" (in the real world: people who buy a product when it is just released) is modelled by a low need to conform to others' behaviour. On the opposite end of the spectrum, some agents buy a good only because a lot of their acquaintances have already acquired it.

In (Janssen and Jager 2001) the influence of network size and topology is tested, as well as the influence of the utility brought by the product consumption and the way agents choose their action. One aspect that is studied is the type of cognitive process that can be used by the agent (repetition of the same action; deliberation to find a new action; imitation – where other agents' consumption is imitated; social comparison – where other agents are imitated based on their utility). This indicator is quite rare and shows the psychological grounding of the paper. It is interesting to observe that the cognitive process changes with the utility gained by the consumption of the considered product. Agents with a lot of links are very important for the spreading of product adoption: in a small world network, many more products get adopted than in a scale free network. A discussion is open here about the type of products, which certainly influence the way people copy others – it will be different for milk and for computers or clothes.

This last question is actually developed in a different paper: in (Delre et al. 2007), the question that is at stake is to determine how to advertise efficiently depending on the type of good. Is it better to advertise a lot at the beginning of a campaign, or after a moment; is it better to advertise to a large number of people or to disseminate information among only a few agents?

Two products are differentiated: a brown product, which is a high-tech and quite fancy product that can be compared with other agents of the network (e.g. CD, DVD player), and a white product which is related to basic needs and is not really compared (fridge or washing machine). Agents are gathered in networks of different topologies. In this model, the heterogeneity in the utility formula is similar to the one in the preceding paper: each one is defined by a propensity to be influenced by others and to be an adopter of a new technology.

The first finding of this model is that the timing of promotion is important to the success of the campaign. For a first launch, the best strategy is to "throw gravel", i.e. to do a little advertising to many distant small and cohesive groups of consumers, who will then convince their network neighbours. Another element is that not too many people must be reached at first, since if they see that others have not adopted the good yet, they might not want it and become impossible to convince afterwards. This is mainly true for the white good, where it is better to advertise broadly when at least 10 % of agents have already adopted the good, whereas with brown goods adoption is much faster and a campaign helps the takeoff.

The issue of the adoption of a practice within a social context has also been studied to understand the adoption of electronic commerce for consumers (Darmon and Torre 2004). The issue at stake is that it should be logical that everyone turns to electronic commerce, which radically reduces transaction and search costs, but we observe that a very small proportion of items are as yet traded via the internet. This is mainly because consumers have not developed special abilities that are associated with this form of interaction, and do not know how to reduce the risk of performing a bad transaction. To study the dynamics of adoption of electronic commerce and learning of agents in a risky setting, a simulation model has been built. The market is composed of agents who can produce a good and exchange it for the good they want to consume (hence all agents are at the same time producers and end consumers). Agents and goods are located on a circle; the location of an agent defines the "quality" of the good it produces. For consumption, each agent is defined by an interval of quality: when consuming a good whose quality is within this interval (not including its own production good), it will get a strictly positive profit. The cost of production is heterogeneous and can be constant during the simulation or evolving.

When trading on the traditional market, an agent can identify the quality of a product offered by another agent, but it has to talk to many others before finding out who it can exchange with (depending on the number of agents and of its interval of choice). The authors also added a notion of friction, which is a probability of failing to trade when two agents meet. In the electronic market, an agent sees all other agents at once (no search cost), but cannot identify precisely the quality that is offered and evaluates it with an error interval. Hence it potentially accepts goods with zero utility. Agents are heterogeneous in their ability to perceive the quality and their learning algorithm. If an agent learns via individual learning, then it eliminates an agent from its list of potential sellers whenever the previous trade brought no utility. If agents learn through collective learning, then a part of the whole society belongs to a community that shares information about the quality (location on the circle) of the seller they met at this time step; the agents not belonging to the community learn individually. In some simulations, for both types of learning, agents forget a randomly chosen part of what they learnt at each time step.

In the case of individual learning, the dynamics produced depend on the production cost, which can either change at all time steps (and hence all agents have the same average production cost over the simulation) or which can be constant and delineate populations with high or low production costs. When the production cost changes at each time step, the main result is that eventually all agents switch to the electronic market, with the move happening in stages. Those that have a good appreciation of quality go to the electronic market very fast because it is more profitable for them. Their departure from the traditional market reduces the probability of exchange for the remaining agents, who eventually move to the electronic market as well. When production costs are heterogeneous, some agents cannot switch from traditional to electronic because of their inadequate production cost. Hence they never learn how to identify quality and stay in the traditional market. When agents forget part of what they have learnt, then the size of the electronic market does not get as large as with perfect learning, and a number of agents do not switch.

When agents participate in a community and exchange their information, the highest number of agents will switch to the electronic market and overall the lowest number of agents is excluded from exchange. Three groups are created: agents belonging to the community, who get the highest pay-off; agents with low production cost or high expertise who can go on the electronic market and make a high profit; and the remaining agents, which sometimes cannot exchange This result is rather coherent with what could be expected, but it is interesting to have it created with this location-based representation of quality that each individual wants to attain. It is especially clear that there is little risk that traditional markets should disappear if the main assumption of the model – that agents need an expertise that takes long to acquire before switching to the electronic market – is true.

### *23.4.3 Decentralized Supply Chain*

Supply chains are an important aspect of economics, and they are often difficult to consider, mainly because their dynamics spread in two directions: (1) along the length of the chain, suppliers have to adapt to demand and buyers have to adapt to the speed of production so as to be able to provide the end market with the right quantity of goods; (2) another dimension is the fact that suppliers as well as buyers in the chain are substitutable and that each actor is itself in a market position and can choose between several offers or demands. In existing agent-based models, only the first issue is treated. The structure of these models is a series of agents (firms) that are each linked to two agents: a supplier and a client (except for first supplier and end consumer, of course, who are linked to only one firm). Each agent has to decide on its production level at each time step, knowing that it needs to use goods from the preceding firm in the production process. It must then anticipate the demand to order enough, before being able to transform the needed quantity. Of course, each firm takes some time (number of steps) to transform the product and be able to sell and there is a cost in storing its own production when it is not completely sold.

One very important issue of these chains is at the centre of most research: how to avoid the so-called *bullwhip effect*. This effect is a mechanical dynamics that comes from the slow spreading of information and delay in answer because of the length of the production process in each firm. When there is variability in demand coming from end consumers, this variability increases a great deal when it goes up the chain, right up to the first producer who exhibits the highest variability. It can be very annoying for organizations to be trapped in such negative dynamics. Several authors propose algorithms for artificial agents that have to deal with the issue of anticipating demand at each stage of the chain. For example Lin and Lin (2006) describe a system where artificial agents can interact with real agents (and hence be integrated in a real-life company to help decision makers) in order to choose the right level of production and order to reduce costs. Several learning algorithms are tested and their efficiency attested, even in situations where the environment is dynamically evolving. The same issue is dealt with by others, for example Kawagoe and Wada (2005), who propose another algorithm. They also propose a method to statistically evaluate the bullwhip effect. Their method is different from the usual frequency-based statistical measurement (like stochastic dominance) but is based on descriptive statistics.

## 23.5 Financial Markets and Auctions

Financial markets have been one of the first examples that were developed to prove the relevance of agent-based modelling. Arthur et al. (1997b) indeed reproduced some important stylized facts of asset exchanges on a market and this paper is always cited as the first important use of this modelling technique for studying markets. Contrary to models that were presented before, there is no direct interaction among agents in these models, only observation of price patterns. One rare example presented here is an attempt to link a financial market to a consumer market such as the ones seen in previous sections. Another type of market that does not integrate any interaction in the economy is the representation of auctions.

### 23.5.1 Financial Markets

The literature on financial markets is very important in agent-based simulation, and dates back to the 1990's (Arthur 1991, 1994; Arifovic 1996; Arthur et al. 1997a). This holds also true for the related branch of research, which is called *econophysics*: the use of physics techniques to deal with economic issues in systems that are composed of a huge number of simple interacting actors (Levy et al. 2000). A comprehensive review of this topic is (Lux 2009). It describes at the same time the main stylized facts that can be found in financial markets (and hence are meant to be reproduced by simulation) and some models that are candidate for explaining these facts. Another review (Samanidou et al. 2007) describes several agent-based simulations models dealing with financial markets; unfortunately, these models do not achieve to reproduce the very general statistical regularities of these markets. As usual, in the following I will describe only a selection of models and ways to represent agents' learning in the context of financial markets. The basic structure of the market, which defines the type of choice the agent has to make, can vary as well as the aim and methodology of the researcher building these models and this is why discussing a few representative examples in detail seems a better idea than presenting very generic results.

One reason for using agent-based models is to be able to represent populations of heterogeneous agents. What is very often found is the representation of two types of agents with different reactions to information: chartists and fundamentalists. Fundamentalists base their investment decisions upon market fundamentals such as dividends, earnings, interest rates or growth indicators. In contrast, technical traders pay no attention to economic fundamentals but look for regular patterns in past prices and base their investment decision upon simple trend following trading rules. Computer simulations such as those of the Santa Fe artificial stock market (LeBaron et al. 1999; but see also e.g. Kirman 1991; Lux and Marchesi 1999, 2000) have shown that rational, fundamental traders do not necessarily drive out technical analysts, who may earn higher profits in certain periods. An evolutionary competition between these different trader types, where traders tend to follow strategies that have performed well in the recent past, may lead to irregular switching between the different strategies and result in complicated, irregular asset price fluctuations. Brock and Hommes (1998) have shown in simple, tractable evolutionary systems that rational agents and/or fundamental traders do not necessarily drive out all other trader types, but that the market may be characterized by perpetual evolutionary switching between competing trading strategies. Non-rational traders may survive evolutionary competition in the market; see for example (Hommes 2001) for a survey.

In (Hommes and Lux 2008), the chosen market model is the so-called *cobweb model*, which is a prediction model on a market, not an actual model of selling and buying for agents. The model offers, however, a rational expectation value, which serves as a benchmark. The methodology is to try to fit agents' behaviour in an artificial world to real behaviours of individuals in experiments. The game is such that participants of the experiments have no clear idea of the structure of the market but still have to predict the price of the next period. They neither know how many other agents are present, nor the equation that calculates the future price based on the realised price and the expectations of all participants. The simulations are made based on rather simple models of agents including a genetic algorithm, simple learning that copies past prices, and reinforcement learning. What interests the authors most is the GA learning, which is the only one to fit stylized facts in different treatments. What the GA learns about is a 40 bit string of 0 and 1 representing two values, $\alpha$ (the first 20 bits) and $\beta$ (the remaining 20 bits), that predict the price at $t + 1$ depending on the price at $t$ with $p(t + 1) = \alpha + \beta (p(t) - \alpha)$.

There are three runs both for experiments and for simulations, with one parameter defining the stability of the price (high, medium or low). The genetic algorithm being varied for different mutation rates is proven to be largely better than other learning procedures that have been implemented. “Better” means here that it fits the stylized facts that have been produced by humans in experiments: (1) the mean price is close to rational expectation, and the more stable the market, the closer the mean price is to this rational expectation value; (2) there is no significant linear autocorrelation in realized market prices. The reason for the good fit of the GA given by the authors is really interesting because it is not obvious to imagine how GAs, which are random learning processes with selection, should be similar to human learning. The authors assume that the good fit is based on two facts: the fact that successes are selected positively, and that there is heterogeneity in the strategies among the set that agents can use. Once the assessment of the model is done, it is used to question the stability of the results of the learning process. One question that arises is to wonder whether humans would adapt the same way when interacting in a very large group as they do in a small group of six. This opens many questions about the scalability of results concerning market dynamics.

In our second example the interaction of agents is direct and not necessarily via the price system, as is usual in financial markets. Hoffmann et al. (2007) indeed consider that many agent-based simulations still take little interest in representing actual behaviours of decision makers in financial markets. They argue that (Takahashi and Terano 2003) is the first paper to integrate theories that come from behavioural finance and represent multiple types of agents, such as overconfident traders. In their own paper, Hoffmann et al. (2007) present their platform SimStockExchange™, with agents performing trades and making decision according to news they perceive and prices they anticipate. They argue that their model is based on several theories that are empirically sound and that they validated their model results against data over several years from the Dutch market. As usual, the platform allows many variations (increase the number of different shares, of agents, change the characteristics of agents) but is tested only with some values of parameters.

Agents receive news that they forget after one time step, and then can perform two types of action: either sell their stock (if they expect to lose at the next time step) or buy more shares (in the opposite case). To make sure that they are not making mistakes, agents can use risk-reducing strategies, which can be clarifying strategies (such as collecting more data) or simplifying strategies (i.e. imitating other agents), as well as purely individual (the first one) or social (the latter). In the presented simulation, strategies are always social, and hence agents' confidence, $C$, determines their use of risk-reducing strategies; the confidence values were deduced from empirical studies. Each agent is also defined by a tendency $R$ to perform a simplifying strategy or a clarifying one. $R$ and $C$ are evaluated on the basis of surveys made with investors. Agents are imbedded in networks of two different topologies (torus, scale-free); agents may acquire information from their links or choose to imitate them. The market itself is designed as an order book, where proposals for sells and buys are written down with quantity and price, and are erased as soon as an agent answers positively to a particular proposal. The market price is the average of all proposed bids and asks of the order book – hence it is not a realized price (average transactions' price) but an aggregation of desired prices for agents.

In the results, some statistical properties of the stock exchange have been reproduced. For example, with weekly data of Dutch stock exchange, linear auto-correlation can be observed, and this is better reproduced when the torus-shaped network is used rather than the scale-free one is used. With regard to volatility clustering, the torus network differs from both the scale-free network and the real data. This can be due to the high speed of information circulation reducing the shocks that it can cause. The main aspect of the SimStockExchange that needs improvement is the news arrival, which is a normal distribution around the present price. This might have a large impact since the use of different networks integrates the importance of information spreading.

### *23.5.2 Relation Between Two Markets*

Sallans et al. (2003) report a model integrating two types of markets: a financial market and a goods market in the same system. Consumers, financial traders and production firms are interacting and the aim is to understand how these two markets influence each other. The good is perishable, and hence needs to be purchased regularly. Consumers make purchase decisions; firms get income from sales and update products and pricing policies based on performances; traders have shares, which they can hold, sell or buy. Firms decide upon the features of their products, which are represented as two binary strings of 10 bits. In choosing actions, the firm agent uses an action-value function that integrates expectations about future rewards (firms are not myopic agents) by taking into account the evolution of the price of its share in the financial market and the profit made by selling products on the goods market. Consumers have preferences for particular features of a product and its price, and compare any available product to these preferences: they can choose not to buy if the product is too different from their preferences. In the financial market, agents build expectations and built representations of future values by projecting actual and past values into the future. They are divided into two groups: fundamentalists (use past dividend for projection) and chartists (use the history of stock prices); they are also heterogeneous regarding their time horizon. The market clearing mechanism is a sealed bid-auction and the price is chosen to maximize the number of exchanges (and randomly among different prices if they produce the same trade volume).

Agents from the financial market and firm agents have different views on the future of a firm, and evaluate future gains in a different way, which might impact the firm's performance negatively. The simulations' aim is to prove that the model can be used, in certain parameter settings, to reproduce stylized facts of markets.

Although the central issue is very interesting, the paper itself is not as helpful as it could be to understanding the dynamics of two markets. In particular, the stylized facts are not very explicit in the paper (appear only once at the end, when obtained results are given). They are classical in financial market analysis, but not clearly shown here: low autocorrelations in stock returns, high kurtosis in marginal return, volatility clustering. Hypotheses on behaviour are never explained, hence there is no understanding of why the stylized facts can be achieved, apart from doing some random exploration of the parameter space. Thus, while the main issue of the paper is fascinating, the results are a bit frustrating and the reciprocal influences of these two markets, so important in our real world, stay hidden.

### 23.5.3 Double Auctions

In economics double auction is a very fascinating topic, since it is an extremely stable market protocol in which predictions can be translated from theory to real life, which is not really the case for most economic systems. When putting real people in a double-auction setting, one can observe that the convergence to equilibrium price occurs. This does not mean that this protocol is efficient, since a lot of exchanges take place out of equilibrium price, but at least there is a tendency for the group to converge to a price where the highest number of exchange can be performed, and hence the highest global profit can be extracted. Many authors have therefore wanted to reproduce a double-auction market in an artificial society in order to understand the source of this high efficiency.

The CDA (continuous double auction) is a two-sided progressive auction. At any moment in time, buyers can submit *bids* (offers to buy) and sellers can submit *asks* (offers to sell). Both, buyers and sellers may also accept an offer made by others. If a bid or ask is accepted, a transaction occurs at the offer price. An improvement rule is imposed on new offers entering the market, requiring submitted bids (asks) at a higher (lower) price than the standing bid (ask). Each time an offer is satisfying for one of the participants, she announces the acceptance of the trade at the given price, and the transaction is completed. Once a transaction is completed the market is cleared (meaning there is no standing bid or ask any more) and the agents who have traded leave the market. At that moment, similar to the opening of the market, the first offer can take any value, and this proposed price imposes a constraint on any following offer. When the market closes, after a time decided beforehand, agents who have not yet traded are not allowed to continue. In this market protocol all market events are observed by all (bid, ask, acceptance and remaining time before market closing) and hence are said to be common knowledge.

Using this double-action setting, a seminal paper by Gode and Sunder (1993, 2004) shows the strength of institutional constraints on the actions of agents. In their model agents are perfectly stupid from an economics point of view, since they have no understanding of their own interest and only follow a simple rule without any strategic planning. These so-called "zero-intelligence" agents are not allowed to sell (buy) lower (higher) than their reservation price, and they have to bid within the limits that have been put by others. With this rule, convergence of prices is obtained very fast. The approach in this paper is quite original in the behavioural economics literature in the sense that it is close to an "artificial life approach". The authors do not pretend to study human rationality, but instead focus on the abstract reproduction of phenomena. It is interesting to note that is not so easy to design a double-auction market, especially in its continuity. Indeed, in a real situation, if two individuals have close reservation prices, they will often be able to buy or sell at the same moment. Who will be first is not obvious, since people have different aspiration for profit. Gode and Sunders randomly choose an agent between all buyers who can buy or make a bid, and then randomly pick a seller among those who can sell or make an offer. After trying several methods they decided on random selection, explaining that this is a good approximation to continuous double auctions.

Their work is widely criticized because (a) they are not interested in rationality but in a specific market protocol and (b) it cannot be generalized to other protocols (Brenner 2002). However, their result is important and led a lot of researchers to question it. For example (Brewer et al. 2002) shows that humans are able to have markets converge when the context changes a lot, which Gode and Sunders' agents cannot do. They organize a double-auction market, in which agents participate in the public market but also receive offers from the experimenter privately. Only one offer is made at a time, and it is the same for all agents that are proposed the offer, since the equilibrium has to stay the same. The global equilibrium (which value is described in the paper) is thus constant, but individuals can have incentives not to participate in the public market if the offer is interesting. This does change the performance of zero-intelligence a lot, since the prices do not converge anymore in simulations led with this new protocol. On the opposite, humans performing experiments attain convergence, which could mean that only very specific institutions constrain participants enough so that they have no choice but to converge, even while not understanding more than the rules (zero-intelligence).

Inspired by Gode and Sunder, but also by the theoretical model of Easley and Ledyard (1993), Rouchier and Robin (2006) tried to establish the main elements a rational agent would need to be able to choose the right action in the context of a double auction. To differentiate among different possible learning procedures, a comparison with some experimental results was made. The learning procedure chosen is a simple algorithm that consists of making the agent revise its reservation price towards past average perceptible prices, depending on two variable elements.

First, the duration after which an agent would change its reservation price (i.e. a buyer (seller) accepting higher (lower) prices), called the "stress time", could change – increasing after a successful transaction and decreasing after a day with no transaction. Second, the agent could either only perceive its own transactions or those of any successful transactions in the market. The paper demonstrated that agents learn faster to converge to the equilibrium price (making the highest global pay-off) if they did not revise their stress-time and had a global perception of prices. This quick learning would at the same time correspond best to the speed of convergence that could be found in experiments. What is a bit surprising in this result is that more "clever" agents (reacting to risk and failure from one day to another) would neither copy human behaviour well nor get to the equilibrium price very fast.

## 23.6 Market Design/Agent Design

In a chapter of the handbook for computational economics (Tesfatsion and Judd 2007), Robert Marks (2007) reviews recent work in market design using agent-based simulation. *Market design* is the branch of economic research aiming to provide insights about which protocol, i.e. interaction structure and information circulation rules, is the best to obtain certain characteristics of a market. As said repeatedly in this chapter, this choice is crucial in having certain parts of a population gain more power than others, or having efficiency attained in a short time. Hence, many scientists have been thinking about this issue, using game theory (Roth et al. 1991), as well experimental economics, and more recently computational exploration. As seen before, sophisticated agents are not the ones who do best in market situations or copy human behaviour closest.

When designing a market protocol it is important to see two challenges. First the "aim" of the protocol needs to be clear since not all positive aspects can be achieved in a single protocol (see for example Myerson and Satterthwaite 1983). For example, using Dutch auction has the advantage of being fast, whereas double auction is good because it extracts the highest global profit for all. On the other hand, one might wish to extract the highest profit for buyers only, for example. LeBaron (2001) explains that the fitness of a model is as important as all other elements (what is traded, the motivations of agents, how the interaction and information circulation is organised, etc.). To achieve this trade-off between different characteristics is already a huge choice before starting the design.

Then one has to think on how to achieve this aim. It is indeed not easy to know how individuals will react to an interaction and information constraint. The basic use of agent-based simulation can then be to either test a certain agent behaviour and compare protocols to see what difference it makes in prices or other indicators (Moulet and Kirman 2008) or to test different learning algorithms in the same setting (Chan and Shelton 2001). Both approaches are uniquely developed using agent-based simulation, and can indeed help understand the relation between participant behaviour and market protocol.

Many models, be it for computer scientists or economists, were designed to fit the context of the *electricity market*, which is crucial since problems can be very severe for society (when there are huge unpredicted shortages) and the variations in price can be very fast. The agents in those models are not designed to represent human rationality but to try to be as optimal as possible in the adaptation to the electricity market. Many market protocols can be used, although auctions (which are theoretically the most efficient of all market protocols) are most common. Bidding behaviours, but also the number of sellers and buyers, and the capacity to produce and sell (Nicolaisen et al. 2001) have an impact on the efficiency and this can be explored. As said before, what is explored is the impact of the protocol on efficiency and market power. Two ways of learning are commonly used for the agents, and authors sometimes disagree over which one to choose: either social with a Genetic Algorithm, or individual with reinforcement learning. While it is already well known that this has a huge impact on global results (Vriend 2000), in this chapter we cannot decide on the best choice to make. However, to our view, most results cannot really be extended to real life design since the representation of learning for agents can be badly adapted to the application context (necessity to have long learning in case of GA or even reinforcement learning).

One original approach that is cited by Marks (2007) is the "evolutionary mechanism design" (Phelps et al. 2002), where the strategies of three types of actors – sellers, buyers and auctioneers – are all submitted to evolution and selection (the fitness of the auctioneer's strategy being linked to the total profit of the participants). This approach is logically different since the protocol itself (via the auctioneer) is what evolves to get to a better result, with the characteristics of the participants being fixed (relative number of each and relative production and demand).

It is interesting to note that another branch of research deals with the representation of individual agents on large markets, and is also quite close to an idea of design of markets, but from the opposite perspective: by introducing agents into real markets. Computer scientists interested in the analysis of cognition have the goal of making artificial agents as efficient as possible in a context of bidding in auctions, both from the point of view of the seller and the buyer (Kephart and Greenwald 2002). They are usually not interested in understanding human behaviour and decisions, but rather in explaining the properties that can emerge in markets in which many artificial learning agents interact (with each other or humans), differentiating their strategies, getting heterogeneous payoffs and creating interesting price dynamics. The focus lies mainly on information treatment. This applied approach is interesting in that many of its algorithms can also be used for economic analysis in the framework of models of the type that have been explored here. However, the aim is slightly different, since the indicator in the latter case is the individual success of a strategy, whereas the indicators for the previous works on markets are based on global properties of the system.

## 23.7 Concluding Remarks

This chapter is not a general review of market simulation in recent research; instead of giving many examples, we focused on a few to show the diversity of questions, models, rationality and eventual results that can be found in the literature, coming from different backgrounds (classical economy, experimental economy, computer science). The representation of a market is always linked to the purpose of the simulation study, and there is never just one way forward. The quantity and substitutability of goods, the possibility to interact with one or several sellers, with other buyers, the memory of the agents themselves – all depends on the type of issue, and this is why we have built the chapter in this manner: to give some ideas of the issues that have been addressed up until now with agent-based simulation. What is noticeable is the real difference between this approach and the classical approach in economics, where the dynamics are not regarded as a central question. However, the achievements with this new approach are now numerous enough to prove that agent-based simulation can participate in a better understanding of market protocols and behaviours of individuals on the market, and enhance the institutional choices of politics. What can be noted in conclusion is that several issues are still at stake when it comes to the representation of markets.

First, like with most simulation models, the temporal issue is huge. Most models use discrete time to advance the simulation. This can lead to problems, for example in an auction, where different agents might act precisely at the same time and have a different impact on prices than when they act sequentially. Some people are specifically working on this issue and build platforms that support a simulated continuous time[4] (Daniel 2006).

Another technical issue is the one of learning sequences of actions. In a situation where agents evaluate their actions with profit, if they have to perform several actions in a row (i.e. choosing a seller and then accepting a price or not), it is impossible to decide which of these actions is the reason for a success or a failure. Facing this issue, economic papers describe agents that associate the profit to all actions, as if they were separated. This is clearly not very satisfying in terms of logic, but no alternative modelling has been proposed yet.

Finally, there is a conceptual gap in all the cited models. As yet, another element has never been taken into account in the representation of agents' reasoning on markets, which would fit in models where agents try to maximize their profit by choosing the best strategies. In this case, they can scan past actions and the following profits, or their past possible profit with all actions they could have undertaken and then select the best action in all contexts. While the latter strategy is a bit more general than the first one, neither lets the agents imagine that a change in their action will modify other agents' behaviour as well. This is strange enough, since a lot of people interested in game theory have been working on agents in markets, but none of them have produced models of anticipation of others' choices.

[4] Natlab, which can be found at: http://www.complexity-research.org/natlab.

In markets where bargaining is central, it could however be a central feature in the understanding of real human behaviour.

**Acknowledgements** I wish to thank Bruce Edmonds for his patience, and Scott Moss and Sonia Moulet for their advices.

## Further Reading

Arthur (1991) is one of the first models incorporating learning agents in a market. Lux (1998) describes a model of speculation on an asset market with interacting agents. Duffy (2001) was the first attempt to link experimental data to simulation results in order to evaluate the kind of learning within a speculative environment. Jefferies and Johnson (2002) give a general overview of market models including their structures and learning by agents. Moulet and Rouchier (2007) use data on negotiation behaviours from a real market in order to fit the parameters of a two-sided learning model. Finally, Kirman (2010) summarises many interesting dimensions that can be captured using agent-based models.

## References

Arifovic J (1996) The behaviour of the exchange rate in the genetic algorithm and experimental economies. J Polit Econ 104(3):510–541

Arthur WB (1991) On designing economic agents that behave like human agents: a behavioural approach to bounded rationality. Am Econ Rev 81:353–359

Arthur WB (1994) Inductive reasoning and bounded rationality. Am Econ Rev 84:406

Arthur WB, Durlauf S, Lane D (eds) (1997a) The economy as an evolving complex system II (Santa Fe institute studies in the sciences of complexity, XXVII). Addison-Wesley, Reading

Arthur WB, Holland J, LeBaron B, Palmer R, Taylor P (1997b) Asset pricing under endogenous expectations in an artificial stock market. In: Arthur et al (1997a), pp 15–44.

Axtell R (2005) The complexity of exchange. Econ J 115(June):193–210

Basci E (1999) Learning by imitation. J Econ Dyn Control 23:1569–1585

Becker G (1962) Irrational behaviour and economic theory. J Polit Econ 70:1–13

Bendor J, Mookherjee D, Ray D (2001) Aspiration-based reinforcement learning in repeated interaction games: an overview. Int Game Theory Rev 3(2–3):159–174

Boylan R, El-Gamal M (1993) Fictitious play: a statistical study of multiple economic experiments. Game Econ Behav 5:205–222

Brenner T (2002) A behavioural learning approach to the dynamics of prices. Comput Econ 19:67–94

Brenner T (2006) Agent learning representation: advice on modelling economic learning. In: Tesfatsion and Judd (2006), Chapter 18

Brewer PJ, Huang M, Nelson B, Plott C (2002) On the behavioural foundations of the law of supply and demand: human convergence and robot randomness. Exp Econ 5:179–208

Brock WA, Hommes C (1998) Heterogeneous beliefs and routes to chaos in an asset pricing model. J Econ Dyn Control 22:1235–1274

Camerer C, Ho TH (1999) Experience-weighted attraction learning in normal form games. Econometrica 67:827–874

Chan NT, Shelton CR (2001) An electronic market-maker. Technical report, AI Memo 2001–005, MIT AI Lab, Cambridge

Cyert RM, March JG (1963) A behavioural theory of the firm. Prentice-Hall, Englewood Cliffs

Daniel G (2006) Asynchronous simulations of a limit order book. PhD thesis, University of Manchester, http://gillesdaniel.com/PhD.html

Darmon E, Torre D (2004) Adoption and use of electronic markets: individual and collective learning. J Artif Soc Soc Simul, 7(2), http://jasss.soc.surrey.ac.uk/7/2/2.html

Delre SA, Jager W, Bijmolt THA, Janssen MA (2007) Targeting and timing promotional activities: an agent-based model for the takeoff of new products. J Bus Res 60:826–835

Duffy J (2001) Learning to speculate: experiments with artificial and real agents. J Econ Dyn Control 25:295–319

Duffy J, Ochs J (1999) Emergence of money as a medium of exchange: an experimental study. Am Econ Rev 89:847–877

Easley D, Ledyard J (1993) Theories of price formation and exchange in double oral auction. In: Friedman D, Rust J (eds) The double-auction market: institutions, theories, and evidence. Addison-Wesley, Reading, pp 63–97

Erev I, Bereby-Meyer Y, Roth A (1998) Prediction how people play games: reinforcement learning in games with unique strategy equilibrium. Am Econ Rev 88:848–881. http://ie.technion.ac.il/Home/Users/erev/erev_roth_aer98.pdf

Friedman M (1953) The methodology of positive economics. In: Friedman M (ed) Essays in positive economics. The University of Chicago Press, Chicago, pp 3–43 [Republished in Breit W, Hochman HM (eds) (1971) Readings in microeconomics, 2nd edn. Holt, Rinehart and Winston, New York, pp 23–47]

Geertz C, Geertz H, Rosen L (1979) Meaning and order in Moroccan society: three essays in cultural analysis. Cambridge University Press, Cambridge

Giardini F, Conte R, Paolucci M (2013): Reputation. Chapter 15 in this volume

Gintis H (2006) The emergence of a price system from decentralized bilateral exchange. B. E. J Theor Econ 6(1):1–15, http://works.bepress.com/hgintis/1/

Gode DK, Sunder S (1993) Allocative efficiency of markets with zero intelligence traders: markets as a partial substitute for individual rationality. J Polit Econ 101:119–137

Gode DK, Sunder S (2004) Double-auction dynamics: structural effects of non binding price controls. J Econ Dyn Control 28:1707–1731

Hoffmann AOI, Jager W, von Eije JH (2007) Social simulation of stock markets: taking it to the next level. J Artif Soc Soc Simul 10(2), http://jasss.soc.surrey.ac.uk/10/2/7.html

Hommes C (2001) Financial markets as complex adaptive evolutionary systems. Quant Financ 1:149–167

Hommes C (2007) Bounded rationality and learning in complex markets (Working paper, 07–01), Center for Nonlinear Dynamics in Economics and Finance (CeNDEF), University of Amsterdam, http://www1.fee.uva.nl/cendef/publications/papers/Handbook_Hommes.pdf

Hommes C, Lux T (2008) Individual expectations and aggregate behavior in learning to forecast experiments (Working paper, 1466), Kiel Institute for the World Economy http://www.ifw-members.ifw-kiel.de/publications/individual-expectations-and-aggregate-behavior-in-learning-to-forecast-experiments/KWP_1466_Individual%20Expectations.pdf

Izquierdo LR, Gotts NM, Gary Polhill J (2004) Case-based reasoning, social dilemmas, and a new equilibrium concept. J Artif Soc Social Simulat 7(3). http://jasss.soc.surrey.ac.uk/7/3/1.html

Jager W (2007) The four P's in social simulation, a perspective on how marketing could benefit from the use of social simulation. J Bus Res 60:868–875

Janssen M, Jager W (2001) Fashion, habits and changing preferences: simulations of psychological factors affecting market dynamics. J Econ Psychol 22:745–772

Jefferies P, Johnson NF (2002) Designing agent-based market models. http://arxiv.org/abs/cond-mat/0207523

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 2:263–292
Kawagoe T, Wada S (2005) A counterexample for the bullwhip effect in a supply chain. In: Mathieu P, Beaufils B, Brandouy O (eds) Artificial economics, agent-based methods in finance, game theory and their applications (Lecture notes in economics and mathematical systems, 564), Springer, Berlin, pp 103–111
Kephart J, Greenwald A (2002) Shopbot economics. Auton Agent Multi-Agent Syst 5(3):255–287
Kirman A (1991) Epidemics of opinion and speculative bubbles in financial markets. In: Taylor M (ed) Money and financial markets. Macmillan, London, pp 354–356, Chapter 17
Kirman A (2001) Some problems and perspectives in economic theory. In: Debreu G, Neuefeind W, Trockel W (eds) Economic essays: a Festschrift for Werner Hildenbrand. Springer, Berlin, pp 231–252
Kirman A (2010) Complex economics: individual and collective rationality. Routledge, London
Kirman A, Schulz R, Hardle W, Werwatz A (2005) Transactions that did not happen and their influence on prices. J Econ Behav Org 56:567–591
Kirman A, Vriend NJ (2001) Evolving market structure: an ACE model of price dispersion and loyalty. J Econ Dyn Control 25(3–4):459–502
Kiyotaki N, Wright R (1989) On money as a medium of exchange. J Polit Econ 97:924–954
Kopel M, Dawid H (1998) On economic applications of the genetic algorithm: a model of the cobweb type. J Evol Econ 8(3):297–315
LeBaron B (2001) Evolution and time horizons in an agent-based stock market. Macroecon Dyn 5 (2):225–254
LeBaron B, Arthur WB, Palmer R (1999) The time series properties of an artificial stock market. J Econ Dyn Control 23:1487–1516
Leloup B (2002) Dynamic pricing with local interactions: logistic priors and agent technology. In: Arabnia HR, Mun Y (eds) Proceedings of the international conference on artificial intelligence, IC-AI '02, vol 1. CSREA Press, Las Vegas, 24–27 June 2002, pp 17–23
Levy M, Levy H, Solomon S (2000) Microscopic simulation of financial markets: from investor behaviour to market phenomena. Academic, San Diego
Lin F-R, Lin S-M (2006) Enhancing the supply chain performance by integrating simulated and physical agents into organizational information systems. J Artif Soc Soc Simul 9(4), http://jasss.soc.surrey.ac.uk/9/4/1.html
Lux T (1998) The socio-economic dynamics of speculative markets: interacting agents, chaos and the fat tails of return distribution. J Econ Behav Organ 33(2):143–165
Lux T (2009) Stochastic behavioral asset-pricing models and the stylized facts. In: Hens T, Schenk-Hoppé KR (eds) Handbook of financial markets: dynamics and evolution. Elsevier, Burlington, pp 161–215
Lux T, Marchesi M (1999) Scaling and criticality in a stochastic multi-agent model of a financial market. Nature 397:498–500
Lux T, Marchesi M (2000) Volatiliy clustering in financial markets: a micro-simulation of interacting agents. Int J Theor Appl Financ 3:675–702
Macy M, Flache A (2002) Learning dynamics in social dilemmas. Proc Natl Acad Sci 99:7229–7236
Macy MW, Benard S, Flache A (2013) Learning. Chapter 17 in this volume
Marks R (2007) Market design using agent-based models. In: Tesfatsion L, Judd KL (eds) Handbook of computational economics volume 2: agent-based computational economics. Elsevier/North Holland, Amsterdam, pp 1339–1380
Moss S (2002) Policy analysis from first principles. PNAS 99(3):7267–7274
Moulet S, Kirman A (2008) Impact de l'organisation du marché: Comparaison de la négociation de gré à gré et les enchères descendantes (Document de travail, 2008–56). GREQAM, Marseille
Moulet S, Rouchier J (2008) The influence of sellers' beliefs and time constraint on a sequential bargaining in an artificial perishable goods market. J Econ Dyn Control 32(7):2322–2348

Myerson RB, Satterthwaite MA (1983) Efficient mechanisms for bilateral trading. J Econ Theory 29:265–281

Nicolaisen J, Petrov V, Tesfatsion L (2001) Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. IEEE Trans Evolut Comput 5:504–523

Oliver J (1997) Artificial agents learn policies for multi-issue negotiation. Int J Electron Commer 1 (4):49–88

Palmer R, Arthur WB, Holland JH, LeBaron B, Taylor P (1993) Artificial economic life: a simple model for a stock market. Physica D 70:264–274

Phan D (2003) From agent-based computational economics towards cognitive economics. In: Bourgine P, Nadal J-P (eds) Cognitive economics. Springer, Berlin, pp 371–398

Phelps S, Parsons S, McBurney P, Sklar E (2002) Co-evolution of auction mechanisms and trading strategies: towards a novel approach to microeconomic design. In: Smith RE, Bonacina C, Hoile C, Marrow P (eds) Proceedings of ECOMAS-2002 workshop on evolutionary computation in multi-agent systems, at genetic and evolutionary computation conference (GECCO-2002), International society for genetic and evolutionary computation

Pinyol I, Paolucci M, Sabater-Mir J, Conte R (2007) Beyond accuracy: reputation for partner selection with lies and retaliation. In: Antunes L, Paolucci M, Norling E (eds) Multi-agent-based simulation VIII, International workshop, MABS 2007, Honolulu, 15 May 2007, Revised and invited papers. (Lecture notes in computer science, 5003). Springer, Berlin, pp 128–140

Roth A, Prasnikar V, Okuno-Fujizara M, Zamie S (1991) Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. Am Econ Rev 81(5):1068–1095

Rothschild M (1974) A two-armed Bandit theory of market pricing. J Econ Theory 9:185–202

Rouchier J, Bousquet F, Requier-Desjardins M, Antona M (2001) A multi-agent model for describing transhumance in North Cameroon: comparison of different rationality to develop a routine. J Econ Dyn Control 25:527–559

Rouchier J (2003) Reimplementation of a multi-agent model aimed at sustaining experimental economic research: the case of simulations with emerging speculation. J Artif Soc Soc Simul 6(4), http://jasss.soc.surrey.ac.uk/6/4/7.html

Rouchier J, Hales D (2003) How to be loyal, rich and have fun too: the fun is yet to come (Working Paper, 03B13). Groupement de Recherche en Économie Quantitative d'Aix-Marseille (GREQAM), France, http://www.greqam.fr/IMG/working_papers/2003/03b13.pdf

Rouchier J, Mazaud J-P (2004) Trade relation and moral link in transactions among retailers and wholesale sellers on the Arnavaux market. In: 11th world congress for social economics, Albertville, 8–11 June 2004

Rouchier J, Robin S (2006) Information perception and price dynamics in a continuous double auction. Simul Gaming 37:195–208

Rouchier J, Bousquet F, Requier-Desjardins M, Antona M (2001) A multi-agent model for describing transhumance in North Cameroon: comparison of different rationality to develop a routine. J Econ Dyn Control 25:527–559

Rouchier J (2013) The interest of having loyal buyers in a perishable market. Comput Econ 41(2):151–170

Rubinstein A, Wolinksy A (1985) Equilibrium in a market with sequential bargaining. Econometrica 53(5):1133–1150

Sallans B, Pfister A, Karatzoglou A, Dorffner G (2003) Simulation and validation of an integrated markets model. J Artif Soc Soc Simul 6(4), http://jasss.soc.surrey.ac.uk/6/4/2.html

Samanidou E, Zschischang E, Stauffer D, Lux T (2007) Agent-based models of financial markets. Rep Prog Phys 70:409–450

Simon H (1955) A behavioural model of rational choice. Q J Econ 69:99–118

Takahashi H, Terano T (2003) Agent-based approach to investors' behavior and asset price fluctuation in financial markets. J Artif Soc Soc Simul 6(3), http://jasss.soc.surrey.ac.uk/6/3/3.html

Tesfatsion L, Judd KL (2007) Handbook of computational economics volume 2: agent-based computational economics. Elsevier/North Holland, Amsterdam
Vriend NJ (2000) An illustration of the essential difference between individual and social learning and its consequences for computational analyses. J Econ Dyn Control 24:1–19
Vriend NJ (2007) ACE models of endogenous interactions. In: Tesfatsion L, Judd KL (eds) Handbook of computational economics volume 2: agent-based computational economics. Elsevier/North Holland, Amsterdam, pp 1047–1079
Weisbuch G, Kirman A, Herreiner D (2000) Market organization and trading relationships. Econ J 110(463):411–436
White H (1988) Varieties of markets. In: Wellman B, Berkowitz SD (eds) Social structures: a network approach. Cambridge University Press, New York

# Chapter 24
# Movement of People and Goods

**Linda Ramstedt, Johanna Törnquist Krasemann, and Paul Davidsson**

**Why Read This Chapter?** To gain an overview of approaches to the simulation of traffic and transportation by way of representative examples. Also, to reflect the characteristics and benefits of using social simulation as opposed to other methods within the domain. The chapter will inform both researchers and practitioners in the traffic and transportation domain of some of the applications and benefits of social simulation and relevant issues.

**Abstract** Due to the continuous growth of traffic and transportation and a thus increased urgency to analyze resource usage and system behavior, the use of computer simulation within this area has become more frequent and acceptable. This chapter presents an overview of modeling and simulation of traffic and transport systems, and focuses in particular on the imitation of social behavior and individual decision making in these systems. We distinguish between *transport* and *traffic*. *Transport* is an activity where goods or people are moved between points A and B while *traffic* is referred to as the collection of several transports in a common network such as a road network. We investigate to what extent and how the social characteristics of the users of these different traffic and transport systems are reflected in the simulation models and software. Moreover, we highlight some trends and current issues within this field and provide further reading advice.

L. Ramstedt (✉)
Transport and Society, Vectura Consulting, Solna, Sweden
e-mail: linda.ramstedt@vectura.se

J.T. Krasemann
School of Computing, Blekinge Institute of Technology, Karlshamn, Sweden
e-mail: johanna.tornquist@bth.se

P. Davidsson
Department of Computer Science, Malmö University, Malmö, Sweden
e-mail: paul.davidsson@mah.se

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_24,

## 24.1 Introduction

The continuous growth of traffic and transportation is increasing the interest in limiting their negative impact on society. Moreover, the different stakeholders involved in traffic and transportation are interested in utilizing the available resources in the best way possible. This has stimulated the development and use of both various policies and advanced transport infrastructure systems, as well as the deployment of information and communication technologies. In order to examine the effects of such developments prior to, or during implementations, computer simulations have shown to be a useful approach. This chapter addresses modeling and simulation of traffic and transport systems, and focuses in particular on the imitation of social behavior and individual decision-making.

Traffic and transport systems typically involve numerous different stakeholders and decision-makers that control or somehow affect the systems, and the prevalence of social influence is thus significant. In order to study the behavior of such systems and model them it then becomes necessary to capture and include the significant social aspects to some extent, in addition to the flow of traffic or transport units that constitute the backbone of such systems. In this context, we consider a traffic or transport system as a *society*, consisting of *physical components* (e.g., cars, buses, airplanes or parcels) and *social components* (e.g., drivers, passengers, traffic managers, transport chain coordinators, or even public authorities) where the interactions between them may play an important role. The social components determine the physical flow in their common *environment* (e.g., road or rail networks, and terminals) in line with *external restrictions*, *internal intentions* and so forth. Depending on the purpose of the simulation study and the sophistication and detail that is desired in the model, there are different approaches to incorporate social influence in such systems.

The purpose of this chapter is to present an overview of when and how social influence and individual behavior within the domain of traffic and transportation have been modeled in simulation studies. We also provide further reading advice to related approaches and highlight current issues. We divide the domain into transport and traffic. *Transport* is an activity where goods or people are moved between points in one or several traffic modes (road, rail, waterborne, or air). The types of vehicles we consider are train, truck, car, bus, ship, ferry, and airplane. While transport refers to the movement of something from one point to another, *traffic* refers to the flow of different transports within a network. One train set is thus a transport (and possibly also part of a transport chain) that takes part in the train traffic flow. Hence, a transport can be part of several traffic networks (air, waterborne, road or rail) and a traffic network comprises several transports.

Typical points of interest related to transportation in this context are, for instance, to predict consequences of transport policies aimed at mediating governmental goals or objectives of companies, and to design and synchronize bus timetables at certain stations to facilitate passenger flows. A typical issue that is interesting to study in the traffic context is the impact of driving assistance systems where the driving behavior in road networks is studied.

In the following sections, the types of simulation approaches that exist within the traffic and transportation domain, and the motivation behind them, are presented.

The social aspects of the studied systems and models, and how they have been accounted for are also described here. Finally, a concluding discussion is presented along with some further reading advice.

## 24.2 Traffic

This section addresses the modeling and simulation of individual and social behavior within traffic systems. There are several types of simulation models developed and applied which have different granularity; i.e., macro-, meso- and micro-level models. The first two typically represent the traffic behavior by the use of equations which are based on aggregated data. Thus, the traffic is modeled as a collection of rather homogenous entities in contrast to the microscopic models which more in-depth consider the *individual* characteristics of the traffic entities and how these influence each other and the traffic system. Since this handbook mainly addresses the modeling of social aspects, we will focus on microscopic models.

Within the domain of traffic system simulation, the dominating focus is on simulation of road traffic and car driver behavior to evaluate the quality-of-service in road traffic networks (Tapani 2008). The development and implementation of ADAS (Advanced Driver Assistance Systems), ATIS (Advanced Traveler Information Systems), and road traffic control regimes, such as congestion taxation, stipulates an increasing interest in sophisticated simulation models. There are also approaches that study the behavior of traffic systems during extraordinary situations such as urban evacuations. Since users of road traffic systems normally do not communicate and interact with each other directly but rather indirectly due to the restrictions of the traffic system, the design of agent communication and protocols is mostly not considered in these simulation models. Focus is instead on individual behavior models.

The attention given to the simulation of social aspects in other modes of traffic such as air and railway traffic is, however, very limited and we have not found any publications focusing on this specific subject. One reason may be that the traffic in these modes to a large extent is managed by central entities, such as, traffic controllers. In this perspective, the behavior and interaction of the individual vehicles are less interesting to study. However, the interaction between traffic controllers in these traffic modes seems to be a domain that needs more attention.

Below we will provide a more in-depth presentation of social simulation in road traffic. Since the movement of people associated with a vehicle is the focus in this chapter, simulation of pedestrians and crowds will only be presented briefly in the last section on related research.

### *24.2.1 Road Traffic Simulation*

In road traffic simulations, there are several distinctions made. First, there is a distinction made with respect to the problem area in focus and whether it concerns *urban, rural* or *motorway* traffic. Furthermore, depending on the infrastructure modeled, there is a distinction between *intersection, road* and *network* model. As an example, urban traffic is often modeled as a network (Pursula 1999).

One can see that there are three main categories of simulation models with increasing level of detail in driving behavior. First, there are the empirical macroscopic traffic models (e.g., software like TRANSYT provided by TRL Software) that focus on traffic flow analysis. Then we have the extended microscopic models with capabilities of representing individual driver behavior by use of different sub-models or rules for speed adaptation, car-following, lane change, as well as intersection and roundabout movements if relevant (e.g., VISSIM for urban and freeway environments and TRARR, TWOPAS, and VTISim for rural road environments (Tapani 2008)). The third type of models has an even more complex representation of driver behavior by use of, e.g., neural networks (NN) (Lee et al. 2005; Panwai and Dia 2007) or discrete choice models (Dia 2002; Lee et al. 2005). In some cases these behavior models can be dynamically configured to imitate driver behavior adaptation and the effect of learning from experience, see, e.g., the work by Rossetti et al. (2000).

These three types of models do complement each other, but with the growing need to evaluate the impact of investments in ITS (Intelligent Transport Systems) such as ATIS and ADAS, and policies such as road pricing, the need to reflect the complexity of driver behavior in more detail becomes apparent (Tapani 2008). The current behavior sub-models for acceleration, lane changing, car-following and so forth are mainly equation-based with threshold values. Toledo (2007) claims in a review of state-of-the-art in road traffic driver behavior modeling that, in many cases these sub-models are insufficient to adequately capture the sophistication of drivers and the impact of long-term driving goals and considerations. Henceforward, we will focus on the approaches that emphasize such advanced driving behavior modeling and refer to Pursula (1999), Mahmassani (2005), Toledo (2007) and Tapani (2008) for more information about related research.

In the third category of approaches mentioned, the individual drivers are usually modeled as individual autonomous vehicles represented by *intelligent agents*. The main differences between the equation- or rule-based behavior sub-models and the agent-based models are the increased reasoning capabilities and planning horizon considered. The traffic network flow could in either case be based on techniques like *cellular automata (CA)* (Nagel and Schreckenberg 1992; Esser and Schreckenberg 1997) and *queuing theory* while the decision-making of the agent is based on a possibly more diverse set of objectives and influencing parameters. Sometimes the agents use simple decision rules acting in a reactive manner rather than having sophisticated reasoning capabilities. We refer to Ehlert and Rothkrantz (2001), El hadouaj et al. (2000), Wahle and Schreckenberg (2001), and Kumar and Mitra (2006), for examples of such reactive behavior. However, these approaches are not considered part of the third category in this context and this illustrates that agent technology does not necessarily imply, but only *enables*, a representation of more complex behavior.

The approaches and their models often focus on a certain traffic infrastructure depending on which aspect to study. Below we address three main topics: Simulation of road traffic in intersections, evaluation of ATIS and ADAS, and benchmarking of behavior modeling techniques.

#### 24.2.1.1 Simulation of Road Traffic in Intersections

Urban traffic networks typically involve intersections with a complex coordination and interplay between vehicles. This is challenging to model and it is difficult to prevent deadlocks from occurring during the simulation. One way to handle this is to use game theory and agents. Mandiau et al. (2008) propose a distributed coordination and interaction mechanism for the simulation of vehicles in T- and X-shaped intersections in urban road traffic networks. The vehicles are represented by agents who have a choice of one of two actions: brake or accelerate. The coordination between the vehicles crossing the intersection is based on game theory and a $2x2$ decision matrix for each pair of vehicles is used to compute the decisions. The approach is also extended to involve a larger traffic volume of $n$ agents (vehicles), where the memory requirements then increases rapidly due to the $n(n - 1)/2$ decision matrices. The mechanism does not prevent occurrence of deadlocks, but is able to resolve them.

Bazzan (2005) does also propose a distributed game theory-based approach for coordination of road traffic signals at X-shaped intersections where vehicles are represented by autonomous intelligent agents. The agents have both local and global goals and are adaptive in the sense that they are equipped with a memory to register the outcome (i.e. payoff) of executed actions and via learning rules they are able to incorporate their experience in the decision-making. Game theory has also been used in this context to model the route choice behavior (Schreckenberg and Selten 2004).

#### 24.2.1.2 Evaluation of ATIS and ADAS

The benefits from and application of ATIS and ADAS have become more known and common in modern road traffic systems. An implementation is, however, often associated with large investments and limited insight in how the system and its users would respond to such an implementation in the short and long term. The use of simulation may then offer some information on possible consequences to guide the stakeholders. A typical question to address is how the individual drivers would change their choice of traveling with an increased access to traffic information. Dia (2002) investigates the implications of subjecting drivers to real-time information in a road traffic network. Each vehicle is represented by a BDI agent (Beliefs-Desires-Intentions), which has a static route-choice multi-nomial logit model (Ben-Akiva and Lerman 1985). Static refers to that the multi-nomial logit model is not updated during simulation and thus not making use of driver/vehicle experience. Panwai and Dia (2005) extend the approach using NN and then further improve it by giving each agent a memory and a dynamic behavior model based on NN (Dia and Panwai 2007). The behavior model is updated accordingly during the agent's journey allowing the agent to act on the information provided by an ATIS by re-evaluating its strategy en-route and possibly changing route if possible and beneficial.

Rossetti et al. (2000) focus on the departure time and route choice model of commuters in a road traffic network using the DRACULA (Dynamic Route Assignment Combining User Learning and microsimulAtion) simulation model. The drivers are represented by autonomous BDI agents, some of which have access to ATIS and thus more information about the current traffic situation than other drivers. The agent behavior model is dynamic in the sense that it incorporates the experience of the commuting driver on a daily basis, but does not allow the driver to change strategy en-route. Before the commuter starts its journey and decides when to depart and which route to choose, it compares the predicated cost of choosing its usual daily route with the cost of the alternative route. If the cost of any of the alternatives is significantly lower than the cost of the daily route, the driver chooses that. Otherwise, it stays with its usual route. Once the agent has decided on departure time and route, it will follow that guided by car-following and lane-changing rules and cannot change its mind.

Chen and Zhan (2008) use the agent-based simulation system Paramics to simulate the urban evacuation process for a road traffic network and the drivers' route choice and driving behavior. The vehicles are modeled as individual agents that have local goals to minimize their travel time, i.e. to choose the fastest way. Based on traffic information, a car-following model, network queuing constraints and a behavior profile (e.g. aggressive or conservative driving style) each agent dynamically, en-route, re-evaluates its driving strategy and route choice.

#### 24.2.1.3 Benchmarking of Behavior Modeling Techniques

Since a number of alternative methods to model and simulate road traffic exist, a few studies have focused on comparisons to evaluate the strengths and weaknesses of some alternatives. Lee et al. (2005) compare and evaluate the use of three different route choice models for road traffic drivers with access to trip information: A traditional multinomial logit model, a traditional NN approach and NN combined with a genetic algorithm (GA). The initial attitude of the authors indicates a preference for the combined NN-GA solution proposing that it is better suited to consider the influence of non-linearity and obscurity in drivers' decision-making. Based on their simulation experiments and the mean square error of the different route choice models, the NN-GA approach was said to be most appropriate.

Hensher and Ton (2000) make a similar comparison of the predictive potential of NN and nested logit models for commuter mode choice. They could not make any judgment about which method that is most appropriate but concluded that both requires a lot of historic data to train or construct the models. The characteristics of NN are that they are better at handling noisy or lack of data than discrete choice models were discussed.

For more in-depth information about human behavior in traffic we refer to Schreckenberg and Selten (2004) while Boxill and and Yu (2000) present an overview of road traffic simulation models and Koorey (2002) an overview of software.

## 24.3 Transportation

In this section, approaches to simulating transportation systems are described, which include both transportation of freight and people. Transportation is often described as road, rail, waterborne or air transportation, but transportation can also be *intermodal*. Intermodal transportation refers to a transport chain of two or more modes of transport where some modal shift activity takes place, for instance at a terminal where loading and unloading of goods is done, or at a train station where passengers transfer from train to bus. *Supply chains* are related to freight transportation, even if the focus mainly is on the product and its refinement processes, in contrast to freight transportation where the focus is on the vehicle and its operations.

Issues in the field of passenger transportation typically concern evaluation of policies for more efficient bus timetables and pricing policies. Another field within the domain is *emergency transportation*, which often concerns the planning of resources, such as ambulances or fire engines, in order to serve people in need efficiently with respect to costs, coverage equity and labor equity. Other issues are the evaluation of different policies, such as dispatching policies. Goldberg (2004) has reviewed operations research approaches for emergency transportation, and claims that mathematical programming currently is the dominating method used. He also states that simulation is a promising approach for future work related to emergency transportation, especially for vehicle relocation and dispatching, due to the complexity of the problem domain. However, we have not found any papers describing approaches of social simulation in the field of emergency transportation.

In papers describing simulation approaches to transportation with a focus on social aspects and individual behavior, we have only found models for road, rail, and waterborne transportation, i.e., we have not found any air transportation models.

### *24.3.1 Freight Transportation*

Different approaches are used when modeling and simulating transportation. A common approach when analyzing transportation is the so-called four-step approach (production/attraction, distribution, modal split, and assignment) which primarily is developed for passenger transportation, but also used for freight transportation (de Jong and Ben-Akiva 2007). This approach is on the macroscopic level, i.e. averaged characteristics of the population are in focus and aggregated data is used, in contrast to a microscopic level, which would include more details on the individuals of the population. A trend in the field of freight transport simulation for predicting probable consequences of transport policies is to include more details, see e.g. (de Jong and Ben-Akiva 2007) and (Hunt and Gregor 2008). However, most of these approaches are still macroscopic approaches, but with some microscopic characteristics (Ramstedt 2008).

Traditional simulation approaches for simulating supply chains are discrete-event simulation and dynamic simulation (Terzi and Cavalieri 2004). In such models the behavior of the individuals is often only modeled as a set of actions related to a probability function of being executed, not capturing causal behavior of the simulated system.[1] Moreover, interactions between the individuals, such as negotiations, are not explicitly modeled in traditional supply chain models. Since modeling and simulating social aspects is the main concern here, we focus on simulation approaches that address interactions between individuals.

Most simulation approaches of freight transportation have a descriptive purpose, such as predicting the effects of different kinds of policies. For instance, Gambardella et al. (2002) who simulate intermodal transportation make use of multi-agent-based simulation to examine policies aimed at improving the operations at terminals, while simulation approaches of supply chains (Swaminathan et al. 1998; van der Zee and van der Vorst 2005) often focus on evaluating strategies such as VMI (Vendor Managed Inventories). Such studies are mainly of interest for private companies, even if they can be of interest for public authorities as well. Another example is provided by Holmgren et al. (2012) which studies the possible effects of transport policies in transport chains, which is of interest for public authorities. In this approach the decision making of actors in transport chains, for instance regarding traffic mode choice, selection of supplier, etc., are simulated. New prerequisites as a consequence of transport policies have the potential to change these decisions so that a different system behavior occurs. There are also examples of models with a prescriptive purpose, such as to support the transport planning in order to improve the efficiency of the usage of transport resources (Fischer et al. 1999).

The simulated system in transport approaches typically consists of a network of links and nodes served by resources such as vehicles, which have a spatial explicitness and are time-dependent. In supply chains the focus is more on the nodes and their processes, while the links are not explicitly considered.

In the domain of freight transportation and supply chains, the decision making of stakeholders is typically modeled and simulated. To model the decision making, agents representing real-world roles in transportation, such as customers, transport planners, transport buyers, and producers are typically implemented. Only in a few cases are physical entities, such as vehicles, also modeled as agents (Gambardella et al. 2002). Tasks which are commonly performed by the agents are selecting (1) *which* resources (e.g. terminals, transport and production resources) to use, and (2) *how* these resources should be used considering time, cost, availability, etc. These tasks are often performed as a consequence of a customer request with an aim of satisfying the demand based on cost, time, and availability. The behaviors of the agents are often implemented in terms of various types of algorithms or decision rules. Typically the agents try to minimize their costs, e.g. labor or fuel costs, which occur as a consequence of performing a transport task between two nodes.

[1] See Chap. 3 in this handbook (Davidsson and Verhagen 2013) for further general discussion.

Of course restrictions and requirements from other agents – concerning for instance time of delivery – are also taken into account. The implemented algorithms in the agents can be rather complex, with optimization techniques and heuristics used to compute the decisions and actions (Holmgren et al. 2012; Fischer et al. 1999).

If some of the simulated agents represent physical entities, the locations of the entities as well as their characteristics are typically modeled. If the agents represent decision makers, the responsibilities of the agents are typically modeled, e.g. the responsibility of certain types of vehicles on certain infrastructure segments.

The interactions between the agents often take place as negotiations concerning, for instance, the cost and time of performing a task, such as transportation between two nodes. The negotiations are then carried out according to interaction protocols and the corresponding information exchange between the agents is also modeled. As an example, a customer agent requests information concerning possible transport solutions, or a transport planning agent requests information concerning available vehicles for the transport task (Holmgren et al. 2012).

The agent-based simulation models are implemented in different ways, in some cases multi-agent-based simulation platforms are used (e.g. Strader et al. 1998), while multi-agent system platforms are used in other cases (e.g. Holmgren et al. 2012; Fischer et al. 1999). It is also possible to implement the agent model without any platform, see e.g. (Swaminathan et al. 1998; Gambardella et al. 2002; van der Zee and van der Vorst 2005). Using such platforms may often facilitate the implementation of the model. However, if the model is very complex it can also cause problems due to the structure of the platform; see for example (Davidsson et al. 2008) for further discussion.

### *24.3.2 Passenger Transportation*

While more work has been done regarding modeling and simulation of freight transportation, there are some approaches concerning the transportation of passengers using rail and road. These are described here.

One example of the simulation of *bus transportation* is Meignan et al. (2007) where the main purpose of the simulations is to evaluate bus networks. The bus networks are assessed based on the interests of travelers, bus operators and authorities with respect to, for instance, accessibility, travel time, passenger waiting time, costs, and profit. Therefore, a potential type of user of the simulation model is the manager of a bus network. The model includes a road network, a bus network, and a traveler network. Bus stops, bus stations, bus lines and itineraries are part of the modeled system. There are two agent types: buses and passengers. A typical task of the bus agent is to perform a bus trip. The networks include spatial explicitness, and time is important for the agents since the bus routes are determined by timetables, and the passengers have a need for travel at certain points in time. The model combines micro- and macroscopic approaches since the global traffic situation is taken into consideration in addition to the individual transports. Traffic assignment and modal choice of the overall demand is made on the macroscopic level with a discrete choice model. Interactions take place between the buses and passengers. For instance, the bus agents have to consider the travel demand of passenger agents, and the passenger agents have to consider the available bus agents. Moreover, the actual loading and unloading of passengers is one kind of interaction. A similar approach is presented by Gruer et al. (2001) where the mean passenger waiting time at bus stops is evaluated. In this approach buses, stops, and road sections are modeled as agents, with the focus on the activities at the bus stops.

Work is also done in the *taxi domain* where for instance Jin et al. (2008) present a simulation model for planning the allocation of taxis to people requesting taxi transportation. The model can be used for evaluating different planning policies for how the allocation should be made taking issues like vehicle usage, passenger waiting time, travel time, and travel distance into consideration. Four types of agents are included in the model; the user agent, the node-station agent, the taxi agent, and the transport administration agent. The taxi agent represents the physical taxis, while the other agents have different planning functions. The different agent types have different goals and agreements are reached by negotiations between the agents through the user agent.

An example of a social simulation of *rail passenger transportation* is presented by Li et al. (2006). In this, an activity-based model is outlined for the evaluation of pricing policies and how they affect the traveler behavior. This approach is similar to an agent-based approach where passengers are modeled as agents. The focus in the model is on the traveler behavioral model, where the characteristics and preferences of travelers and their activities, in terms of activity schedules, are modeled. Typical tasks the traveler agent performs are scheduling and planning the journeys as well as executing these activities. The decisions are typically made based on generalized costs. In the presented model the traveler agents can interact with for instance a tool that provides information regarding available travel options by sending requests of possible travel options.

### *24.3.3 Related Research*

Related to the simulation of passenger transportation is the simulation of pedestrians and crowds. Traditionally, pedestrian simulation has used techniques such as flow-speed-density equations thus aggregating pedestrian movement into flows rather than a crowd of possibly heterogeneous individuals (Rindsfüser and Klügl 2007). Rindsfüser and Klügl (2007) propose an agent-based pedestrian simulation model for the simulation of the movement and behavior of train passengers at the railway station in Bern, Switzerland. A similar approach is Zhang et al. (2008), in which the alighting and boarding of passengers in the Beijing metro is modeled and simulated. A cellular automata approach is used in this model. Pelechano et al. (2005) also discuss social aspects in the simulation of crowds and provide a review of different crowd simulation approaches.

One special case of crowd simulation is the simulation of emergency evacuation situations where animal flocking behavior models are one type of models that have been applied (see Chap. 22 in this handbook (Hemelrijk 2013) for an overview of modeling animal social behavior). For further information on simulation of emergency evacuation we refer to (Santos and Aguirre 2004) which provides a review of current simulation models.

Another type of related work is agent-based simulation of seating preferences of passengers in buses and airplanes (Alam and Werth 2008). The different agents are then representing different categories of people characterized by ethnicity, age, cultural background, and their respective seating preferences.

## 24.4 Discussion

In papers concerning simulation of traffic and transportation, different arguments are given to support the use of social simulation. A common argument for making use of multi-agent-based simulation in the transportation domain is that it enables capturing the complex interactions between individuals, such as coordination and cooperation (Fischer et al. 1999; Meignan et al. 2007; Li et al. 2006), and consequently the emergent behavior of the system. Moreover, including autonomous and heterogeneous individuals and their behavior is also supported with the agent approach, as well as modeling and simulating distributed decision making which are important in the transportation domain (Meignan et al. 2007). Since multi-agent systems provide a modular structure of the system, the possibility to easily exchange or reuse different parts of the simulated system for different cases is facilitated (Swaminathan et al. 1998; van der Zee and van der Vorst 2005).

As pointed out earlier, there are new phenomena in the road traffic networks imposed by the increasing level of technologies facilitating driving as well as new motivations behind controlling and supervising networks and the effects of such support systems when the infrastructure capacity becomes a scarce resource. Toledo (2007) argues that to capture the level of sophistication, the modeling capabilities need to improve where the use of agents can make a contribution.

In contrast to well-established software and methods for traffic and transport simulations, the data requirements are different and data is not available to the same extent for the newer approaches mentioned in this chapter. In addition, the more novel approaches are all different and are often developed, used and evaluated only by the researchers themselves. The level of maturity and acceptance reached by the traditional approaches and software by being used and evaluated by a large number of researchers during a long time is naturally difficult to compete with at this point. However, using social simulation where individuals and their interactions are explicitly modeled provides opportunities for validation due to the natural, structure-preserving representation of the system. For instance, the behavioral models of drivers or decision makers can be validated by the actual drivers or real-world decision makers.

It is possible to identify some general differences between traffic and transportation applications. In traffic approaches, the agents typically represent physical entities actually involved in the movement, i.e. vehicles or drivers are modeled as agents. In the simulation of freight transportation, on the other hand, the agents typically represent decision makers such as customers or transport planners. Therefore, the physical representation of the agents is not of the same importance in these approaches. In freight transportation approaches several agent types are typically necessary, as opposed to traffic approaches where typically only one agent type is modeled. Trends in the approaches to passenger transportation are not as obvious. However, passengers are typically represented as agents (one exception is the taxi transportation approach), and in the bus and taxi transportation approaches, vehicles are also represented as agents. Gruer et al. (2001) represent also bus stops and road sections as agents. Thus, physical entities may be represented as agents, like in the traffic approaches. The number of different agent types is smaller than in freight transportation, but there are more agent types than in traffic approaches.

The agent behavior models are typically of a lower level (more detailed) in the traffic approaches than in the transport approaches. One reason for this difference is that in traffic approaches the personalities of the drivers often have a larger impact on the system and are therefore modeled with corresponding driving behaviors such as aggressive or calm driving style. The decision making of the agents in traffic approaches is typically based on different rules where the choices or the planning in transport approaches are made based on the best performance metrics such as cost or time. The agent behaviors sometimes also include learning aspects, which provide the agents with a dynamic behavior by the use of, e.g., NN.

In transport approaches, the focus is not on modeling the different personalities of decision makers, but rather on modeling the different types of decision-making roles and their associated rational behavior. Moreover, in (freight) transport approaches to model the negotiations and the interactions between the decision makers is crucial in order to reach a solution. In traffic approaches the interactions between the individuals are secondary, while the models of the individuals, their individual behavior, and consequently the system behavior are focused.

As far as we have seen, social simulation is not commonly applied for all modes in traffic and transportation. In the traffic domain, mainly road traffic is studied. Road traffic includes social aspects in terms of interactions between the vehicles. In air and rail traffic, the control typically takes place in a more centralized way with common objectives, which explain why these traffic modes include less social aspects in this context and therefore do not benefit as much from social simulation studies. For waterborne traffic, the infrastructure is not a scarce resource in the same sense as in the other modes; instead the bottlenecks appear in the ports or other terminals. In the freight transportation domain the social aspects mainly concern the interactions and decision making of the actors in transport chains. The most common modes that are included here are road and rail, but also waterborne transportation is sometimes included. The types of decisions that are most often studied in the freight transport domain are related to planning and mode choice, which are a consequence of the interactions between actors in transportation. In passenger transportation, planning decisions are also sometimes simulated, but sometimes operational behavior, such as loading of passengers, is also simulated which is a consequence of the interactions between passengers and vehicles.

## Further Reading

For further information about traffic simulation we refer the interested reader to (Chung and Dumont 2009), (Tapani 2008), (Toledo 2007) and (Koorey 2002). Terzi and Cavalieri (2004) provide a review of supply chain simulation, while Williams and Raha (2002) present a review of freight modeling and simulation. For general information about transport modeling, we suggest to read (Ortúzar and Willumsen 2001). For further information on how agent technologies can be used in the traffic and transport area, see (Davidsson et al. 2005).

## References

Alam SJ, Werth B (2008) Studying emergence of clusters in a bus passengers' seating preference model. Transport Res C Emerg Technol 16(5):593–614

Bazzan AL (2005) A distributed approach for coordination of traffic signal agents. Auton Agent Multi-Agent Syst 10:131–164

Ben-Akiva M, Lerman SR (1985) Discrete choice analysis: theory and application to travel demand. MIT Press, Cambridge, MA

Boxill S, Yu L (2000) An evaluation of traffic simulation models for supporting ITS development (Technical report, SWUTC/00/167602-1). Center for transportation training and research. Texas Southern University, Houston

Chen X, Zhan FB (2008) Agent-based modeling and simulation of urban evacuation: relative effectiveness of simultaneous and staged evacuation strategies. J Oper Res Soc 59(1):25–33

Chung E, Dumont A-G (2009) Transport simulation: beyond traditional approaches. EFPL Press, Lausanne

Davidsson P, Verhagen H (2013) Types of simulation. Chapter 3 in this volume

Davidsson P, Henesey L, Ramstedt L, Törnquist J, Wernstedt F (2005) An analysis of agent-based approaches to transport logistics. Transport Res C Emerg Technol 13(4):255–271

Davidsson P, Holmgren J, Persson JA, Ramstedt L (2008) Multi agent based simulation of transport chains. In: Padgham L, Parkes D, Müller J, Parsons S (eds) Proceedings of the 7th international conference on autonomous agents and multiagent systems (AAMAS 2008), 12–16 May2008, Estoril. International Foundation for Autonomous Agents and Multiagent Systems, Richland, pp 1153–1160

de Jong G, Ben-Akiva M (2007) A micro-simulation model of shipment size and transport chain choice. Transport Res B 41(9):950–965

Dia H (2002) An agent-based approach to modelling driver route choice behaviour under the influence of real-time information. Transport Res C 10(5–6):331–349

Dia H, Panwai S (2007) Modelling drivers' compliance and route choice behavior in response to travel information. Nonlinear Dynam 49:493–509

Ehlert PAM, Rothkrantz LJM (2001) Microscopic traffic simulation with reactive driving agents. In: Proceedings of intelligent transportation systems 2001. IEEE, Oakland (CA) USA, pp 860–865

El hadouaj S, Drogoul A, Espié S (2000) How to combine reactivity and anticipation: the case of conflict resolution in a simulated road traffic. In: Moss S, Davidsson P (eds) Multi-agent-based simulation: second international workshop, MABS 2000, Boston; Revised and additional papers (Lecture notes in computer science, 1979). Springer, Berlin, pp 82–96

Esser J, Schreckenberg M (1997) Microscopic simulation of urban traffic based on cellular automata. Int J Mod Phys C 8(5):1025–1036

Fischer K, Chaib-draa B, Müller JP, Pischel M, Gerber C (1999) A simulation approach based on negotiation and cooperation between agents: a case study. IEEE Trans Syst Man Cybern C Appl Rev 29(4):531–545

Gambardella LM, Rizzoli A, Funk P (2002) Agent-based planning and simulation of combined rail/road transport. Simulation 78(5):293–303

Goldberg JB (2004) Operations research models for the deployment of emergency services vehicles. EMS Manage J 1(1):20–39

Gruer P, Hilaire V, Koukam A (2001) Multi-agent approach to modeling and simulation of urban transportation systems. In: Proceedings of the 2001 I.E. international conference on systems, man, and cybernetics (SMC 2001), vol 4. Tucson, 7–10 Oct 2001, pp 2499–2504

Hemelriajk C (2013) Animal social behaviour. Chapter 22 in this volume

Hensher DA, Ton TT (2000) A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. Transport Res E 36(3):155–172

Holmgren J, Davidsson P, Ramstedt L, Persson JA (2012) TAPAS: a multi-agent-based model for simulation of transport chains. Simulation Model Pract Theory 23:1–18, Elsevier

Hunt JD, Gregor BJ (2008) Oregon generation 1 land use transport economic model treatment of commercial movements: case example. In: Hancock KL (Rap.) Freight demand modeling: tools for public-sector decision making; Summary of a conference, Washington, DC, 25–27 Sep 2006 (TRB conference proceedings, 40). Transportation research board, Washington, DC, pp 56–60

Jin X, Abdulrab H, Itmi M (2008) A multi-agent based model for urban demand-responsive passenger transport services. In: Proceedings of the international joint conference on neural networks (IJCNN 2008), Part of the IEEE world congress on computational intelligence, WCCI 2008, 1–6 June 2008, IEEE, Hong Kong, China, pp 3668–3675

Koorey G (2002) Assessment of rural road simulation modelling tools (Research report, 245). Transfund New Zealand, Wellington. http://ir.canterbury.ac.nz/bitstream/10092/1561/1/12591251_LTNZ-245-RuralRdSimulatnTools.pdf

Kumar S, Mitra S (2006) Self-organizing traffic at a malfunctioning intersection. J Artif Soc Soc Simul 9(4):3. http://jasss.soc.surrey.ac.uk/9/4/3.html

Lee S, Kim Y, Namgung M, Kim J (2005) Development of route choice behavior model using linkage of neural network and genetic algorithm with trip information. KSCE J Civil Eng vol 9, (4):321–327

Li T, Hofker F, Jansma F (2006) Passenger travel behavior model in railway network simulation. In: Perrone LF, Wieland FP, Liu J, Lawson BG, Nicol DM, Fujimoto RM (eds) Proceedings of the 2006 winter simulation conference, Baltimore

Mahmassani HS (ed) (2005) Transportation and traffic theory: flow, dynamics and human interaction. In: Proceedings of the 16th international symposium on transportation and traffic theory. Elsevier, Oxford, Great Britain

Mandiau R, Champion A, Auberlet J-M, Espié S, Kolski C (2008) Behaviour based on decision matrices for a coordination between agents in a urban traffic simulation. Appl Int 28 (2):121–138

Meignan D, Simonin O, Koukam A (2007) Simulation and evaluation of urban bus-networks using a multiagent approach. Simulat Model Pract Theory 15:659–671

Nagel K, Schreckenberg M (1992) A cellular automaton model for freeway traffic. J de Physique I 2(12):2221–2229

Ortúzar JD, Willumsen LG (2001) Modelling transport, 3rd edn. Wiley, Chichester

Panwai S, Dia H (2005) A reactive agent-based neural network car following model. In: Proceedings of the 8th international IEEE conference on intelligent transportation systems, Vienna, Austria, 13–16 Sept 2005. IEEE, Vienna, pp 375–380

Pelechano N, O'Brien K, Silverman B, Badler N (2005) Crowd simulation incorporating agent psychological models, roles and communication (Technical report). Center for Human Modeling and Simulation, University of Pennsylvania

Pursula M (1999) Simulation of traffic systems – an overview. J Geogr Inf Decis Anal 3(1):1–8

Ramstedt L (2008) Transport policy analysis using multi-agent-based simulation. Doctoral dissertation No 2008:09, School of Engineering, Blekinge Institute of Technology

Rindsfüser G, Klügl F (2007) Large-scale agent-based pedestrian simulation. In: Petta P et al (eds) MATES 2007, LNAI 4687, Leipzig, pp 145–156

Rossetti R, Bampi S, Liu R, Van Vleit D, Cybis H (2000) An agent-based framework for the assessment of driver decision-making. In: Proceedings of the 2000 I.E. intelligent transportation systems, IEEE, Dearborn, (MI) USA, IEEE, 387–392

Santos G, Aguirre BE (2004) A critical review of emergency evacuation simulation models. In: Peacock RD, Kuligowski ED (eds) Proceedings of the workshop on building occupant movement during fire emergencies. National Institute of Standards and Technology, Gaithersburg, 10–11 June 2004, pp 27–52

Schreckenberg M, Selten S (2004) Human behaviour and traffic networks. Springer, Berlin

Strader TJ, Lin FR , Shaw M (1998) Simulation of order fulfillment in divergent assembly supply chains. J Artif Soc Soc Simul 1(2). http://jasss.soc.surrey.ac.uk/1/2/5.html

Swaminathan J, Smith S, Sadeh N (1998) Modeling supply chain dynamics: a multiagent approach. Decis Sci J 29(3):607–632

Tapani A (2008) Traffic simulation modeling of rural roads and driver assistance systems. Doctoral thesis, Department of science and technology, Linköping University, Linköping

Terzi S, Cavalieri S (2004) Simulation in the supply chain context: a survey. Comput Ind 53(1):3–17

Toledo T (2007) Driving behaviour: models and challenges. Transport Rev 27(1):65–84

van der Zee DJ, van der Vorst JGAJ (2005) A modelling framework for supply chain simulation: opportunities for improved decision making. Decis Sci 36(1):65–95

Wahle J, Schreckenberg M (2001) A multi-agent system for on-line simulations based on real-world traffic data. In: Proceedings of the 34th annual Hawaii international conference on system sciences (HICSS'01). Hawaii, USA, IEEE

Williams I, Raha N (2002) Review of freight modelling. Final report, DfT Integrated Transport and Economics Appraisal, Cambridge

Zhang Q, Han B, Li D (2008) Modeling and simulation of passenger alighting and boarding movement in Beijing metro stations. Transport Res C 16:635–649

# Chapter 25
# Modeling Power and Authority: An Emergentist View from Afghanistan

**Armando Geller and Scott Moss**

**Why Read This Chapter?** To understand how an evidence-driven approach using agent-based social simulation can incorporate qualitative data, and the effects of social complexity, to capture some of the workings of power and authority, even in the absence of sufficient statistical data. This is illustrated with a model of Afghan power structures, which shows how a data collection process, intuitive behavioral models and epistemological considerations can be usefully combined. It shows how, even with a situation as complex as that of Afghanistan, the object under investigation can shape the theoretical and methodological approach rather that the other way around.

**Abstract** The aim of this chapter is to provide a critical overview of state of the art models that deal with power and authority and to present an alternative research design. The chapter is motivated by the fact that research on power and authority is confined by a general lack of statistical data. However, the literal complexity of structures and mechanisms of power and authority requires a formalized and dynamic approach of analysis if more than a narrative understanding of the object of investigation is sought. It is demonstrated that evidence-driven and agent-based social simulation (EDABSS) can contend with the inclusion of qualitative data and the effects of social complexity at the same time. A model on Afghan power structures exemplifying this approach is introduced and discussed in detail from the data collection process and the creation of a higher order intuitive model to the derivation of the agent rules and the model's computational implementation. EDABSS not only deals in a very direct way with social reality but also produces

A. Geller (✉)
Scensei, LLC, 107 S. West St., Alexandria, VA 22314, USA

The School for Conflict Analysis and Resolution, George Mason University, 3330 N. Washington Blvd., Arlington, VA 22201, USA
e-mail: armando@scensei.com; ageller1@gmu.edu

S. Moss
Scott Moss Associates, Brookfold, The Wash, Chapel-en-le-Frith, SK23 0QW, UK
e-mail: scott@scott.moss.name

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_25,

complex artificial representations of this reality. Explicit socio-cultural and epistemological couching of an EDABSS model is therefore essential and treated as well.

## 25.1 Introduction

Notions such as "power" and "authority" are redolent with meaning yet hard to define. As a result, the number of definitions of power and authority is overwhelming (Neumann 1950). Weber (1980:53) defines power as the "probability that one actor within a social relationship will be in a position to carry out his own will in spite of resistance." Giddens (1976:111) understands power in a relational sense as a "property of interaction, and may be defined as the capability to secure outcomes where the realization of these outcomes depends upon the agency of others." He implies a major distinction between two types of resources in connection with power, (1) control over material resources and (2) authoritative resources. Parsons (1952:121–132) underlines the pragmatic character of the notion of power even more by stating that power is the capacity to achieve social and societal objectives, and as such can be seen as analogous to money. Power, consequently, is the basis of a generalized capacity to attain goals and to invoke consequences upon another actor (Moss 1981:163).

Neither these nor any other definitions predominate and the decision to apply a particular one is subjective and, if not based on normative grounds, most likely context-dependent. Hence, in this research it is not aimed at applying one particular theoretical approach to power and authority, but instead it is argued that such an approach can also be founded on available and contextual evidence. Evidence is understood as information that is derived from case-studies, empirically tested theories, the high quality media and engagement with stakeholders, domain experts and policy analysts and makers.

For heuristic reasons – and in awareness of the plethora of conceptual approaches to power – it is for the time being assumed that the social phenomena of power and authority occur in a two- (or more-) sided relationship. It is also assumed that power serves a purpose. What should be of interest to students of power has been identified by Lasswell (1936): "Who gets what, when, how". Who and what describe and explain structures; when and how describe and explain mechanisms[1] and processes. However, Lasswell ignores an important aspect: "why". Why does someone get something at a particular moment in time in a particular way? And more generally: Why did a particular condition of power form?

[1] Schelling (1998) understands a "social mechanism [...][as] a plausible hypotheses, or set of plausible hypotheses, that could be the explanation of some social phenomenon, the explanation being in terms of interactions between individuals and other individuals, or between individuals and some social aggregate." Alternatively, a social mechanism is an interpretation, in terms of individual behavior, of a model that abstractly reproduces the phenomenon that needs explaining.

Castelfranchi (1990) already noted in the year 1990 that social power is a lacuna in social simulation and (distributed) artificial intelligence. This chapter shows that although power relations are ubiquitous in social systems, only a small number of relevant models have been developed. This is despite the fact that social simulation and in particular evidence-driven and agent-based social simulation (EDABBS) are valuable complementary techniques to orthodox approaches to the study of power and authority.

A prime virtue of EDABSS is that it imposes on the modeler a requirement to devise an unambiguous formal meaning for such notions as power and authority and their corresponding concepts. The modeling process and thus formalization should begin by formulating questions that structure the rich body of narratives that are available to describe the *explanandum*:

- Under what conditions would you label someone as powerful or as being in a position of authority?
- Having labeled someone as being powerful or in a position of authority, how would you expect that person to behave?
- Having labeled someone as powerful or in a position of authority, how would you expect yourself/others to behave towards that person?

These questions are not abstract. They form part of a data collection strategy and aim at enriching general accounts of power and authority, such as that power is a form of relationship that is exercised between two or more individuals, by more specific and context-dependent forms of descriptions of power and authority. Often these descriptions concern actor behavior. Models based on such descriptions can be closer to the evidence, because this evidence can be straightforwardly translated into “if-then” rules that can be implemented in logic-like fashion in rule-based or standard procedural languages using their native if-then structures. The if part is determined by the first question and the then part by instances of the second or third questions. Our own preference is for an evidence-driven and declarative representation of power and authority in order to preserve as much as possible of the rich data drawn from case-studies and evidence in general while maintaining conceptual clarity.

The computational modeling procedure described in this article is inspired by the idea to represent reality by means of modeling; it is driven by shortcomings and advantages of other methodological approaches; and it has matured out of research on cognitive decision making as declarative and thus mnemonic implementations (Moss 1981, 1998, 2000; Moss and Edmonds 1997; Moss and Kuznetsova 1996; Moss et al. 1996) and on contemporary conflicts (Geller 2006b). Classical hermeneutic approaches, although they may be strong in argument, are often methodologically weak. However, they have an important “serendipity” function and creative role, very much like that of intuitive models (Outhwaite 1987). Traditional empirical approaches, such as statistical and econometric modeling, do not represent reality and have difficulties to produce insight into mechanisms, processes and structures (cf. Shapiro 2005; Hedström and Swedberg 1998).[2] Moreover, regular incorporation of poor proxies and the use of inadequate data does not contribute to the plausibility of research results. However, rigorous formalization furnishes desirable clarity and comparability. Finally, qualitative and case-study-based analysis produces deep structural and processual insight as well as "thick description", although at a high – for some too high – price of idiography and thus lack of generalizability.

None of these approaches is incompatible with EDABSS. But we argue here that a natural modeling procedure starts with an informed but intuitive and theoretical model that needs to be validated against reality. The intuitive model is to be validated at micro-level against largely qualitative evidence, so that a rich model develops. This qualitatively validated and enriched model is then formalized as a computational model. The social simulation's output should enhance our understanding of reality, which, in turn, necessitates adjustments in the model design. This process is the hermeneutic cycle of EDABSS (cf. Geller 2006b; Geller and Moss 2008b).

A selection of models dealing with power and authority is reviewed in Sect. 25.2. A critical appraisal of these models reveals the strengths and weaknesses of past modeling approaches and underlines the necessity of the research design presented here. The selection criteria are evidently influenced by our own modeling approach and are therefore subjective. However, models have also been chosen on a functional basis: How successful are they in describing and explaining their target system? We have also tried to choose exemplary models from a wide range of modeling philosophies, to elucidate conceptual differences amongst the discussed models. Section 25.3 comprises the dialectical results of Sect. 25.2 and discusses the analytical concepts and methodological tools applied here to analyze power and authority. The materialization of our approach is presented in Sect. 25.4, where the model implementation and simulated output is discussed. The model's context is conflict-torn Afghanistan. Section 25.5 concludes by embedding our modeling framework into a broader context of comparable social phenomena, such as conflict and organized crime, and promotes agent- and evidence-based social simulation as an efficient approach for the study of power and authority.

## 25.2 What Can We Learn from (a Selection) of Models on Power and Authority?

The development of a social simulation model can be informed by intuitive ideas, theory or observation. For many simulations, the respective borderlines cannot be drawn unambiguously. Nevertheless, such a classification is superior to a more traditional one which only distinguishes between micro- and macro-models.

[2] See for a promising corrective Sambanis (2004).

Although agent-based models entail explicit micro level foundations, for example, the micro foundations of econometric models are at best implicit. More importantly in relation to complexity, agent-based simulations often generate emergent phenomena at macro level. From a modeling point of view it is therefore more interesting to understand the level and nature of data that has guided the researcher to conceptualize a model in a particular way, to what model output this conceptualization has led and how a design helped to better understand mechanisms, processes and structures in a target system.

### *25.2.1 Modeling Ideas*

A variety of models on power and authority are implemented not strictly based upon theory but rather on a mixture of intuition and existing theoretical research in a particular field. These models promise to lend insight into a usually only little defined social phenomenon in an explorative, and likely to be abstract, way and therefore operate as an *entrée* into the object of investigation's broader field. Robert Axelrod's emerging political actor model has been chosen because it epitomizes the prototype of an explorative model; Rouchier et al.'s model still exemplifies the want to explore, however on a more evidence-oriented basis.

#### 25.2.1.1 Emerging Political Actors

In a well known agent-based model Axelrod (1995) reasons about the emergence of new political actors from an aggregation of smaller political actors. His motivation is to explain the restructuring of the global political landscape after the end of the cold war and the fact that, although political scientists have a number of concepts and theories to analyze the emergence of new political actors, they lack models that account for this emergence endogenously.

In short, Axelrod's model of emerging actors is a well structured and intelligible but empirically ungrounded, conceptual model. The core of his model is a simple dynamic of "pay or else" resulting in a "tribute system in which an actor can extract resources from others through tribute payments, and use these resources to extract still more resources." (Axelrod 1995:21) The model consists of ten agents distributed on a circle. Wealth is the only resource and is distributed to each agent randomly at the beginning of the simulation. In each simulation iteration three agents are randomly chosen to become active. When active, agents can ask other agents for tribute. When asked, an agent has the choice between paying the demanded tribute or to fight, depending on his and the demander's resources. In case of paying the tribute, a specified amount of wealth is transferred to the demander; in case of fighting, each agent loses wealth equal to 25 % of his opponents' wealth. After three iterations all agents exogenously receive again wealth.

The core of any agent-based model is the rules according to which the agents behave. In Axelrod's tribute model the agents have to decide when to demand tribute and how to respond to such a demand. First, an active agent needs to decide whom it should address. "A suitable decision rule [. . .] is to choose among the potential targets the one that maximizes the product of the target's vulnerability multiplied by its possible payment" (Axelrod 1995:24). If no agent is vulnerable enough, then no demand is made. The chosen agent responds by fighting only if *t* would cost less than paying the tribute.

Agents can form alliances. During the course of the simulation, agents develop commitments towards each other. Commitment between two agents increases if one agent pays tribute to the other agent and vice versa and if two agents fight another agent together. Commitment decreases if one agent fights at the opposite side of another agent. Alliance building indirectly increases an agent's resources and outreach, as it can only make a demand to an agent if it is either adjacent or indirectly connected via allied agents.

Axelrod gets six characteristic results: (1) The model does not converge to an equilibrium. (2) The model's history is fairly variable in terms of combinations of wealthy actors, fighting frequency and overall wealth accumulation. (3) Agents do not only share the resources with their allies, they also share the costs and thus the risks. Major conflict can occur as agents can be dragged into fighting. (4) Because of asymmetric commitment, fighting can erupt among allies. (5) An alliance can host more than one powerful agent. (6) The initial wealth distribution is not an indicator for an agent's success. As an overall result, Axelrod reports that he was able to breed new political actors of a higher organizational level. These new political actors are represented in the model as so called clusters of commitment.

Axelrod's model demonstrates the realist notion that states do not have friends, but only interests. (Agents initially make commitments out of rational calculations – "pay or else" –, not out of ideological considerations.) These interests, the model reveals, are attended most efficiently by joining coalitions. Thus, an effective powerful agent seeks cooperation of some form or the other.

Axelrod's model is convincing as long as he concentrates on his main task – letting new political agents emerge in the context of an explorative setup. But interpreting a positive feedback effect resulting from a total of ten agents as "imperial overstretch", or interpreting fighting between two agents of the same cluster of commitment as civil war rings a little hollow. As much as we can learn from Axelrod about how to setup and present a simple, but innovative and explorative model, as little can we learn about how to discuss it, as Axelrod continually blurs the distinction between the model and reality which leads to over- and misinterpretation of his results. Hence, a number of open questions remain, the most important of which is whether the model would withstand even circumstantially founded cross-validation. This is naturally related to the question of how much explanatory power Axelrod's model holds. With agents so abstract from reality, agent behavior in Axelrod's model simply cannot provide any explanatory insight into real world actor behavior. And while it makes sense to claim that complex macro level behavior can emerge from simple micro level behavior – as is the case in Axelrod's model – this macro outcome, again, is so abstract from reality that it does not allow for any insight into the real international system. Consequently, the emergent interplay between the micro- and the macro-level, so typical for complex systems, cannot have the effects it would have in reality. This accounts for another loss of explanatory power. Axelrod claims that his model's objective is to think in new ways about old problems. However, its lack of evidence-based foundations and its degree of abstraction from reality might well foster stereotyped perceptions of the international system instead of triggering critical reflection about it.

An alternative is for agents and their environment to be derived either from an empirically well-tested theory or from qualitative real-world observations. Cederman (1997), discussed below, has advanced Axelrod's work into this direction.

#### 25.2.1.2 An Artificial Gift-Giving Society

Rouchier et al. (2001) reported a model of an artificial gift-giving society that is loosely founded on ethnographic research. Gifts structure society and reproduce habits and values. The giving of gifts can also be a means of redistribution. The donation, reception and reciprocation of a gift create relationships, which can become competitive. Thus, gifts can be a means to create authority, establish hierarchies and uphold power structures.

The model's goal is to create an artificial society in which reputation emerges. The gift-giving society's agents are either occupied with working to accumulate resources, which enables them to give away gifts or by giving away gifts themselves. The motivation to give away gifts is twofold: On one hand gifts are given away because agents act according to their self-esteem, i.e. the desire to be able to make gifts that are acceptable to the group. On the other hand, agents give away gifts because they want to increase their reputation by swaggering. Agents are fully informed and share the same decision process to determine what actions to take. The artificial society's population consists of 50 agents.

Each agent has to decide at each time step what gift it wants to make to whom. This decision process stands at the core of the model. Agents can give away gifts, either for the sake of reputation or sharing. So called "sharing gifts" are less costly than prestige gifts. The decision of whom to give what gift is socially embedded and the agent's rationality depends on its social position. Making a sharing gift to any agent adds to the social inclusion of an agent; making a prestige gift to an agent who is considered as being prestigious fosters a hierarchy among the agents. Therefore, receiving a sharing gift represents social acceptance; receiving a prestige gift represents the acceptance of social stratification. The better an agent is socially integrated, i.e. the higher its self-esteem, the higher is its motivation to give away gifts. At the same time, high social integration increases the likelihood that an agent is able to give away prestige gifts.

All agents exchange their gifts after each agent has decided to whom to give its gift. Subsequently the agents evaluate their ranks within the group and their reputation on the basis of the gifts they have received. Self-esteem and reputation are adapted according to the outcome of the gift-giving round. Then the next round starts.

The authors discuss their findings with regard to donation-reception dynamics evolving during the simulation, of which most are positive feedback loops (Rouchier et al. 2001:5.1). An agent's esteem increases with the number of gifts it has made and received respectively. The higher this agent's esteem is, the higher the likelihood is that it receives even more gifts and that he can give gifts away. The same holds true with an agent's reputation, which increases with the reception of prestige gifts. The higher this agent's reputation is, the higher is the likelihood that it receives even more prestige gifts. Moreover, the authors find that esteem and prestige "go together". "Both help the circulation of gifts, and then the creation of prestige reputation." (Rouchier et al. 2001:5.3) The gift-giving model provides insight into the emergence of an elite group based on socially contextualized decision processes. Within the gift-giving model the emergence of social power is explained by two factors: (1) the willingness and ability to become socially accepted and (2) the ambition to accumulate reputation. Gift-giving has been introduced and computationally implemented by Rouchier et al. (2001) as a process of resource accumulation and redistribution, an important variant of strategic behavior in many other contexts.

The authors, while creating a naturalistic model, do not attempt to derive their agent rules directly from qualitative, in this case likely anthropological research. This would have allowed them to gain narrative insight into the dynamics of gift giving and receiving and would have enabled them to directly cross-validate their findings with accounts made by anthropologists. Micro-level explanation was not one of the modelers' main interests. Since the emergence of complex macro level outcomes results from micro-level behavior and social interaction, the model is of limited usefulness in the analysis of real world social complexity. Formulation of agent rules from anthropological evidence would also have enabled the authors to avoid the assumption that agents are totally informed, both, spatially and with regard to the internal state of other agents. The cognitive processing of information about other agents is an especially difficult task. For example, the model does not address the question of when, as well as to whom, agents give gifts. It was not a purpose of the Rouchier et al. (2001) model to capture strategic decision-making using incomplete information within a realistic context. To have done so would usefully have entailed reliance on more evidence-based information available in the anthropological literature and might then have provided an empirically better grounded account of social status and power.

### 25.2.2 *Testing Theory*

#### 25.2.2.1 Power Games Are Not Games

Game theoretical applications comprise perhaps the most formalized, coherent and theoretically unified approach to power, especially in international relations.[3] Regularly applied in military-political analysis and international political economy, game theory has a straightforward conception of the nation state as interdependent, goal seeking and rational actor embedded in a context free of centralized, authoritative institutions (Snidal 1985). The application of game theoretical models in international relations raises questions such as "Who are the relevant actors?", "What are the rules of the game?", "What are the choices available to each actor?", "What are the payoffs in the game?" and "Is the issue best characterized as single-play or repeated-play?" (Snidal 1985:26). Abstract and simplified as they are, game-theoretical models nevertheless intend replicating a particular social situation and aim at – if the actors' preferences, strategies and payoffs are accurately modeled – generating testable predictions and understanding of fundamental processes governing the system of international relations.

If states are conceived as rational power maximizers in an anarchic system, then we talk about the realist paradigm of international politics. "Rationality in this Realist world centers on the struggle for power in an anarchic environment" (Snidal 1985:39). However, compared with reality it would be misleading to conceive of states as self-defending entities of purely opposing interests. Rather, game theory teaches us that states exhibit a strategic rationality that incorporates the awareness that no state can choose its best strategy or attain its best outcome independently of choices made by others. This awareness is the birth of cooperation. Moreover, from so called iterated and dynamic games we learn that while states have incentives to defect in the short-run, in the longer-run they can achieve benefits from cooperation through time (Snidal 1985).

Game theory has often been criticized on the grounds that it is unrealistic. Game theory can, of course, analyze very particular examples of world politics, such as the Cuban Missile Crisis, nuclear war in a bipolar world or the General Agreement on Trade and Tariffs (GATT). What can be expected, in more general terms, are explanations for meta-phenomena or for abstract representations of particular circumstances. We criticize game theoretical approaches in the social sciences from an evidence-based point of view. The difference between a game theoretic model and the target system is enormous. Everything that is a natural representation

[3] Models not discussed in this subsection but of further interest to the reader are Alam et al. (2005), Caldas and Coelho (1999), Guyot et al. (2006), Lustick (2000), Mosler (2006), Rouchier and Thoyer (2006), Saam and Harrer (1999) and Younger (2005). Particularly highlighted should be the work of (Mailliard and Sibertin-Blanc 2010) who merge a multiagent and social network approach to the complexity and transactional nature of power with approaches to power from the French school of sociology, and develop against this background a formal logic system.

of the international system is a special and difficult case in game theory, for example, non-symmetric $n$-person games involving intermediate numbers of states, and everything that is straightforwardly implementable in game theory is unrealistic. Yet analogies are quickly drawn between model results and the model's target system. Hence, findings rely on oversimplified model ontologies, which may lead to over-interpretation.

Game theory's simplistic ontologies also affect a model's simulation output, as Moss (2001) argues. An analysis of state of the art game theory as represented by 14 game theoretic papers published in the Journal of Economic Theory in the year 1999 indicates that the game theoretic models' assumptions (perfect information) and implementations (e.g., Markov transition matrices and processes) preclude the emergence of self-organized criticality as reported by Bak (1997) and cannot capture the necessary interaction as a dynamic process. Game theoretic models on markets do not entail statistical signatures found in empirical data on markets, such as power law distributed data, a characteristic of self-organized criticality. This critique applies to game theoretic models in international relations as well, as one of the observed regularities in the international system is the power law distribution of wars (Richardson 1948). Whereas Cederman (2003) replicated Richardson's (1948) findings by means of agent-based modeling, to our knowledge there exists no game theoretical reproduction of this statistical signature.

Although game theory can make statements of the "when" and "why", it cannot say anything about the "how" and cannot give insight into the mechanisms and processes underlying the emergence of power structures [54]. Therefore, game theoretical modeling is, like most statistical or econometric modeling, a type of black box modeling. A rare exception to this is Hirshleifer's work (Hirshleifer 1991, 1995).

#### 25.2.2.2 Exploring the Limits of Equation-Based Modeling

Hirshleifer (1991) published a seminal paper on the paradox of power. The paradox of power states that in case of conflict, a weaker contestant is able to ameliorate his position relative to the stronger actor because his inferiority makes him fight harder. In other words: "[N]on-conflictual or cooperative strategies tend to be relatively more rewarding for the better-endowed side." (Hirshleifer 1991:178) Hirshleifer's modeling solution for this problem is based on the assumption that if there exists an equilibrium market economy, then there must also exist an equilibrium outcome if contestants in two-party interactions compete by struggle and organized violence. The model's assumptions are full information, a steady state, indifference towards geographical factors and the non-destructive nature of fighting (Hirshleifer 1991:198).

Hirshleifer's econometric model leads to a well-specified outcome from which a number of unequivocal results can be derived. When the paradox of power applies the "rich end up transferring income to the poor" and this "tends to bring about a more equal distribution of [. . .] income." (Hirshleifer 1991:197) However, he also states that "the comparative advantage of the poor in conflictual processes can be overcome when the decisiveness of conflict is sufficiently great that is, when a given ratio of fighting effort is very disproportionately effective in determining the outcome of conflict." (Hirshleifer 1991:197) Hirshleifer validates his analytical results by providing circumstantial evidence for a number of examples from class-struggle within nation-states to firms (labor-management conflicts) and protracted low-level combat.

If one accepts economic theory's underlying assumptions, Hirshleifer's results are indeed compelling and apply to a wide range of social issues related to power. The paradox of power identifies, perhaps correctly, the expectations weak and strong actors can have when mutually entering power politics: When the stakes are high, the weak are likely to get crushed by the strong. This is why the paradox of power is likely to apply to more limited contests than to full fledged conflict.

What can we learn from Hirshleifer with regard to modeling power and authority? He elegantly exemplifies state of the art model conceptualization, presentation and discussion of results, including the model's limitations. His presentation of the analytical tools and his disclosure of the model's underlying assumptions are exemplary and render his work amenable to critique. The same applies to the rigid and straightforward formalizations of the mechanisms underlying the paradox of power, which are, moreover, well-annotated and embedded in theory. Last, but not least, the model's scope and its delimitations are clearly marked by referring to a number of examples ranging from interstate war to the sociology of the family.

Hirshleifer informs us precisely of the outcomes from power struggles and of the factors that cause these outcomes; he fails to produce analytical insight into the structural arrangements and processes that lead to these outcomes as a consequence of the methodology itself as is demonstrated in another paper of his.

In "Anarchy and Its Breakdown" Hirshleifer (1995) models anarchy, a social state lacking *auctoritas*, as a fragile spontaneous order that may either dissolve into formless amorphy or a more organized system such as hierarchy. Interesting for our task is the fact that Hirshleifer produces results that indeed allow insight in terms of structure. He can for example demonstrate that the state of anarchy is stable if no actor is militarily dominant and if income is high enough to assure one's own and the group's survival. However, concrete processual insight can again not be delivered and it is referred to circumstantial evidence to concretize particular aspects of the model. In fact circumstantial evidence is arguably the only validation that is feasible with these kinds of very abstract models as the statistical signature of the model's output is so ideal-typical that it is not comparable with empirical data. In short, orthodox theoretically informed models, such as econometric or statistical models, often do address those issues in which social scientists are really interested, but cannot provide an explanation in involving complexity arising from social interaction.

One reason for this has already been identified above with relation to game theory, i.e. the preclusion of emergence of self-organized criticality. Another reason is highly unrealistic general assumptions, i.e. rational choice, and more specific unrealistic assumptions, for example the non-destructive nature of fighting (as such also identified by Hirshleifer (1991:196–199) himself) or monolithic actors lacking an internal state. Finally, methodological individualism ignores the micro–macro link as well as the heterogeneous nature of political actors. While the particular assumptions as well as the (homogeneous) agents could be chosen more realistically, methodological individualism is together with most statistical and econometric modeling relying on homoskedasticity of data points and variances. Such models cannot be used to describe or explain the evolution of power structures. (For statistical models this holds true only, of course, if there is sufficient statistical data that describes power relations.) With this regard, Cederman's (1997) model is a paradigmatic shift.

#### 25.2.2.3 When, Why and How Nations Emerge and Dissolve

Cederman's (1997) model has been chosen, because it applies the analytical rigor of formalized approaches to agent-based modeling without relinquishing the latter's virtues. He raises fundamental questions regarding realist and methodologically orthodox approaches to international relations. He introduces an agent-based simulation of the emergence and dissolution of nations. His model is based on a $20 \times 20$ grid that is initially inhabited by "predator" and "status quo" states. Each period of time a state is randomly assigned to receive resources for attack and defense. Given they have an advantage in overall and local resources respectively, predator states can attack and conquer neighboring states. Although Cederman applies neorealist principles to a dynamic and historical meta-context, his findings challenge the orthodox neorealist belief that applying alternative methods does make an epistemological difference. He reports that defense alliances and the dominance of defense in the international system are paradoxically destabilizing. While defensive systems of cooperation deter predator states from growing to predator empires, at the same time they make possible the build-up of a hegemonic predator actor, because once a predator has reached a critical mass it has, due to the defensive nature of the system, no rivals.

Cederman (1997:136) asserts that "[s]tates have been mostly modeled as internally consolidated monoliths, albeit with emergent outer boundaries." He confronts this simplifying assumption by supplying his agents with an internal decision making mechanism representing a nationalistic two-level-politics mechanism. State leaders cannot be concerned only with foreign affairs anymore, but must also take into consideration domestic issues (cf. Putnam 1988). Lazer (2001) has stated that the insight gained from Cederman's nationalist implementation is not as striking as the one dating from his emergent polarity implementation. Perhaps this is true in terms of contents, but in an epistemological perspective Cederman makes an important point: His nationalist model inspirits previously dead states. It is not enough to know that states do something, but from a social scientific point of view it is essential to know why they do it and how. This affords realistic, i.e. evidence-based assumptions and implementations.

Although Cederman (2003) later on introduces technological change and power projection and Weidmann and Cederman (2005) introduce strategy and self-evaluation into the decision making process, Cederman's models conform to traditional empirical perceptions of international relations. Consequentially, a number of issues relevant to the study of power and authority in contemporary conflicts remain untouched. The state, territorial conquest and consequentially the redrawing of borderlines have been important explanatory factors for conflicts since the emergence of territorial states, but they misconceive the nature of a great number of conflicts throughout history and consequentially can only partially explain the emergence of power structures in contexts where the nation state or any other type of centralized political power has only played a marginal role. And even where such well defined territory existed, Cederman cannot explain the causes for conflict if they have been for example of ethnic, religious or economic character. Other issues that should be taken into consideration are neo-patrimonialism, anomic states, genocide, transnational organized crime and external intervention. Models analyzing such a reality must be multivariate and causal, allowing for an explorative framework, and – contrary to Cederman (1997) – be able to include atheoretical and evidence-based information, which is often of a qualitative type.

### *25.2.3 Toward Implementing Reality*

Modeling reality is not just about modeling particular cases. Modeling reality is about the development of models that have explanatory power on both, the micro- and macro-level and therefore give also insight into mechanisms, processes and structures. A model can hardly claim to exhibit explanatory power when lacking pivotal aspects of a perceived reality and when abstracting too much from this reality. Every reasonable model of reality – i.e. a model that describes not only the who, what and when, but also the why and how – must entail construct validity.[4]

#### 25.2.3.1 Explaining the Causes for Social Unrest and Organized Violence

Kuznar and Frederick (2007) propose a model in which they explore the impact of nepotism on social status and thus power. They rely on an innovative model architecture supported by relevant research results, which are not framed by dogmatic theory. An agent-based model is employed "to model the origins of the sort of wealth and status distributions that seem to engender political violence. In particular, we explore the minimum conditions required to evolve a wealth

[4] A model not discussed in this section but that is of excellent quality, both in terms of content and innovation is (Guyot et al. 2006). The authors analyze and discuss the evolution of power relations on the basis of participatory simulations of negotiation for common pool resources.

distribution from the mild inequalities seen in ancestral human societies of hunter-gatherers to the more extreme wealth inequalities typical of complex, stratified societies." (Kuznar and Frederick 2007:31)

Wealth is implemented as a variable that takes an ideal cultural form over which actors would compete in a particular society. For hunter-gatherer societies wealth is distributed in a sigmoid fashion, where agents in the convex parts of the distribution have more to gain than to lose when taking chances and therefore are risk prone. The model consists of three building blocks: the distribution of wealth, agent interaction and nepotism. Wealth distributions in complex societies are, by contrast, exponential, with sigmoid oscillations around the exponential curve. Kuznar and Frederick (2007:32) term this an expo-sigmoid curve. Agent behavior is modeled along the lines of a coordination game with two equilibria (either both players defect, or both players cooperate) and a Nash optimum which is to play a mixed strategy of join and defect. Kinship is inherent to nepotism. Thus, agents with many kin and offspring perform better, i.e. have higher payoffs, in the coordination game and exhibit higher fertility due to the effect of cultural success.

The result Kuznar and Frederick are getting is that the effects of nepotism transform a sigmoid wealth distribution into an approximate expo-sigmoid distribution. The explanation for this is the emergence of a distinct three class society: the poor and their offspring get poorer, a small middle class gets richer without changing status and the elites get richer and richer. Thus, the authors conclude the positive feedback loop working between nepotism and cultural success increases the structural inequality between a powerful elite and the rest of the population and thus escalates social unrest and potential organized violence.

Whereas Axelrod (1995) over-interprets his model, Kuznar and Frederick explore the full potential of their research design: A simple, intuitive and at the same time thoroughly grounded model is presented well, specified together with moderate but auspicious conclusions, which advance research and shed new light on a problem. The model, however, would be even more compelling if the authors would have had presented cross-validational results.

Model output should be, if possible, cross-validated against empirical data originating from the target system (Moss and Edmonds 2005). There are three strategies: (1) If the simulation leads to statistical output, this output is statistically analyzed and the resulting significant signatures are compared with the statistical signatures gained from data originating from the target system. Such signatures can for example be a leptokurtic data distribution, clustered volatility, power laws or distinct (e.g., small world) patterns in a social network. If the model yields output of statistical nature but statistical data is not available for the target system, then validation must rely on qualitative data. In this case, validation must either (2) seek systematic structural and processual similarities between the model and the target system, for example cross-network-analysis, or (3) find circumstantial evidence in the target system that can also be found in the simulation. In case of empirical data scarcity (3) is often the last resort.

#### 25.2.3.2 Power, Resources, and Violence

Geller (2006a,b) developed an agent-based model of contemporary conflict informed by evidence. The lack of a unified theory of contemporary conflict motivated an intuitive and explanatory model of contemporary conflict. This model is based on three types of interacting actors: a politician, businessman and warrior. They engage in six interactions: (1) the politicization of the economy and (2) the military, (3) the economization of politics and (4) the military, (5) the militarization of politics and (6) the economy. To ascertain if this intuitive and simple ontology can capture the main structural and processual characteristics of contemporary conflicts, ten cases, such as Afghanistan, Chechnya and Sierra Leone, have been analyzed in a primary validation procedure against the backdrop of the intuitive model. The analytical results in the form of mini case-studies based on secondary literature have been sent out to subject matter experts for a critical review. None of the experts requested an essential revision of the analytical tool, the intuitive model.

As a next step, Geller enriched the theoretical background of the primarily validated intuitive model by consulting more relevant literature for further specification of the model's structure and agency aspects in order to be able to model the computational model's agent rules. The basic idea is that politicians affiliate with businessmen and warriors to make good business and get protection while businessmen affiliate with politicians to get access to the lucrative political arena and warriors seek political representation by politicians. Businessmen affiliate with warriors for the same reason politicians do, to get protection, while warriors get money for their provided services. Warriors can kill warriors affiliated with other politicians or businessmen, whereas civilians are considered as being non-constitutive to the intuitive model, as they are introduced into the computational model in a reactive way, meaning that although they can affiliate themselves to politicians, they can be forcibly recruited and ultimately killed by warriors.

The model offers insight into the dynamics of power in contemporary conflicts. Contrary to the prevailing "greed and grievance" approach in current conflict studies, Geller's model demonstrates that a powerful agent is dependent on businessmen and warriors at the same time. Powerful is who is most socially embedded. He can also show that the main organizers of violence are the politicians and that the warriors need not exhibit enough organizational capacity for a fully fledged campaign of organized violence. Hence, the more fragmented the political landscape is, the greater is the magnitude of organized violence. Geller's results gain importance as they are cross-validated against statistical data describing the number of conflict related victims on a daily basis in Northern Ireland, Iraq and Afghanistan. Both the simulation output and the real world data suggest that conflict related victims are log-normally distributed over time (right-skewed), exhibiting outbreaks of violence unpredictable in magnitude and timing.

Modeling always involves a degree of arbitrariness. A modeler's task, then, should be to reduce arbitrariness by making the model's design as inter-subjectively comprehensible as possible. Axelrod's emerging actors model is a good, but nevertheless simple, example of this. The more evidence oriented a model becomes, the more difficult it becomes to justify the various omissions, inclusions and abstractions. Procedural programming is cumbersome in responding to idiographic challenges. As described in the next section, a declarative, rule-based approach is better suited to the translation of evidence-based information of actor behavior into agent rules.

### *25.2.4 Discussion*

The synopsis presented above has revealed the many approaches through which the social phenomena of power and authority can be scrutinized: in models based on ideas interest oriented states and gift giving individuals have been implemented; in highly formalized theoretical models agents are conceived as rationalist power maximizers or as neo-realist states internally and externally struggling for survival; detailed evidence collected from secondary literature is used for modeling processes and structures that entail a high construct validity. Power structures are complex as well as dynamic and emerge as a result of a multitude of structure generating interactions among self- and group-oriented actors encompassing manifold interests.

The discussed models allow for insight into dynamic model behavior, as well as drawing structural conclusions with regard to their object of investigation whether it is theoretical or empirical by nature. Those models that feature an individual agent architecture also lend insight into aspects of agency. Only these cope with our stipulation that the *explanandi* of social simulations of power and authority must deal with structure and agency. Nevertheless, in most cases the agent rules have been implemented on a basis, which is under-representing evidence, bringing about the problem that structural emergence cannot be related clearly (i.e. intra-contextually) to agent behavior. Consequently, the analysis of agent behavior cannot be related to actor behavior in the target system. This lack of realism in model design renders validation attempts of simulation results against reality less plausible and informative. By contrast, homologue models of the type advocated for in this chapter enable the researcher to gain insight into structure and agency of the sort that is more directly linkable to actor behavior. As a result, validation results become more plausible and research can enter the hermeneutic cycle of EDABSS.

## 25.3 Evidence-Driven and Agent-Based Social Simulation

EDABSS models seek homology.[5] The modeled mechanisms, structures and processes aim at resembling the mechanisms, processes and structures identified in the target system. This has two reasons: (1) Construct validity renders validation more

[5] For a meta-theoretical discussion of what follows see Bhaskar (1979), Boudon (1998), Cruickshank (2003), Outhwaite (1987), Sawyer (2005), Sayer (1992, 2000).

expressive. (2) An agent-based implementation of the type presented in the following sections is more than a mere input–output model. It is an "exhibitionist" model that allows to analytically focus on internal mechanisms, processes and structures. From a socio-scientific standpoint this can only be of interest, if the modeled mechanisms, processes and structures exhibit construct validity in comparison with the target system – otherwise the model is just an arcade game (cf. Boudon 1998).

The key to homology lies in the agent design. It is the agents and their interactions respectively that trigger the evolution of emergent phenomena. Thus, at the bottom of EDABSS agent design lays an evidence gathering process. Posing the right questions leads to a collection of data (evidence), which directly informs the modeling of agent behavior and cognition. We have presented such questions in Sect. 25.1.

### 25.3.1 Evidence-Based Modeling

The source of information for homologue models must be evidence-based. This refers to the fact that all information that is used during the process of model design, whether derived from a single case or from a theory, must be empirically valid (see also Boero and Squazzoni 2005). The bulk of this data is of qualitative nature, stemming from one or a number of case-studies. Case-studies that give concrete information of actor behavior, in particular social circumstances, are of best use to EDABSS modelers. Such a presupposition excludes assumption-laden concepts such as rational choice or Belief-Desire-Intention (BDI). EDABSS's higher rational is to found models of social simulation on what is social reality and not what is methodologically convenient or theoretically desirable (cf. Shapiro 2005). It would be wrong to stipulate that all the details entailed in the dataset must also be recognized in an EDABSS model. Modeling is an intellectual condensation process and it is the modeler who decides what particular aspects of a social phenomenon are crucial and need to be represented in a stylized way in the model.[6]

The extensive use of qualitative data in EDABSS can be a virtue of its own, when statistical data is scarce or not available at all. This applies to a variety of important topics in the social sciences, such as elites, power structures, conflict or organized crime. Logically systematic statistical data collection in these areas of research is difficult. Although the same holds true for qualitative data collection as well, it is, nevertheless, better feasible. For example, researchers, journalists or humanitarian aid workers very often have the opportunity to conduct an interview or to make an observation. Often this data becomes available to the public.

[6] The term condensation is alternatively denoted by Stachowiak (1973) as reduction and by Casti (1997) as allegorization. Other important modeling principles are simplicity and pragmatism (Lave and March 1975).

EDABSS therefore fills an important lacuna that is set between abstract statistical modeling and idiographic case-study research as it incorporates the advantages of both, formalization and context-sensitivity.

The integration of stakeholders in the modeling process plays an important role in EDABSS. Integrating stakeholders in the modeling process can be rewarding and delusive at the same time. Stakeholders are keepers of information that others do not have. For example, if a stakeholder is a powerful person, then he can be motivated in a semi-structured interview to reflect on why he thinks he is powerful, how he is acting as a powerful person and how he expects others to behave towards her/him. On the other hand, stakeholder's accounts can be deliberately misleading. Consequentially, EDABSS modelers have to be familiar with qualitative data collection and analysis techniques.[7]

To presume that evidence-based modeling ignores theory is not justified. Evidence-based modeling is guided by theory in many respects. First, critical realism clearly defines a research project's *explanandi*: mechanisms, processes and structures. Second, it highly depends on the researcher – and is not generic to evidence-based modeling – how much the research process is guided by theory. Third, evidence based modeling seeks generalization by intensively studying a single unit for the purpose of understanding a larger class of (similar) units (Gerring 2004).[8]

### *25.3.2 Endorsements: Reasoning About Power and Authority*

Whereas evidence-based models of social simulation incorporate a variety of structural and processual information of the target system, the actor's actual reasoning process cannot be derived from the data. Alternatively, the concept of endorsements is applied to couch an agent's reasoning process.

Power relations, as aforementioned, are interactions between at least two actors. The computational implementation of these interactions must be based on certain grounds. This can be knowledge an actor has about another actor; it can also be experiences an actor has made in the past with his environment. Endorsements are a "natural" way of implementing reasoning about this knowledge or experience.[9] They were introduced by Cohen (1985) as a device for resolving conflicts in rule-based expert systems. Endorsements can be used to describe cognitive trajectories aimed at achieving information and preferential clarity over an agent or object from the perspective of the endorsing agent himself. We use endorsements exactly in

[7] The literature on qualitative data research has grown considerably in the last years and the interested reader is referred to, amongst many others, Silverman (2004).

[8] We are well aware of the ongoing discussion on induction with regard to case-study research and the interested reader may refer, amongst others, to Gomm et al. (1992), Eckstein (1992) and Stakes (1978).

[9] See for a more complete treatment of endorsements Alam et al. (2010).

this sense, namely to capture a process of reasoning about preferences and the establishment of a preferential ordering (Moss 1995, 1998, 2000; Moss and Edmonds 2005).

Endorsements capture an agent's (the *endorser*'s) reasoning process about other agents (the *endorsees*). That process projects the endorser's internal preferences onto the endorsee. These preferences are represented by an endorsement scheme which is a collection of categories of possible characteristics of other agents. These categories of endorsements amount to a partial preference ordering of characteristics perceived in other agents. The ranking of collections of endorsements is an essentially arbitrary process. Cohen (1985) used a lexicographic ordering so that the preferred object (in this case, agent) would be that with the largest number of endorsements in the most valuable category. If several objects were tied at the top level, the second level of endorsements would be used to break the tie and then, if necessary, the third or fourth levels, etc. An alternative is to allow for a large number of less important endorsements to dominate over a small number of more important endorsements. One way of achieving this is to calculate endorsement values $E$ for each endorsee as in Eq. 25.1 where $b$ is the number base (the number of endorsements in one class that will be a matter of indifference with a single endorsement of the next higher class) and $e_i$ is the endorsement class of the $i$th endorsement token (Moss 1995). In choosing from among several other agents, an agent would choose the endorsee with the highest endorsement value $E$.

$$E = \sum_{e_i \geq 0} b^{e_i} - \sum_{e_i < 0} b^{|e_i|} \tag{25.1}$$

The process of endorsing an agent must be thought of as being embedded in an agent's environmental context, i.e. his neighboring agents (see Fig. 25.1). The endorsement process allows an agent to find the agent most appropriate to him – it does not (and cannot) seek the best of all agents.

The main advantage in applying the idea of endorsements lies in the fact that they allow for combining the efficiency properties of numerical measures, with the richness and subtleties of non-numerical measures of interest or belief (Moss 1995).

The choice of endorsements and the conditions in which each endorsement will be attached is entirely context dependent. Agents concerned with critical incidence management in water supply (Moss 1998) obviously have different criteria than agents embedded in the context of contemporary conflicts (Geller and Moss 2008a). While the former might be interested in actions and information models leading to the successful resolution of a complicated allocation problem, agents in models of contemporary conflict might be interested in with whom they should cooperate and whom they should shun or even fight. For example, it is of importance to an agent to know if its *vis-à-vis* is of the same ethnicity, religion and kin; if it has lived a similar past; if it is reliable, corrupt or wealthy. Power and authority relations depend on knowledge about these kinds of questions. Section 25.4 addresses this as well as the question of how to translate the evidence into an adequate endorsement scheme in more detail.

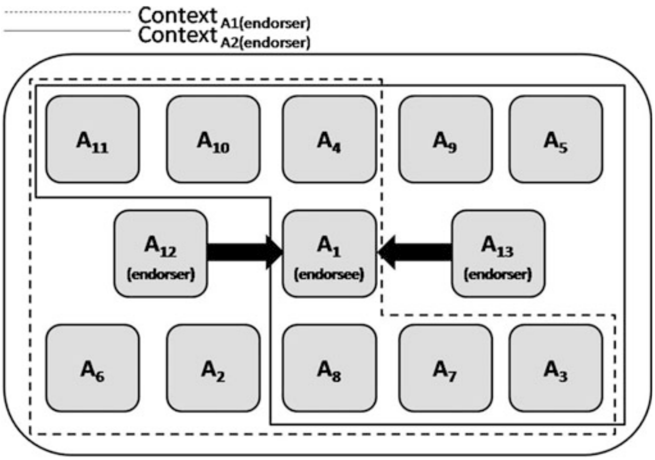



**Fig. 25.1** Schematic representation of the embeddedness of the endorsement process

### *25.3.3 Declarative Implementations of Agent-Based Models*

A program is declarative if there is a set of statements on a database, rules have a set of conditions, which are statements with some values left open as variables, and consequents exist, which are another set of statements. When all of the statements in the conditions of a rule are matched by statements on the database, then the variables are given their specific values from the database statements and the consequent statements are added to the database. When a set of conditions are satisfied and a rule fires (that is puts its consequents on the database), then the state of the environment as represented by the database is changed and perhaps other rules will now be able to fire and so on until all rules have fired and no further matches of conditions can be found on the database. The sequence of rules that will fire and the particular instantiations of their variable values are determined only as the program is running. The sequence of actions represents the process of agent behavior and leads in each case to a new state of the environment. If all agents are implemented declaratively, then they will be changing the state of the environment for one another and the pattern of rules, and therefore actions of all the agents taken together will be influenced by one another.

In these circumstances, the outcomes of such a model are usually impossible to predict with any exactitude.[10] Frequently, such models exhibit the sort of episodic volatility associated in the first section with complexity. The same effect can be achieved by other means, but declarative representations of agents have a number of virtues in terms of ease of development as new evidence becomes available and in terms of yielding comprehensible outputs stored as statements on the databases.

[10] Hence, a declarative model architecture does not allow easily for exact simulation replication. Other disadvantages are that declarative models tend to be computationally expensive and ontologically complex.

The assumption that the fulfillment of conditions triggers the execution of consequents marks, in a homologue model, a natural representation of actor behavior. Each actor behaves according to a defined set of rules. A rule fires only when the set of conditions attributed to this rule is satisfied. Accordingly, an agent's behavior is governed on the basis of the fulfillment of conditions. It is fairly straightforward to translate information on actor behavior obtained during the data collection process into conditions.

Recall the following question from the opening paragraph: "Under what conditions would you label someone as powerful?" A possible answer to this could be "If a person belongs to a well known family." The condition for an agent providing the above-mentioned answer to label another agent as powerful is fulfilled if this other agent belongs to a well-known family (whereas well known could translate into socially well connected). Similarly, an agent might provide the answer "If the agent has won five consecutive battles." In this case an agent has to follow another agent's battle history to be able to tell if it can label him powerful or not. The translation for a possible answer to the second question is analogous. "Having labeled someone as being powerful, how would you expect that person to behave?" could be answered by saying "That the person is generous." However, the translation now includes two conditions. Firstly, an agent must have been already labeled powerful. Secondly, an agent must have experienced the powerful agent being generous to it.

If the conditions of a rule are satisfied, then its consequents are put into effect. Recall the third question stated in the introduction: "Having labeled someone as powerful, how would you expect yourself/others to behave towards that person?" A possible answer could be "Then I would subordinate myself to that person." The translation reads as follows: If an agent has labeled another agent as powerful, it then (as a consequence) subordinates itself to this agent.

Analogously to the examples given, all the collected information that bears relevance to the modeling process can be translated into declarative program code. This translation process is essentially an operationalization and formalization process of (sometimes vague) bits of qualitative information. Power and authority are not implemented as predefined entities, but are "grown" artificially from a number of evidence-based rules that are proxies for dimensions of power and authority. Agents become powerful as a consequence of a variety of causally interconnected conditions and consequents.

## 25.4 Modeling Power and Authority: A Case from Afghanistan

This section presents an implementation of what has been discussed above theoretically. Reflections of power and authority in contemporary conflict are presented, from which an intuitive, but evidence-informed model of power and authority in Afghanistan is derived. Against the background of this model and on the basis of qualitative data, answers in the form of evidence to the questions posed in the opening paragraph are presented. From these answers the agent rules are being developed and translated into program code.

### *25.4.1 Power and Authority in Contemporary Conflicts*

The anthropogenic nature of power structures (Popitz 1992) has been shown for a variety of conflict regions, including Afghanistan (Bayart et al. 1999; Reno 1998; Roy 1994, 1995).[11] Sofsky (2002) argued that conflict societies are societies *sui generis*. They function according to their own social laws and are structurally and processually disjointed from societies lacking a comparable degree of organized violence. In conflict-torn societies virtually anything goes. This can be illustrated by the concept of anomie. Anomie is the situation in which the upper and lower normative boundaries for the aspirations of members of a society are thrown awry (Marks 1974). An anomic situation emerges when the means to attain a specific goal, such as accumulation of wealth or power, run out of social control (Merton 1938). Accordingly, in a space emptied of restricting norms, i.e. an anomie, virtually everything goes along with the creation of power structures to one's own ideas and interests.

Anomic spaces are political spaces lacking strong modern institutions, such as the state's monopoly on organized violence, stability of the law and protection of property rights. In these circumstances only highly adaptive stakeholders prevail. The socio-structural outcomes of this organizational process are manifold and so are the adopted means that serve one's interests.

In contemporary conflict societies this outcome is neo-patrimonialism (Geller 2006a; Medard 1990; Reno 1998). Weber (1980) understands patrimonial power as being based on authority, suppressed subjects and paid military organizations, by virtue of which the extent of a ruler's arbitrary power, as well as grace and mercy increases. Stakeholders interested in gaining power in contemporary conflict settings have to act neo-patrimonially to accumulate and redistribute material as well as social resources. The range of related activities is broad and includes corruption, clientelism, patronage, nepotism, praebendism and so forth (cf. Medard 1990).

### *25.4.2 An Intuitive Model of Power and Authority in Afghanistan*

Anthropogeneity, anomie and neo-patrimonialism – or any other theoretical context relevant to a particular research project – have eminent ramifications for the

[11] Parts of this and the next paragraph have been taken from Geller and Moss (2007).

perception of power and authority and henceforth for the development of the model at hand. The evidence presented below should therefore corroborate the implicit claim that anthropogeneity, anomie and neo-patrimonialism amalgamate to a framework describing Afghan power structures and functioning as an evidence-informed theoretical framework that can be filled with the intricacies of the Afghan case. In the beginning of a research project, such a framework model provides a theoretically informed ontological *entrée* for the object of investigation.

The actual information the model at hand rests upon is derived either from data collected by ourselves or from relevant secondary data sources. The collected primary data stems from semi-structured interviews conducted with urban Afghan elites between May 2006 and October 2007. The secondary data stems from case studies, most of which are of anthropological type, reports published by Non Governmental Organizations (NGOs), International Organizations, such as the United Nations (UN) and the International Committee of the Red Cross (ICRC), or the print media.

Although 27 years of conflict accentuated two important factors in Afghan society, namely ethnicity and religion, the traditional organizational principle of the *qawm* rested sound (Azoy 2003; Roy 1994, 1995; Shahrani 1998). Less mentioned, however, is a decline of norms and values in Afghan society leading to a Hobbesian form of society (Tarzi 1993). Today's Afghanistan can be characterized as an anomie (Geller 2010).

The causes for this development are complex, but nevertheless directly linked to the Jihad of 1979–1989. Although trends of neo-patrimonial politics are already recognizable in the very beginning of the Jihad – and are indeed a characteristic of Afghan politics throughout history – the war's fundamental goals started to mutate with its increasing duration. Some of the adopted means of warfare have been traditional, such as organized violence, intrigue, alliance formation and dissolution; others have been "imported", such as religious extremism and radicalization of ethnicity (Geller 2010; Roy 1998).

The concept of *qawm* is context dependent, defined by such social dimensions as family, kinship, ethnicity and occupational groups, and also more abstract but related concepts such as solidarity, rivalry, cooperation and conflict. The notion of *qawm* also underlines the fluidity and contextual dependency of social relations in Afghanistan. Hence, *qawm* ontologies are devised to codify individual actors, their behavior and relations between actors, as well as social processes and structures arising as a result of social interaction (Dorronsoro 2005:10–11).[12] Each of those aspects is pertinent to the development of our model, which represent and clarify social processes associated with these overlapping identity spheres and the actors acting within them.

[12] Monsutti (2004) "explores the basis of cooperation in a situation of war and migration" amongst the Hazara in Afghanistan through the concepts of solidarity and reciprocity. Nancy Tapper (1991) "reveals the structure of competition and conflict for the control of political and economic resources" through the concept of marriage.

The notion of *qawm* varies not only in the literature, but also amongst Afghans themselves. It can mean (extended) family, tribe, descent group, ethnicity, "people like us" (Tapper 2008:88), an "occupational group" (Roy 1992:75), "persons who mutually assist each other" (Canfield 1973:35) and it can connote a complex interpersonal "network" (Roy 1995:22; Dorronsoro 2005:10) of political, social, economic, military, and cultural relations (Mousavi 1997:46–48; Tapper 2008; Glatzer 1998:174; Rasuly-Paleczek 1998:210–214; Roy 1995:21–25; Shahrani 1998:218–221).[13] In fact, our interview data suggests that these meanings are not mutually exclusive: *qawm* do not have clear boundaries nor do they divide Afghan society into mutually exclusive groups. "[A]n individual always belongs to more than one [*qawm*]." (Canfield 1988:194)

*qawm* face competition with other *qawm* and internal competition amongst members of a *qawm* (Azoy 2003; Mousavi 1997:46–48; Roy 1994:74, 1995:21–22). *qawm* need to be sustained, and it is an Afghan leader's ability to redistribute resources that makes him powerful and eventually successful (Roy 1994:74). The ability to create a *qawm* for a particular aim is also perceived as a demonstration of power (Azoy 2003:36). *qawm* still "have a powerful and pervasive effect on contemporary political discourse and the behavior of Afghans" (Shahrani 1998:220) and have during the years of conflict often been misused by new elites for the pursuit of conflict and criminal aims (Canfield 1988; Rasuly-Paleczek 1998:210–214; Roy 1994; Rubin 1992; Shahrani 2002; Tapper 2008). Manifestations of such abusive behavior are for example corruption, drug production and smuggling, nepotism, massive organized violence, crime, ethnic, political and religious radicalization (Giustozzi 2007; Glatzer 2003; Rubin 1992, 2007; Schetter et al. 2007). *qawm* are not the cause for conflict in Afghanistan, as these causes are manifold (Shahrani 2002:716; Dorronsoro 2005), but we will explore the usefulness of the notion of *qawm* in analyzing and understanding conflict in Afghanistan.

Figure 25.2 depicts an informed intuitive and ideal-typical representation of a *qawm*. It consists of ten actor types: politicians, religious leaders, commanders (meritocratic title for a militia leader), businessmen, warriors, civilians, farmers, drug farmers, organized criminals and drug dealers. An important abstraction from reality is that in our model each actor has its distinct role, whereas in reality actors may incorporate a variety of roles. For example, a commander can be a (military) commander, a politician and a drug lord at the same time. We proxy individual role pluralism by mutual interdependence, i.e. each actor has virtues another actor may be in need of and vice versa, leading to mutual cooperation and interdependence. This, of course, is also a common pattern in reality, where there is no clear distinction between role incorporation and cooperation.

[13] Whether a *qawm* denotes a group or a network is not clear from the evidence. Following Tapper's (2008) argument, a *qawm* can take the form of a group or a network, depending on the context.

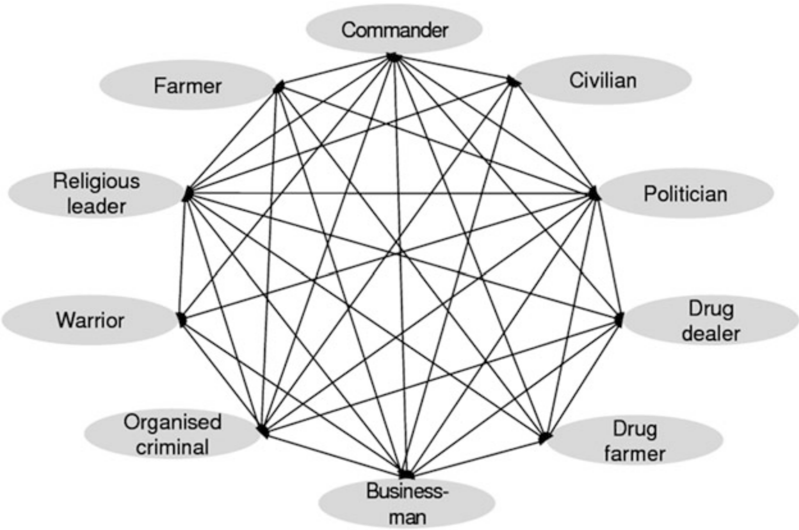



**Fig. 25.2** A case-study informed intuitive model of a *qawm*

The following examples explain the *qawm*-model in terms of agency. If a politician is in need of military protection, he approaches a commander. In return, a commander receives political appreciation by mere cooperation with a politician. If a businessman wants to be awarded an official construction contract by the government, he relies on a politician's political connections. In return, the politician receives a monetary provision, for example, bribes. If a politician wants beneficial publicity, he asks a religious leader for support. The religious leader, in return, becomes perceived as a religious authority. If a warrior seeks protection and subsistence for his family, he lends his services to a commander, who, in return, provides him with weapons, clothes, food and/or money. If an organized criminal wants to carry drugs, he relies on the transport business of a businessman who, in return, receives a share of the drugs sold. If a drug farmer needs protection for his poppy fields, he a affiliates with a commander, who, in return, receives a tithe of the drugs sold to a local drug dealer. According to Azoy (2003), such interactions are also guided by the following four social categories: kinship, residence, class, and religion. Our model represents this neo-patrimonial behavior. The links between the agents in Fig. 25.2 can also represent such categories.

The continuing existence of the *qawm* in times of severe social change as a means to organize and manage power cannot baffle the fact that the *qawm* itself has undergone configurational alteration. Protracted conflict deteriorated not only social structure but also obliterated moral boundaries. Corruption, fraud, mistrust, crude materialism, and the like systematically found their way into Afghan society. The power of the *qawm* and of its members to constantly adapt to new states of anomie epitomizes the anthropogenic nature of power and authority (Geller 2010).

### 25.4.3 An EDABSS Model of Power and Authority in Afghanistan

#### 25.4.3.1 Evidence

The collected interview data explicitly highlights three sources of power in current Afghanistan: Ownership, reputation, and *qawm* (cf. Azoy 2003). Their meaning manifests when mirrored against the notions of *hisiyat* and *e'tibar*. *hisiyat* and *e'tibar*, two Dari words that roughly translate into "character" and "credit". *hisiyat* denotes qualities such as piety and wisdom; *e'tibar* is about meritocracy. A powerful actor must dispose of *hisiyat and e'tibar*.

Traditionally, ownership can be defined as land, access to water, livestock and women. References to landownership were made often during interviews, whereas water and women have never been mentioned. Ownership of livestock was mainly mentioned to serve reputational means in order to increase one's own or someone else's reputation. The interview data and observations made during the field trips suggest that a modern comprehension of ownership has become more materialistic and less subsistence oriented; mundane symbols of power such as money, houses and cars have increased in importance. This raises the issue of the sources for these goods. While some Afghans undoubtedly were able to build up prospering businesses or brought assets with them from exile, other sources remain dubious and are likely to include organized crime, corruption and clandestinely working for foreign countries. Thus, a generalized answer to "When would you label someone as being powerful?" includes that, either being traditional or modern, or, more likely, a combination of the both, a powerful individual must have access to ownership resources. "[. . .][P]ower without wealth is all but possible." (Azoy 2003:30) But "[w]ealth is just a means to achieve prestige." (Roy 1994:74)

In Afghanistan, authority is nourished by reputation. It is the "ultimate source of political authority." (Azoy 2003:31) Reputation exhibits a static and dynamic component. The static component is closely related with ancestry. It is important what ethnicity, family or tribe someone belongs to. Hazaras, one of the four major ethnicities in Afghanistan, are often regarded as working class, while Pashtun, the largest ethnicity, are often perceived as a warrior elite. A family may be regarded as politically powerful and/or religiously influential. To have roots in a political, religious or scholarly family provides authority. This became obvious during interview sessions in a variety of ways. Interviewees have regularly been introduced or have introduced themselves by referring to their family either in a political or religious context. Two outspoken authoritarian interviewees, for example, claimed to be Sayyed, i.e. descendants from the Prophet, and scholars at the same time. The dynamic aspect of reputation relates to an individual's historicity. Individual politico-historical background is important. Individual histories link actors to different social groups and can provide them, depending on the social context, with esteem, such as in the case of having been a Mujahedin, a resistance fighter during the time of Soviet occupation. Thus a generic answer based on the interview data collected to "when would you label someone as being powerful?"

would include that a powerful individual belongs to an important family and/or has played an important role in his past.

The *qawm* is the epitomization of network-based power structures in Afghanistan. A *qawm* is more than mere reference to an actor's ethnicity or tribe; it is more than an extended family. The *qawm* is an actively used instrument in the pursuit of power. An appropriate translation would therefore be "power basis": a number of people that can be mobilized to achieve a particular political aim. With changing political aims, this group of people may change as well. Thus a generalized answer to the question "when would you label someone as powerful?" includes that a powerful individual not only disposes of a *qawm*, but also exerts control over it.

#### 25.4.3.2 Model and Computational Implementation

Based on the evidence, how are a powerful actor and his entourage to be structurally modeled? First, it needs to be defined who is powerful and who is not by disposition. The evidence presented above indicates that the following agent types should be considered as being powerful in our model: Politicians, businessmen, religious leaders, commanders and organized criminals. These agents are computationally created as being powerful by definition, because they are politicians, religious leaders, commanders, etc.[14] Hence, social resources do not have to be distributed explicitly, agents are "born" possessing them – or not.

Secondly, agents must, at the initialization of the model, be equipped with material resources. These assets are given, *pars pro toto*, in the form of money, drugs and land. The absolute amount of money distributed in the model is arbitrary as real numbers are not available. Money is distributed log-normally amongst agents of one particular agent type. A lognormal distribution of wealth also appears for the case of Afghanistan a plausible assumption (Limpert et al. 2001). Hence, a small number of agents are very rich, while most of the people are poor. The data record for land holdings, which is better than the one for wealth, suggests that holdings of land should be distributed log-normally as well (Wily 2004). Again, this means that a small number of agents possess a lot of land, while a large number of agents will only own little land. Last but not least, drug farmers receive some drugs in the beginning of the simulation and harvest drugs during the simulation in specified harvesting periods (UNODC 2006).

Thirdly, agents must be given an internal state, representing what has been identified above as *hisiyat* and *e'tibar*. This internal state is an agent's endorsement scheme, which is depicted in Table 25.1. Some of these labels (the endorsements) are static, others are dynamic. Static endorsements are attributed to each agent at

[14] Note that we are not simulating the genesis of a powerful agent, but the emergence of power as a network-like structure in an evidence-based, artificial society. See for the qualitative description of such a genesis Giustozzi (2006).

**Table 25.1** Crosstabular presentation of an agent's endorsement scheme

| | Static | Dynamic |
|---|---|---|
| *hisiyat* | Intellectual/non-scholarly | Loyal/disloyal |
| | Shared-ethnicity/different-ethnicity | Trustworthy/untrustworthy |
| | Shared-religion/different-religion | Is-neighbor/non-neighbor |
| | Is-kin/non-kin | Pious/sinful |
| | Politico-military-background | |
| *e'tibar* | | Reliable/unreliable |
| | | Successful/unsuccessful |
| | | Capable/incapable |

the beginning of a simulation and cannot be changed during the course of the simulation: an agent is either an intellectual or he is not; he is either a Tajik, another important Afghan ethnicity, or he is not; he is either my brother or he is not, etc. For the time being it is not implemented that an agent's changing relationships are taken into account to form an individual political profile, hence the politico-military background is static. Some *hisiyat* endorsements are dynamic, such as loyalty, trustworthiness, neighborhood or religiousness and can change their respective values during a simulation run. All *e'tibar* endorsements are, by contrast, dynamic, as *e'tibar*, i.e. meritocracy, is inherently dynamic and must call for a dynamic conceptualization. Formerly reliable agents can become unreliable and previously successful agents can become unsuccessful, etc.

The endorsement scheme does not only depict an agent's internal state but also denotes the categories in which an agent reasons about other agents. For example, assume agent A is Tajik, a Mujahedin, successful and neighbor of agent B. Further assume that agent B is Tajik as well, was also a Mujahedin and evidently is also neighbor of A. It is then likely that agent B will look favorably at agent A and vice versa and that the two will establish an affiliation with each other. The endorsement scheme breaks down why an agent is more powerful than another agent: Because it disposes over an internal state that is seen favorable by other agents. Or to put it differently: Because it internalized a number of qualities that are perceived as symbols of power by other agents.

It is important to note that although an individual agent is defined at the moment of its creation as being powerful or not per se, nothing is said about how powerful it is going to be, or spoken differently: How good of a neo-patrimonial agent it will be. The agent's performance as well as the social product of its performance, the *qawm*, are emerging out of the social simulation model and are not computational artifacts.

#### 25.4.3.3 Behaving Powerfully

Having clarified what makes actors powerful in Afghanistan, it is now important to know how these powerful actors behave and how other actors behave towards powerful actors. The corresponding questions from the introduction are: "Having accepted someone being powerful and/or being an authority, how would you expect that person to behave?" And: "Having labeled someone as powerful/as being in a position of authority, how would you expect yourself/others to behave towards that person?" The freedom of choice an actor has to behave towards a powerful agent depends on a number of different factors (*hisiyat* and *e'tibar*) but is foremost based on the distinction whether an actor is powerful himself or not. The two constellations powerful actor/weak actor and powerful actor/powerful actor lead to different outcomes in the social organization. In our understanding these are patron-client-relationships (powerful/weak) or affiliations (powerful/powerful). However, in both cases, accumulating and redistributing resources is in the center of actor behavior.

Weak actors have only little choice of how to behave towards a powerful actor and are forced into a patron-client relationship because of grievance. They are either locked in into economic dependency or are not even a member of the powerful actor's *qawm* and thus cosmos. In both cases weak actors can only submit themselves and must fully depend on the powerful actor's gratitude. Consequentially, in patron-client-relationships, powerful behavior is more of a mundane sort, for example supporting a client's family with food, clothes and housing. Depending on what kind of client it is, the patron might ask for dogs body services in the case of a civilian, for protective services in the case of a warrior, or for *zakhat* (tithe) in the case of a farmer. In general, the patron demands loyalty for his support. The weak actor exerts power over the powerful actor indirectly insofar as it is harmful for a powerful actor's reputation to pay subsidies irregularly or not at all. The weak actor exerts power directly in case of grass-root opposition as a result of insensitive politicization by the powerful actor. This social relationship of dependence is standard in Afghanistan and every powerful man surrounds himself with such a "service force", be it small or big. Nevertheless, supporting the weak should not be considered as being unimportant, as they provide a basis of broad social support. Moreover, supporting the weak increases a powerful man's *e'tibar* as the following ideal-typical characterization of a Pashtun men highlights: He is a men of honor, with prowess and pride, whose dignity does not forbid him to be attentive to authorities as well as the weak (Janata and Hassas 1975:84; translation ours).

Powerful behavior between two or more powerful agents is of a different nature. Powerful actors have the freedom to choose their behavior towards their *vis-à-vis* and can either act cooperatively, conflictously or submissively. Which type of behavior is chosen, is based on a deliberate but nevertheless fragile assessment of who must be considered a supporter and who must be considered a spoiler or even a foe. Hence, whatever the project is that is of concern to a powerful actor, it is intensively discussed with those people from his social circle who he thinks should be included in the decision making process (Azoy 2003).

The evaluation of a project's – and ultimately of a powerful actor's behavior – supporters and opponents constitutes the initialization of a project-based *qawm*. The organization of a *qawm*, whether it is a generic *qawm* based on kinship or a temporal *qawm*, is a delicate operation. Exacerbating is the fact that potential supporters as well as potential enemies are competitors against whom the powerful actor must stand up – at all time. Hence, the creation and maintenance of *qawm* explains the volatile nature of cooperation, the often and sudden changes of alliances and the ubiquity of conflict (cf. Azoy 2003).[15]

In the case of cooperation, powerful actors establish affiliations between each other. Affiliations are relations between qualitatively equals. Consequentially, powerful agents do not give each other material support, but they provide each other with social resources. While powerful actors in patron-client relationships accumulate social resources for redistributing material resources, powerful actors in affiliations mostly accumulate and redistribute social resources. Politicians guarantee that commanders are not denigrated as "warlords", commanders protect politicians, religious leaders openly designate politicians of being pious, politicians declare a religious leader their spiritual leader, businessmen – financially support a politician's campaign, politicians provide businessmen with lucrative state contracts, etc. In short, powerful actors support each other in in-creasing their *hisiyat* and *e'tibar* record. The opposite, of course, exists as well. Powerful actors can actively engage in diminishing another actor's *hisiyat* and *e'tibar* record.

In summary, a powerful actor is able to control his *qawm* economically and socially. He is able to redistribute enough material and social resources to keep his *qawm* alive. If one would have to measure the power an actor in Afghanistan has, then it would not suffice to only count his material assets. A comprehensive measure of power would include this actor's reputation, measured in terms of his ability to "call on the services of supporters to help him in whatever enterprise" (Azoy 2003:32).

#### 25.4.3.4 Model and Computational Implementation

The computational implementation of the behavior of a powerful agent is straightforwardly informed by the evidence presented above. Powerful agents want to accumulate and redistribute material and social resources. For this reason their *hisiyat* and *e'tibar* needs to be relatively superior to their competitors'. Hence, a powerful agent's aim must be to establish as many favorable relationships as possible. He does this in two steps: First he reasons about which agents to endorse and second he takes action on the basis of his decision he has taken in step 1.

Step 1 consists of the endorsement process as explained in Sect. 25.3.3 and as contextualized in Sect. 25.4.3. Each agent continually checks all the agents visible to him – not all of them are –, i.e. he projects his endorsement scheme upon them, rates the corresponding values, calculates $E$, compares all $E$s against each other and chooses the one with the highest $E$ to endorse. According to the evidence, the model implies that during the endorsement process those agents are more likely to

[15] We do not consider the emergence of conflict in this chapter. See for a preliminary discussion Gerring (2004).

**Table 25.2** Crosstabular presentation of a sample endorsement between one civilian and two politicians

| Endorsements | Civilian | $\text{Politician}_1$ | $\text{Politician}_2$ |
|---|---|---|---|
| Same/different ethnicity (±1) | Tajik | Hazara/–2* | Pashtun/–2* |
| Same/different politico-military background (±2) | Mujahedin | Mujahedin/4 | Mujahedin/4 |
| Reliable/unreliable (±3) | – | unreliable/–8* | reliable/8 |
| *E* | – | –6 | 10 |

The values in parentheses indicate the importance of an endorsement. Values on the left side of each slash represent "reality"; values on the right side of each dash represent the "weighed reality". Labels marked with an asterisk have been multiplied with $-1$, because in Eq. 25.1 they are part of the second $\Sigma$ sign and are therefore subtracted

establish a relationship with each other if they are similar with regard to *hisiyat* endorsements and who exhibit higher values with regard to *e'tibar* endorsements. Two short examples shall clarify this, one for patron-client-relationships and one for affiliations.

Powerful agents do not seek ordinary agents; rather they are sought out by the latter. But even though ordinary agents are in misery, they try to choose the powerful agent who is most suitable for them, given the choice. Assume a civilian, who is Tajik, a Mujahedin and in need of material support (see for what follows also Table 25.2). (In the simulation this is the case when the civilian's holdings of money are $\leq 0$.) Before asking every politician for material support the agent can see, it checks which of the visible politicians are most suitable to him. Assume that there are two visible politicians. The civilian agent then projects its endorsement scheme upon each of the two politician agents. Recall that the endorsement scheme not only tells the civilian how important particular characteristics of these two politicians are to it, but also what the categories of its own perceptions are. In Eq. 25.1 $b$ denotes the number base for which every agent is randomly assigned a value $b > 0$. (For $b = 0$ the expression $b^x$ is equal to 0, independently of the exponent $x$.) In this example, the civilian is assigned $b = 2$. In Eq. 25.1 $e_i$ denotes the value of the $i$th endorsement token. This value differs for each endorsement token and for each agent. In this example the endorsement tokens for the civilian are same-ethnicity/different-ethnicity, same-politico-military-background/different-politico-military-background and reliable/un-reliable. Each endorsement token is randomly assigned a value $e_i$. $e_i$ differs for each agent. In the present example the following values for $e_i$ are assigned for the civilian: 1 for same-ethnicity, $-1$ for different-ethnicity, 2 for same-politico-military-background, $-2$ for different-politico-military-background, 3 for reliable and $-3$ for unreliable. These values are on an ordinal scale and represent the civilian's endorsement scheme. The interpretation of this endorsement scheme is that the civilian perceives the labels same-ethnicity/different-ethnicity less important than the labels same-politico-military-background/different-politico-military-background and the labels same-politico-military-background/different-politico-military-background less important than the labels reliable/unreliable. $\text{Politician}_1$ is a Hazara, a Mujahedin and has never been endorsed reliable by the civilian. Politician$_2$ is an Uzbek, a Mujahedin and has been endorsed reliable in the past by the civilian. With regard to politician$_1$ the civilian reasons as follows: Politician$_1$ is not of the same ethnicity, has the same politico-military-background and is unreliable. Depending on this information and the values assigned above, (Eq. 25.1) allows the calculation of $E$, which is: $2^2 - 2^{|-1|} - 2^{|-3|} = -6$. For politician$_2$ the civilian reasons as follows: Politician$_2$ is not of the same ethnicity, has the same politico-military-background and is reliable. Thus, $E$ can be calculated as follows: $2^2 + 2^3 - 2^{|-1|} = 10$. Because politician$_2$ has a higher $E$ than politician$_1$ the civilian decides to choose politician$_2$. This procedure can be extended to as many agents and to as large an endorsement scheme as necessary.

Once the civilian has taken its decision, which politician to ask for support, the agent sends a message to this particular politician, requesting support. As long as the politician has enough money to support the civilian, it accepts the support request. The acceptance of the support request is tantamount to the establishment of a patron-client-relationship. Each simulation iteration, the politician pays the civilian a defined amount of money, as long as it has enough money to do so. In return, the civilian is required to endorse the politician as being trustworthy. In the event that the politician cannot pay the civilian anymore, the latter endorses the politician being unreliable and untrustworthy, leading to a break-up of the patron-client relationship. The civilian then has to seek another politician and the endorsement process starts again. All requests for support by ordinary agents follow this scheme.

Interactions between powerful agents processually do not differ substantially from the scheme described above. Consider the case where a commander wants a trustworthy assertion from a politician in order to not be denigrated as “warlord”. Assume that two politicians are visible to the commander. Again, before sending off a message to the most suitable politician, the commander assesses which of the two politicians are most suitable to it, i.e. which one has the higher $E$. Although different and further endorsements might apply in this case, the endorsement process, as described in the civilian-politician case, does not change in essence. Once the commander has chosen a suitable politician – one, for example, who is of the same ethnicity, the same politico-military back-ground and who has a record of being trustworthy – it then sends a message to this politician requesting to be endorsed trustworthy. Because the commander has chosen a politician with a trustworthy record during its endorsement process, the politician is able to do so, i.e. to endorse the commander trustworthy. However, the politician demands a service in return. When dealing with a commander, this service is naturally protection. Hence, the politician agent sends a message to the commander agent, stating that it will endorse it trustworthy, but only if the commander agent can protect the politician agent. Naturally, the commander can provide protection only, if there is at least one warrior with whom he has a patron-client relationship, established according to the procedure described above. Given he can provide protection, he accordingly answers the politician’s message. This mutual fulfillment of conditions triggers a number of mutual endorsements: The politician endorses the commander as not only being trustworthy, as requested, but also as being capable, as he is capable of providing protection; the commander endorses the politician as being trustworthy. Mutually endorsing each other is tantamount to establishing an affiliation. This affiliation holds as long as the conditions leading to it hold true: “Is the commander still able to provide protection?”, “Is he still capable?”, “Is it still trustworthy?”, “Is the politician still trustworthy?”. Both agents do this every simulation iteration. Moreover, all relationships, whether they are patron-client relationships or affiliations, break up when there is an agent (endorsee) found who fits an endorser better, because agents continually scan their neighborhood for better opportunities.

The example given above is representative for all the interactions between powerful agents: Politicians request armed protection from commanders for a trustworthiness assertion; commanders approach businessmen to invest money; businessmen pay politicians for courtesy services; politicians request a pious assertion from a religious leader for a trustworthiness assertion; religious leaders ask commanders for protection and, in return, assert a commander as being pious. All these affiliation interactions describe the neo-patrimonial usage of social capital as well as the accumulation and redistribution of material resources.

The quintessential mechanism is the endorsement scheme. It is the interface through which two agents communicate with each other and it is the raster that filters those who are “endorsables” from those who are “un-endorsables”. If two agents match to establish a patron-client relationship or an affiliation, is ultimately decided via the mechanism of the endorsement scheme. Hence, the evidence that is enshrined in the endorsement scheme (see Sect. 25.4.3 and there in particular Table 25.1) finally leads to the emergence of social reality.

#### 25.4.3.5 Emerging Structures: Simulation Results

It would be beyond the scope of this chapter to discuss the simulation results and their validation in detail (see Chaps. 8, 9 and 12) in this volume, also Moss 2007 for more on how to do this).[16] Figure 25.3 depicts the output of a representative simulation run at time $t = 100$. There are ten different agent types and the total number of agents is 190: 6 politicians, 6 religious leaders, 6 businessmen, 6 organized criminals, 6 commanders, 10 drug dealers, 35 drug farmers, 35 farmers, 70 civilians and 20 warriors. In summary, these effects represent the emergence of higher order organizational structures or *qawm* out of the micro-processes and -structures introduced above.

Agents affiliated with each other are linked via a line. Two distinct but nevertheless interconnected clusters of agents are apparent in the network. Each cluster consists of a variety of agent types. This means that the depicted clusters are not homogeneous organizations of power, but rather heterogeneous concentrations of power generated by mutually dependent and interacting agents/actors. Agents

[16] This paragraph is a condensed version of Geller and Moss (2008a). See also Geller and Moss (2007; 2008b).

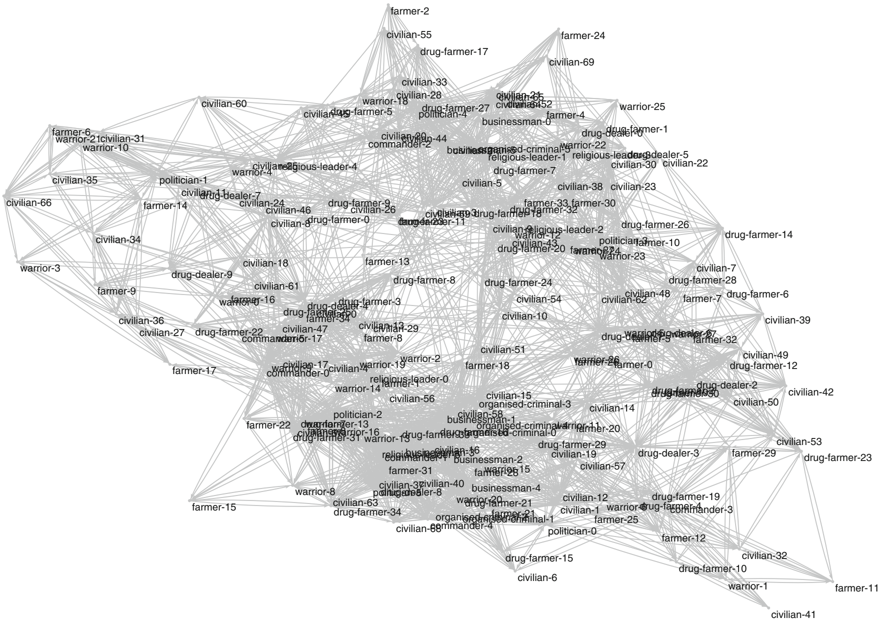



**Fig. 25.3** Artificial Afghan power structures depicted as a relational network of Afghan agents

assumed to be more powerful than others, i.e. politicians, commanders, religious leaders and organized criminals, are prevalent in each of the two dense clusters. The reasons for the evolution of this network of clustered affiliations are manifold: Agents affiliate because they share the same ethnicity or religion; because they have established a business relationship or because they seek protection with a commander. But in general, the clusters can be perceived as emergent properties of agent neo-patrimonial behavior as reified by our agent rules. The model generates data of the sort we expected and its output can therefore be considered as artificial representations of *qawm*.

A cogent argument for this claim is the fact that a cross-validational analysis between the network constructed from simulation output and a real network, constructed by Fuchs (2005), was successfully conducted. Cross-validation provides a link between the model and its target system, i.e. reality (Moss and Edmonds 2005). Fuchs (2005) has collected open source data on Afghan power structures for the years 1992–2004/2005, which can be compared with the AfghanModel network depicted in Fig. 25.3.[17]

[17] Fuchs (2005) only collected data of elites. In order to compare her results with those generated from the simulation presented here, all ordinary agents, i.e. non-elites, had to be removed from the network to meaningfully calculate the desired network measures. Note that the simulation parameters remained unchanged. Note also that the two networks vary in size: 62 agents participate in the Fuchs network and 30 in the AfghanModel network. This can lead to boundary specification problems.

The densities for the Fuchs and the AfghanModel networks are 0.0593 and 0.1943* (only strongmen, marked with *) and 0.0697 (208 agents, marked with †). The clustering coefficient for Fuchs is 0.428; while it is 0.542* and 0.432† respectively for the AfghanModel. Both the Fuchs and the AfghanModel networks tend to be small world. They exhibit sub-networks that are characterized by the presence of connections between almost any two nodes within them. Most pairs of nodes in both networks are connected by at least one short path. There is an over-abundance of hubs in both networks. The average geodesic distances for the Fuchs and the AfghanModel networks, 2.972, 3.331* and 2.628†, are relatively small compared to the number of nodes. Erdős-Rényi random networks of equal size and density as the Fuchs and the AfghanModel networks exhibit lower clustering coefficients than the Fuchs and the AfghanModel networks, namely 0.060, 0.212* and 0.072† – an indication that neither the clustering of the Fuchs nor of the AfghanModel network is random. The two Erdős-Rényi random net-works have geodesic distances (2.094*, 2.276†) which are of a comparable order as the geodesic distances that can be found in the AfghanModel network (3.331*, 2.628†). Thus, a structural and functional equivalence based on qualitative evidence between the model and the target system is observable.

## 25.5 Conclusions

The main task of this chapter was to provide a critical overview of state of the art models that deal in alternative ways with power and authority, and to present an alternative research design that overcomes the inefficiencies and shortcomings of these approaches. The work presented is motivated by the fact that research on power structures is confined on one hand by a general lack of statistical data. On the other hand, the literal complexity of power structures requires a formalized and dynamic approach of analysis if more than a narrative understanding of the object under investigation is sought. The case of Afghanistan has only been instrumentalized to exemplify such an approach, which would work without doubt for any other comparable contexts.

With regard to the case in hand the *explanandum* is power and authority in Afghanistan. The analysis focuses on power relations dominated by elites. The *qawm*, a fluid and goal-oriented solidarity network, has been identified in the data available to us as the pivotal structural and functional social institution to manage power and authority in Afghanistan. The totality of *qawm* in Afghanistan do not form a unified system of power but a cosmos of mutually interacting power systems. This qualitative analytical result has been reproduced by our simulation results and was subsequently cross-validated with independent out-of-sample network-data. This cosmos is a root source for political volatility and unpredictability and ultimately an important explanatory factor for conflict in Afghanistan. For the time being, the latter only outlines an a priori statement that needs further corroboration.

The proposed approach starts with an evidence-informed but intuitive model of power and authority in Afghanistan. Such a model provides an intellectual *entrée*, identifies and defines a target system, isolates relevant actors and generic actor behavior, and helps to address appropriate research questions. Based on this intuitive model, agent structures and rules are developed according to the evidence that is available. The aim is to develop a homologue, i.e. a construct valid model, which allows translation of the information describing the model with regard to structure and processes into program code. Weak agents have been implemented according to the rational that they seek affiliation with a powerful agent because of grievance and their will to survive; a powerful agent is implemented according to the rational that he affiliates himself with other powerful agents in terms of functional complementarity to subsist his solidarity network, i.e. his *qawm*, with the aim of consolidating or increasing his power. The general underlying notion of such behavior is neo-patrimonialism. The simulation results are not self explanatory but need to be validated against reality. This not only provides new insight into the target system but also obliges to restart the research process as new evidence has become available. This is the hermeneutic circle of EDABSS, in which the manner of how to present simulation results meaningfully to stakeholders still needs to be considered as being in its experimental stage (though companion modeling has shown a very fruitful way forward). The approach introduced here does also constitute a consequent implementation of generative social science, for which, nevertheless, the research design still needs further formalization and clarification.

Special consideration was given to a cognitive process called "endorsements". The idea of endorsements serves two aims: *firstly,* to differentiate and define the relevant dimensions agents reason about and, *secondly*, to implement agent cognition in a natural way (being able to use the mnemonic tokens found in the evidence informing the model). Endorsements as they were introduced here lack two important features: that agent types should have randomized, type-specific endorsement schemes, and that Eq. 25.1 to calculate $E$ should allow for continuous data formalization. Alam et al. (2010) have proposed a solution for these problems.

What has been gained by this approach? First, tribute has been paid to reality by taking it seriously into account. Although evidence-driven modeling is about abstraction and formalization like any other modeling technique, it does it on the basis of evidence. Secondly, while agent-based modeling accounts in general for an epistemological shift from an intra-modeling view and disburdens modeling from serving only a mere input–output function, evidence-driven modeling facilitates cross-validation also between model and target system inter-agent mechanisms and on the systems' respective meso-levels. Thirdly, the inclusion and generation of narratives in evidence-driven modeling opens up new ways of engaging stakeholders and domain experts in the modeling process and in informing policy analysts and makers in their decision-making.

However, the heuristic usefulness and value of the applied approach can only be determined against the actual object under investigation, i.e. Afghanistan. What has been found by modeling Afghan power structures evidence-driven and agent-based that would have not been detected by applying only a hermeneutic analysis?

Modeling requires not only abstraction, but also formalization and thus disambiguation of the evidence describing the case at hand. In particular we were forced to dissect the notion of *qawm*, an inherently context-dependent and fuzzy concept, with regard to actor behavior and cognition, assigning clear meaning to all mechanisms deemed to be relevant to the model. This being only a beneficial side effect of evidence-driven social simulation, emerging social processes and structures are in the focus of description and analysis, social dynamics that could hardly be made graspable by pure narrative analysis. Here this is the evidence-driven generation of a model-based demonstration of an autopoietic system of power structures taking the form of a small world network.

From the point of view of complexity science it is interesting to observe how the introduction of localized mechanisms of power generate order on a higher level, or to put it differently: How neo-patrimonial behavior and processes of accumulation and redistribution in a state of anomie create a social structure which can be found in Afghanistan, i.e. *qawm*. Further, EDABSS clarifies an important aspect of emergentism: It is not sufficient to state that in agent-based modeling agents constitute their own environment (cf. Cederman 2001). In lieu thereof it should be replenished that agents and the emergent effects stemming from these agents' interactions as a whole constitute an agent's environment. Agent behavior on the micro-level and social structurization on the macro-level cannot be thought of as disjointed entities, but emblematize the wholeness of what is in essence a complex system. EDABSS therefore is not only a methodological solution to the micro–macro gap problem, but also an implementational one.

EDABSS implies more than inductive reasoning about social phenomena. It constitutes a consequent implementation of the generative social science paradigm. Evidence is the starting point of the research process and evidence denotes its end. It is the nature of the object under investigation that shapes the theoretical and methodological approach and not vice versa. This is perhaps EDABSS' most important virtue that it takes reality and its subjects seriously: during evidence collection, model development and validation.

**Acknowledgments** We would like to thank Bruce Edmonds, Seyed M. Mussavi Rizi, Martin Neumann, and Flaminio Squazzoni for thoughtful and helpful comments. We also thank Zemaray Hakimi for translation and facilitator skills, Sayyed Askar Mousavi for advice in the data collection process, Shah Jamal Alam, Ruth Meyer and Bogdan Werth for modeling support, and the Bibliotheca Afghanica for access to its library.

## Further Reading

Whereas the literature on power and authority is overwhelming, published work on power and authority *and* modeling and simulation is, comparatively speaking, meager. For further reading we suggest Alam et al. (2005) and Rouchier et al. (2001) for models concerned with the emergence of structures and authority in gift exchange; Geller and Moss (2008a) and Alam et al. (2008) for empirical models relevant to power and authority; Axelrod (1995) and Cederman (1997) for applications of modeling power to conflict in international relations; Mailliard and Sibertin-Blanc (2010) for a good discussion of multi-agent simulation and power from a sociological perspective; and finally Guyot et al. (2006) for a participatory modeling approach with relevance to power and authority.

## References


Alam SJ, Hillebrandt F, Schillo M (2005) Sociological implications of gift exchange in multiagent systems. J Artif Soc Soc Simul 8(3), http://jasss.soc.surrey.ac.uk/8/3/5.html

Alam SJ, Meyer R, Ziervogel G, Moss S (2008) The impact of HIV/AIDS in the context of socioeconomic stressors: an evidence-driven approach. J Artif Soc Soc Simul 10(4), http://jasss.soc.surrey.ac.uk/10/4/7.html

Alam SJ, Geller A, Meyer R, Werth B (2010) Modelling contextualized reasoning in complex societies with 'endorsements'. J Artif Soc Soc Simul 13(4), http://jasss.soc.surrey.ac.uk/13/4/6.html

Axelrod R (1995) A model of the emergence of new political actors. In: Gilbert N, Conte R (eds) Artificial societies: the computer simulation of social life. Routledge, London/New York, pp 19–39

Azoy GW (2003) Game and power in Afghanistan, 2nd edn. Waveland Press, Long Grove

Bak P (1997) How nature works: the science of self organized criticality. Oxford University Press, Oxford

Bayart J-F, Ellis S, Hibou B (1999) The criminalization of the state in Africa. Indiana University Press, Bloomington/Indianapolis

Bhaskar R (1979) The possibility of naturalism: a philosophical critique of the contemporary human sciences. The Harvester Press, Sussex

Boero R, Squazzoni F (2005) Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science. J Artif Soc Soc Simul 8(4), http://jasss.soc.surrey.ac.uk/8/4/6.html

Boudon R (1998) Social mechanisms without black boxes. In: Hedström P, Swedberg R (eds) Social mechanism: an analytical approach to social theory. Cambridge University Press, Cambridge, pp 172–203

Caldas JC, Coelho H (1999) The origin of institutions: socio-economic processes, choice, norms and conventions. J Artif Soc Soc Simul 2(2), http://jasss.soc.surrey.ac.uk/2/2/1.html

Canfield RL (1973) Faction and conversion in a plural society: religious alignments in the Hindu Kush. University of Michigan, Ann Arbor

Canfield RL (1988) Afghanistan's social identities in crisis. In: Digard J-P (ed) Le fait ethnique en Iran et en Afghanistan. Editions du Centre National de la Recherche Scientifique, Paris, pp 185–199

Castelfranchi C (1990) Social power: a point missed in multi-agent, DAI and HCI. In: Proceedings of the first European workshop on modelling autonomous agents in a multi-agent world, Elsevier, Cambridge, pp 49–62

Casti JL (1997) Would-be worlds: how simulation is changing the frontiers of science. Wiley, New York

Cederman L-E (1997) Emergent actors in world politics: how states and nations develop and dissolve. Princeton University Press, Princeton

Cederman L-E (2001) Agent-based modeling in political science. Pol Methodol 10(1):16–22

Cederman L-E (2003) Modeling the size of wars: from billiard balls to sandpiles. Am Pol Sci Rev 97(1):135–150

Cohen PR (1985) Heuristic reasoning about uncertainty: an artificial intelligence approach. Pitman Advanced Publishing Program, Boston

Cruickshank J (2003) Introduction. In: Cruickshank J (ed) Critical realism: the difference that it makes. Routledge, London/New York, pp 1–14

Dorronsoro G (2005) Revolution unending. Hurst, London

Eckstein H (1992) Case study and theory in political science. In: Gomm R, Hammersley M, Foster P (eds) Case study method: key issues, key texts. Sage, London, pp 119–164

Fuchs C (2005) Machtverhältnisse in Afghanistan: Netzwerkanalyse des Beziehungssystems regionaler Führer. M.A. Thesis, University of Zurich, Zurich

Geller A (2006a) The emergence of individual welfare in Afghanistan. Paper presented at the 20th international political science association world congress, Fukuoka, 9–13 July 2006

Geller A (2006b) Macht, Ressourcen und Gewalt: Zur Komplexität zeitgenössischer Konflikte; Eine agenten-basierte Modellierung. vdf, Zurich

Geller A (2010) The political economy of normlessness in Afghanistan. In: Schlenkhoff A, Oeppen C (eds) Beyond the “wild tribes” – understanding modern Afghanistan and its Diaspora. Columbia University Press, New York, pp 57–70

Geller A, Moss S (2007) The Afghan nexus: anomie, neo-Patrimonialism and the emergence of small-world networks. In: Proceedings of UK social network conference 2007, University of London, 13–14 July 2007, pp 86–88

Geller A, Moss S (2008a) Growing Qawm: an evidence-driven declarative model of Afghan power structures. Adv Complex Syst 11(2):321–335

Geller A, Moss S (2008b) International contemporary conflict as small-world phenomenon and the hermeneutic net. Paper presented at the annual international studies association conference, San Francisco, 26–29 Mar 2008

Gerring J (2004) What is a case study and what is it good for? Am Pol Sci Rev 98(2):341–354

Giddens A (1976) New rules of sociological method: a positive critique of interpretative sociologies. Hutchinson, London

Giustozzi A (2006) Genesis of a “Prince”: the rise of Ismael Khan in Western Afghanistan, 1979–1992, vol 4, Crisis states working papers series. Crisis States Research Centre, London School of Economics, London

Giustozzi A (2007) War and peace economies of Afghanistan’s strongmen. Int Peacekeeping 14(1):75–89

Glatzer B (1998) Is Afghanistan on the brink of ethnic and tribal disintegration? In: Maley W (ed) Fundamentalism reborn? Afghanistan and the Taliban. New York University Press, New York, pp 167–181

Glatzer B (2003) *Afghanistan* (Studien zur Länderbezogenen Konfliktanalyse). Friedrich-Ebert Stiftung/Gesellschaft für Technische Zusammenarbeit

Gomm R, Hammersley M, Foster P (1992) Case study and theory. In: Gomm R, Hammersley M, Foster P (eds) Case study method: key issues, key texts. Sage, London, pp 234–258

Guyot P, Drogoul A, Honiden S (2006a) Power and negotiation: lessons from agent-based participatory simulations. In: Stone P, Weiss G (eds) Proceedings of the fifth international joint conference on autonomous agents and multiagent systems (AAMAS). ACM, New York, pp 27–33

Guyot P, Drogoul A, Honiden S (2006b) Power and negotiation: lessons from agent-based participatory simulations. In: Stone P, Weiss G (eds) Proceedings of the fifth international joint conference on autonomous agents and multiagent systems (AAMAS). ACM Press, New York, pp 27–33

Hirshleifer J (1991) The paradox of power. Econ Polit 3(3):177–200

Hirshleifer J (1995) Anarchy and its breakdown. J Polit Econ 103(1):26–52

Janata A, Hassas R (1975) Ghairatman – Der gute Pashtune: Exkurs über die Grundlagen des Pashtunwali. Afghanistan J 2(2):83–97

Kuznar LA, Frederick W (2007) Simulating the effect of nepotism on political risk taking and social unrest. Comp Math Org Theory 13(1):29–37

Lasswell HD (1936) Who gets what, when, how. Meridian Books, New York

Lave CA, March JG (1975) An introduction to models in the social sciences. Harper & Row, New York

Lazer D (2001) Review of Cederman LE (1997) Emergent actors in world politics: how states and nations develop. J Artif Soc Soc Simul 4(2), http://jasss.soc.surrey.ac.uk/4/2/reviews/lazer.html

Limpert E, Stahel WA, Abbt M (2001) Log-normal distributions across the sciences: keys and clues. Bioscience 51(5):341–352

Lustick IS (2000) Agent-based modeling of collective identity: testing constructivist theory. J Artif Soc Soc Simul 3(1), http://jasss.soc.surrey.ac.uk/3/1/1.html

Mailliard M, Sibertin-Blanc C (2010) What is power? Perspectives from sociology, multi-agent systems and social network analysis. Paper presented at the second symposium on social networks and multiagent systems (SNAMAS 2010), De Montfort University, Leicester, 29 Mar 2010 ftp://ftp.irit.fr/IRIT/SMAC/DOCUMENTS/PUBLIS/SocLab/SNAMAS-10.pdf

Marks SR (1974) Durkheim's theory of anomie. Am J Sociol 80(2):329–363

Medard J-F (1990) L'etat patrimonialise. Politique Africaine 39:25–36

Merton RK (1938) Social structure and anomie. Am Sociol Rev 3(5):672–682

Monsutti A (2004) Cooperation, remittances, and kinship among the Hazaras. Iran Stud 37 (2):219–240

Mosler H-J (2006) Better be convincing or better be stylish? A theory based multi-agent simulation to explain minority influence in groups via arguments or via peripheral cues. J Artif Soc Soc Simul 9(3), http://jasss.soc.surrey.ac.uk/9/3/4.html

Moss S (1981) An economic theory of business strategy: an essay in dynamics without equilibrium. Martin Robertson, Oxford

Moss S (1995) Control metaphors in the modelling of decision-making behaviour. Comp Econ 8(3):283–301

Moss S (1998) Critical incident management: an empirically derived computational model. J Artif Soc Soc Simul 2(4), http://jasss.soc.surrey.ac.uk/1/4/1.html

Moss S (2000) Canonical tasks, environments and models for social simulation. Comp Math Organ Theory 6(3):249–275

Moss S (2001) Game theory: limitations and an alternative. J Artif Soc Soc Simul 4(2), http://jasss.soc.surrey.ac.uk/4/2/2.html

Moss S (2007) Alternative approaches to the empirical validation of agent-based models. Technical report CPM-07-178, Centre for Policy Modelling, Manchester Metropolitan University, Manchester

Moss S, Edmonds B (1997) A knowledge-based model of context-dependent attribute preferences for fast moving consumer goods. Omega Int J Manage Sci 25(2):155–169

Moss S, Edmonds B (2005) Sociology and simulation: statistical and qualitative cross-validation. Am J Sociol 110(4):1095–1131

Moss S, Kuznetsova O (1996) Modelling the process of market emergence. In: Owsinski JW, Nahorski Z (eds) Modelling and analysing economies in transition. MODEST, Warsaw

Moss S, Gaylard H, Wallis S, Edmonds B (1996) SDML: a multi-agent language for organizational modelling. Comp Math Org Theory 4(1):43–69

Mousavi SA (1997) The Hazaras of Afghanistan: an historical, cultural, economic and political study. St. Martin's Press, New York

Neumann FL (1950) Approaches to the study of power. Polit Sci Quart 65(2):161–180

Outhwaite W (1987) New philosophies of social science: realism, hermeneutics and critical theory. Macmillan Education, London

Parsons T (1952) The social system. Tavistock, London

Popitz H (1992) Phänomene der Macht, 2nd edn. Mohr Siebeck, Tübingen

Putnam RD (1988) Diplomacy and domestic politics: the logic of two-level games. Int Organ 42(3):427–460
Rasuly-Paleczek G (1998) Ethnic identity versus nationalism: the Uzbeks of North-Eastern Afghanistan and the Afghan state. In: Atabaki T, O'Kane J (eds) Post-Soviet Central Asia. Tauris Academic Studies, London/New York, pp 204–230
Reno W (1998) Warlord politics and African states. Lynne Rienner, Boulder/London
Richardson LF (1948) Variation of the frequency of fatal quarrels with magnitude. J Am Stat Assoc 43(244):523–546
Rouchier J, Thoyer S (2006) Votes and lobbying in the European decision-making process: application to the European regulation on GMO release. J Artif Soc Soc Simul 9(3), http://jasss.soc.surrey.ac.uk/9/3/1.html
Rouchier J, O'Connor M, Bousquet F (2001) The creation of a reputation in an artificial society organised by a gift system. J Artif Soc Soc Simul 4(2), http://jasss.soc.surrey.ac.uk/4/2/8.html
Roy O (1992) Ethnic identity and political expression in Northern Afghanistan. In: Gross J-A (ed) Muslims in Central Asia: expressions of identity and change. Duke University Press, Durham/London, pp 73–86
Roy O (1994) The new political elite of Afghanistan. In: Weiner M, Banuazizi A (eds) The politics of social transformation in Afghanistan, Iran and Pakistan. Syracuse University Press, Syracuse, pp 72–100
Roy O (1995) Afghanistan: from holy war to civil war. The Darwin Press, Princeton
Roy O (1998) Has Islamism a future in Afghanistan? In: Maley W (ed) Fundamentalism reborn? Afghanistan and the Taliban. New York University Press, New York, pp 199–211
Rubin BR (1992) Political elites in Afghanistan: Rentier State Building, Rentier State Wrecking. Int J Middle East Stud 24(1):77–99
Rubin BR (2007) Saving Afghanistan. Foreign Aff 86(1):57–78
Saam NJ, Harrer A (1999) Simulating norms, social inequality, and functional change in artificial societies. J Artif Soc Soc Simul 2(1), http://jasss.soc.surrey.ac.uk/2/1/2.html
Sambanis N (2004) Using case studies to expand economic models of civil war. Perspect Polit 2(2):259–279
Sawyer RK (2005) Social emergence, societies as complex systems. Cambridge University Press, Cambridge
Sayer A (1992) Method in social science: a realist approach, 2nd edn. Routledge, London/New York
Sayer A (2000) Realism and social science. Sage, London/Thousand Oaks/New Delhi
Schelling TC (1998) Social mechanisms and social dynamics. In: Hedström P, Swedberg R (eds) Social mechanism: an analytical approach to social theory. Cambridge University Press, Cambridge, pp 32–44
Schetter C, Glassner R, Karokhail M (2007) Beyond warlordism: the local security architecture in Afghanistan. Internationale Politik und Gesellschaft 2:136–152
Shahrani MN (1998) The future of state and the structure of community governance in Afghanistan. In: Maley W (ed) Fundamentalism reborn? Afghanistan and the Taliban. New York University Press, New York, pp 212–242
Shahrani MN (2002) War, factionalism, and the state in Afghanistan. Am Anthropol 104 (3):715–722
Shapiro I (2005) The flight from reality in the human sciences. Princeton University Press, Princeton
Silverman D (ed) (2004) Qualitative research: theory, method and practice, 2nd edn. Sage, London
Snidal D (1985) The game theory of international politics. World Pol 38(1):25–57
Sofsky W (2002) Zeiten des Schreckens: Amok, Terror, Krieg. S. Fischer, Frankfurt a.M
Stachowiak M (1973) Allgemeine Modelltheorie. Springer, Wien/New York
Stakes RE (1978) The case study method in social inquiry. Edu Res 7(2):5–8
Tapper N (1991) Bartered brides: politics, gender and marriage in an Afghan tribal society. Cambridge University Press, Cambridge

Tapper R (2008) Who are the Kuchi? Nomad self-identities in Afghanistan. J R Anthropol Inst 14:97–116

Tarzi SM (1993) Afghanistan in 1992: a Hobbesian state of nature. Asian Surv 33(2):165–174

UNODC (2006) Afghanistan's drug industry: structure, functioning, dynamics, and implications for counter-narcotics policy. United Nations Office on Drugs and Crime/The World Bank

Weber M (1980) Wirtschaft und Gesellschaft: Grundriss der verstehenden Soziologie, 5th edn. J.C.B. Mohr (Paul Siebeck), Tübingen

Weidmann N, Cederman L-E (2005) Geocontest: modeling strategic competition in geopolitical systems. In: Troitzsch K (ed) Proceedings of the European social simulation association annual conference (ESSA 2005). Koblenz, Germany, pp 179–185

Wily LA (2004) Looking for peace on the Pastures: rural land relations in Afghanistan. Afghanistan Research and Evaluation Unit, Kabul

Younger S (2005) Violence and revenge in Egalitarian societies. J Artif Soc Soc Simul 8(4), http://jasss.soc.surrey.ac.uk/8/4/11.html

# Chapter 26
# Human Societies: Understanding Observed Social Phenomena

**Bruce Edmonds, Pablo Lucas, Juliette Rouchier, and Richard Taylor**

**Why Read This Chapter?** To get an overview of the different ways in which simulation can be used to gain understanding of human societies and to gain insight into some of the principle issues that impinge upon such simulation, including the difficulties these cause. The chapter will go through the various specific goals one might have in doing such simulation, giving examples of each. It will provide a critical view as to the success at reaching these various goals and hence inform about the current state of such simulation projects.

**Abstract** The chapter begins by briefly describing two contrasting simulations: the iconic system dynamics model publicised under the "Limits to Growth" book and a detailed model of 1st millennium Native American societies in the south west of the US. These are used to bring out the issues of: abstraction, replicability, model comprehensibility, understanding vs. prediction, and the extent to which simulations go beyond what is observed. These issues and difficulties result in three "dimensions" in which simulation approaches differ. These issues are each rooted in some fundamental difficulties in the project of simulating observed societies that are then briefly discussed. The core of the chapter is a look at 15 different possible simulation goals, both abstract and concrete, giving some examples of each and discussing them. The different inputs and results from such

B. Edmonds (✉)
Centre for Policy Modelling, Manchester Metropolitan University
e-mail: bruce@edmonds.name

P. Lucas
Geary Institute, University College Dublin
e-mail: pablo.lucas@ucd.ie

J. Rouchier
School of Economics, Aix-Marseille University
e-mail: juliette.rouchier@univ-amu.fr

R. Taylor
Oxford Branch, Stockholm Environment Institute
e-mail: richard.taylor@sei-international.org

B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2_26,

simulations are briefly discussed as to their importance for simulating human societies. The chapter ends with a brief critical assessment of the state of the art in terms of using simulation techniques for helping to understand human societies.

## 26.1 Introduction

Understanding social phenomena is hard. There is all the complexity found in other fields of enquiry but with additional difficulties due to our embedding as part of what we are studying.[1] Despite these, understanding our own nature is naturally important to us, and our social aspects are a large part of that nature. Indeed some would go as far as saying that our social abilities are *the* defining features of our species (e.g. Dunbar 1998). The project of understanding human societies is so intricate that we need to deploy any and all means at our disposal. Simulation is but one tool in this vast project, but it has the potential to play an important part.

This chapter considers how and to what extent computer simulation helps us to understand the social complexity we see all around us. It will start by discussing two simulations in order to raise the key issues that this project involves, before moving on to highlight the difficulties of understanding human society in more detail. The core of the chapter is a review of some of the different ways that simulation can be used for, with examples of each. It then briefly discusses a conceptual framework, which will be used to organise the different ways that simulations are used. It ends with a frank assessment of the actual progress in understanding human societies using simulation techniques.

### *26.1.1 Example 1: The Club of Rome's "Limits to Growth" (LTG)*

In the early 1970s, on behalf of an international group under the name "The Club of Rome" a simulation study was published (Meadows et al. 1972) with the attempt to convince humankind that there were some serious issues facing it, in terms of a coming population, resource and pollution catastrophe. To do this they developed a system-dynamics model of the world. They chose a system-dynamics model because they felt they needed to capture some of the feedback cycles between the key factors – factors that would not come out in simple statistical projections of the available data. They developed this model and ran it, publishing the findings – a number of model-generated future scenarios – for a variety of settings and variations. The book ("Limits to Growth") considered the world as a single system, and postulated some relationships between macro variables, such as population, available resources, pollution etc. Based on these relations it simulated what might happen if the feedback between the various variables were allowed to occur.

[1] This embedding has advantages as well, such as prior knowledge.

The results of the simulations were the curves that resulted from this model as the simulation was continued for future dates. The results indicated that there was a coming critical point in time and that a lot of suffering would result, even if mankind managed to survive it.

The book had a considerable impact, firmly placing the possibility that humankind could not simply continue to grow indefinitely. It also attracted considerable criticism (e.g. Cole et al. 1973) mainly based on the plausibility of the model's assumptions and the sensitivity of its results to those relationships. (For example it assumed that growth will be exponential and that delay loops are extended). The book presented the results of the simulations as predictions – a series of what-if scenarios. Whilst they did add caveats and explore various possible versions of their model, depending on what connections there turned out to be in the world-system, the overall message of the book was unmistakeable: that if we did not change what we were doing, by limiting our own economic consumption and population, disaster would result. This was a work firmly in the tradition of Malthus (1798) who, 175 years earlier, had predicted a constant state of near-starvation for much of the world based upon a consideration of the growth processes of population and agriculture.

The authors clearly hoped that by using a simulation (albeit a simplistic one by present standards) they would be able to make the potential feedback loops real to people. Thus this was a use of simulation to illustrate an understanding that the authors of LTG had. However, the model was not presented as such, but as something *more scientific* in some sense.[2] A science-driven study that predicted such suffering was a definite challenge to those who thought the problem was less severe.[3] By publishing their model and making it easy for others to replicate and analyse it, they offered critics a good opportunity for counter-argumentation.

The model was criticised on many different grounds, but the most effective was that the model was sensitive to the initial settings of some parameters (Vermeulen and de Jongh 1976). This raised the question of whether the model had to be finely tuned in order to get the behaviour claimed and thus, since the parameters were highly abstract and did not directly correspond to anything measurable, the applicability of the model to the world we live in was questioned. Its critics assumed that since this model did not, hence, produce reliable predictions that it could be safely ignored. It also engendered the general perception that predictive simulation models are not credible tools for understanding human socio-economic changes – especially for long term analyses – and discouraged their use in supporting policy-making for a while.

---

[2] The intentions of the authors themselves in terms of what they thought of the simulation itself are difficult to ascertain and varied between the individuals, however this was certainly how the work was perceived.

[3] Or those whose vested interests may have led them to maintain the *status quo* concerning the desirability of continual economic growth.

### 26.1.2 Example 2: Modelling 1st Millennium Native American Society

A contrasting example to the Club of Rome model is the use of simulation models to assess and explore explanations of population shifts among the Native American nations in the pre-Columbian era. This has been called "Generative Archaeology" (GA) by Tim Kohler (2009). Here a spatial model of a population was developed where movement was fitted to a wealth of archaeological and climatologically data, in order to find and assess possible explanations of the size, distribution and change in populations that existed in the first millennium AD in the Southwest US. This case offers a picture of settlement patterns in the context of relatively high-resolution reconstructions of changes in climate and resources relevant to human use of these landscapes.

The available data in this case is relatively rich, allowing many questions to be answered directly. However, many interesting aspects are not directly answerable from a static analysis of the data, for example those about the possible social processes that existed. The problem is that different archaeologists can inspect the same settlement pattern and generate different candidate processes (explanations) for its generation. Here agent-based modelling helps infer the social processes (which cannot be directly observed) from the detailed record over time. This is not a direct or certain inference, since there are still many unknowns involved in that process.

In Kohler et al. (2005, 2008)[4] the use of agent-based modelling (ABM) has been mainly to see what patterns one should expect if households were approximately minimizing their caloric costs for access to adequate amounts of calories, protein, and water. The differences through time in how well this expectation fits the observed record and the changing directions of departure from those expectations provide a completely novel source of inference on the archaeological record. Simulations using the hypothesis of local food sharing during periods of mild food shortage may be compared to the fit in a simulation where food sharing does not occur. In this way we can get indirect evidence as to whether food sharing took place.

The ABM has hence allowed a comparison of a possible process with the recorded evidence. This comparison is relative to the assumptions that are built into the model, which tend to be plausible but questionable. However despite the uncertainties involved, one is able to make *a useful* assessment of the possible explanations and the assumptions are explicitly documented. This approach to processes that involve complex interaction would be impossible to do without a computer simulation. At the very least, such a process reveals new important questions to ask (and hence new evidence to search for) and the times when the plausible explanations are demonstrably inadequate. However for any real progress in explanation of such cases, a very large amount of data seems to have been required.

[4] For details of the wider project connected with these papers, see the Village Ecodynamics Project, http://village.anth.wsu.edu.

### 26.1.3 Some Issues That the Aforementioned Examples Illustrate

The previous examples raise a few issues, common to much social simulation modelling of human societies. These will now be briefly defined and discussed as an introduction to the problem of understanding social phenomena using simulation.

1. ***Abstraction***. Abstraction[5] is a crucial step in modelling observed social phenomena, as it involves choices about what aspects are or are not salient in relation to the problem and what level of analysis is appropriate. The LTG example, being a macro-model, assumes that distributive aspects such as geography and local heterogeneity are less important with respect to feedbacks among global growth variables. In this model all the detail of a whole world is reduced to the interaction of a few numeric variables. The GA model was more specific and detailed, including an explicit 2D map of the area and the position of settlements at different times in the past. It is fair to say that the LTG model was driven by the goals of its modellers, i.e. showing that the coming crisis could be sharp due to slow feedback loops, the GA model is more driven by the data that was available, with the model being used to try and answer a number of different questions afterwards.
2. ***Replicability***. Replicability is the extent to which a published model has been described in a sufficiently comprehensive and transparent way so that the simulation experiments can be independently reproduced by another modeller. Replicability may be considerably easier if care is taken to verify the initial model by means of debugging tests, and also if the original source code is effectively available and is well commented. Here the LTG model was readily replicable and hence open to inspection and criticism. The GA models are available to download and inspect, but its very complexity makes it hard to *independently* replicate (although it has been done).
3. ***Understanding the model***. A modeller's inferential ability is the extent to which one can understand one's own model. Evidence suggests that humans can fully track systems only for about two or three variables and five or six states (Klein 1998); for higher levels of complexity, additional tools are required. In many simulations, especially those towards the descriptive end of the spectrum, the agents have many behavioural rules which may interact in complicated and unpredictable ways. This makes simulations very difficult to fully understand and check. Even in the case of a simple model, such as the LTG model, there can be unexpected features (such as the fine tuning that was required). Although the GA was rich in its depiction of the space and environment the behavioural rules of sub-populations was fairly simple and easy to follow at the micro-level. However this does not rule out subtle errors and complexities that might result from the interaction of the micro-elements. Indeed this is the point of such a

[5] Abstraction can be defined as "ignoring or hiding details to capture some kind of commonality between different instances" (Free On-line Dictionary of Computing).

simulation that we can't work out these complex outcomes ourselves, but require a computer program to do it.

4. ***Prediction* vs. *Understanding***. The main lesson to be drawn from the history of formal modelling is that, for most complex systems it is impossible to model with accuracy their evolution beyond an immediate timeframe. Whilst the broad trends and properties may be to some degree forecast the particulars, e.g. the timing and scale of changes in the aggregate variables, generally they cannot (Moss 1999). The LTG model attempted to forecast the future, not in terms of the precise levels but in the presence of a severe crisis – a peak in population followed by a crash. The GA does not aim to strongly predict anything, but rather it seeks to establish plausible explanations for the data that is known. Most simulations of human society restrict themselves to establishing explanations, the simulations providing a chain of causation that shows that the explanation is possible.[6]
5. ***Going beyond what is known***. In social science there are a multitude of gaps in our knowledge and social simulation methods may be well placed to address some of these gaps. Given some data, and some plausible assumptions, the simulations can be used to perform experiments that are consistent with the data and assumptions, and then inspected to answer other questions. Clearly this depends on the reliability of the assumptions chosen. In the GA case this is very clear, a model with a food-sharing rule and one without can be compared to the data, seeing which one fits it better. The LTG model attempts something harder, making severe assumptions about how the aggregate variables relate, it "predicts" aspects of the future. In general: the more reliable the assumptions and data (hence: the less ambitious the attempt at projection), the more credible the result.

A social scientist who wants to capture key aspects of observed social phenomena in a simulation model faces many difficulties. Indeed, the differences between formal systems and complex, multi-facetted and meaning-laden social systems are so fundamental that some criticise *any* attempt to bridge this gap (e.g. Clifford 1986). Simulators have to face these difficulties which have an impact as to how social simulation is done and how useful (or otherwise) such models may be. We briefly consider six of these difficulties here.

- Firstly, there is the sheer difference in nature between formal models (i.e. computer programs) that modellers use as compared to the social world that we observe. The former are explicit, precise, with a formal grammar, predictable at the micro-level, reproducible and work in (mostly) the same way regardless of the computational context. The later is vague, fluid, uncertain, subjective, implicit and imprecise – which often seems to work completely differently in similar situations, and whose operation seems to rely on the rich interaction of meaning in a way that is sometimes explicable but usually unpredictable. In particular the

[6] Although in many cases this is dressed up to look like prediction, such as the fitting to out-of-sample data. Prediction has to be for data unknown to the modeller, otherwise the model will be implicitly fitted to it.

gap between essentially formal symbols with precise but limited meaning and the rich semantic associations of the observed social world (for example as expressed in natural language) is particularly stark. This gap is so wide that some philosophers have declared it unbridgeable (e.g. Lincoln and Guba 1985, Guba and Lincoln 1994).

- Secondly there is the sheer variability, complication and complexity of the social world. Social phenomena seem to be at least as complex as biological phenomena but without the central organising principle of evolution as specified in the neo-Darwinian Synthesis. If there are any general organising principles (and it is not obvious that this is the case) then there are many of these, each with differing (and sometimes overlapping) domains of application. In that sense, it is clear that a model will always capture only a small part of the phenomenon among many other related aspects, hence reducing drastically the possibility to predict with any degree of certainty.
- Then there is the sheer lack of adequate multifaceted data about social phenomena. Social simulators always seem to have to choose between longitudinal studies, narrative data, cross-sectional surveys or time-series data. Having all of these datasets about a single social process or event is to date very unlikely. There does not seem to be the emphasis on data collection and measurement in the social sciences that there is in some other sciences and certainly not the corresponding prestige for those who collect it or invent ways of doing so.
- There is the more mundane difficulty of building, checking, maintaining, and analysing simulations (Galán et al. 2009). Even the simplest simulations are beyond our complete understanding, indeed that is often why we need them, because there is no other practical way to find out the complex ramifications of a set of interacting agents. This presence of emergent outcomes in the simulations makes them very difficult to check. Ways to improve confidence that our simulations in fact correspond to our intentions for them[7] include: unit testing, debugging, and the facility for querying the database of a simulation (see Chap. 8 (David 2013) in this handbook). Perhaps the strictest test is the independent replication of simulations – working from the specifications and checking their results at a high degree of accuracy (Axtell et al. 1996). However such replication is usually very difficult and time-consuming, even in relatively simple cases (Edmonds and Hales 2003).
- Another difficulty is that of the inevitability of background assumptions in all we do. There are always a wealth of facts, processes and affordances to give meaning to, and provide the framework for, the foreground actions and causal chains that we observe. Many of these are not immediately apparent to us since they are part of the contexts we inhabit and so are not perceptually apparent. This is the same as other fields, as it has been argued elsewhere; the concept of causation only makes sense within a context (Edmonds 2007). However it does

[7] In terms of design and implementation, if one has a good reference case in terms of observed data then you can also check one's simulation against this.

seem that context is more critical in the social world than others, since it can not only change the outcomes of events but their very meaning (and hence the *kind* of social outcome). Whilst in other fields it might be acceptable to represent extra-contextual interferences as some kind of random distribution or process, this is often manifestly inadequate with social phenomena (Edmonds and Hales 2005).

- The uncertainty behind the foreground assumptions in social simulation is also problematic. Even when we are aware of all of the assumptions they are often either too numerous to include in a single model or else we simply lack any evidence as to what they should be. For example, there are many social simulation models which include some version of inference, learning or decision-making within the agents of the model, even when there is no evidence as to whether this actually corresponds to that used by the observed actors. It seems that often it is simply hoped that these details will not happen to matter much in the end – thus becoming a rarely checked, and sometimes wrong, aspect of simulations (Edmonds 2001; Rouchier 2001).
- Finally there is a difficulty from the nature of simulation itself. Simulation will demonstrate *possible* processes that might follow from a given situation (relative to the assumptions on which the simulation is built). It does not show all the possibilities, since it could happen that a future simulation will produce the same outcomes from the same set-up in a different way (e.g. using a different cognitive model). Thus simulation differs in terms of its inferential power from analytic models (e.g. equation-based ones), where the simplicity of the model can allow formal proofs of a general formulation of outcomes that may establish the necessity of conditions as well as their adequacy. This difficulty is the same that plagues many mathematical formulations since, in their raw form, they are often unsolvable and hence either one has to use numerical simulation of results (in which case one is back to a simulation) or one has to make simplifying assumptions (in which case, depending on the strength of these assumptions, one does not know if the results still apply to the original case).

These difficulties bring up the question of whether some aspects of societies can be at all understood by means of modelling. This hypothesis asserting that simulation is a credible method to better explore, understand or explain social processes, is implicitly tested in the current volume and is discussed in some detail below. We are not going to take any strong position but will restrict ourselves to considering examples within the context of their use.[8] Agent-based social simulation is not a magic-bullet and is not yet a mature technique. It is common sense in the social simulation community that best results will be achieved by combining social simulation with other research methods.

[8] Obviously we *suspect* it can be a useful tool otherwise we would not be bothering with it.

## 26.2 Styles of Modelling and Their Impact on Simulation Issues

### *26.2.1 Models of Evidence vs. Models of Ideas*

One response to the above difficulties is not to model social phenomena directly, but rather to restrict ourselves to modelling *ideas* about social phenomena. This is a lot easier, since our ideas are necessarily a lot simpler and more abstract than the phenomena itself (which can be formalized with the notion of pattern modelling (Grimm et al. 2005) rather than strict adequacy to data). Some ideas need modelling, in the sense that the ramifications of the ideas are themselves complex. These kinds of models can be used to improve our understanding of the ideas, and later this understanding can be applied in a rich, flexible and context-sensitive way. This distinction is made clear in (Edmonds 2001).

Of course to some extent any model is a compact abstraction of the final target of modelling – there will, presumably, be *some* reason why one conceptualises what one is modelling in terms of evidence or experience by someone, and there will always be *some* level of theory/assumption that motivates the decision as to what can be safely left out of a model. Thus all models are somewhat about ideas and, hopefully, all models have *some* relation to the evidence. However there is still a clear difference between those models that take their primary structure from an idea, and those whose primary considerations come from the available evidence. For example, the former tend to be a lot simpler than the later. The later will tend to have specific motivations for each feature whilst the former will tend to be motivated in general terms. These two kinds of simulation have a close parallel with the theoretical and phenomenological models identified by Cartwright (1993).

Unfortunately these kinds of model are often conflated in academic papers. This seems frequently not deliberate, but rather due to the strong theoretical spectacles (Kuhn 1962) that simulation models seem to provide. There is nothing like developing and playing with a simulation model for a while to make one *see* the world in terms of that model – it is not only that the model is your creation and best effort in formulating an aspect of the social world, but one has interacted with it and changed it to include the features that you, the modeller, think it should have. Nevertheless, whatever the source it can take some careful “reading between the lines” to determine the exact nature of any model, and what it purports to represent.

### *26.2.2 Modelling as Representation of Social Phenomena vs. as an Intervention in a Social Process*

It must be said that some simulation models are not intended to *represent* anything but rather created for another purpose, such as a tool for demonstrating an approach or an intervention in a decision-making process. This may be deliberate and explicit, or not, for various different reasons. Of course if a computer model does not represent anything at all, it is not really a simulation but simply a computer program, which may be presented in the style of a simulation. Also for a simulation to be an effective tool for intervention it will have to have some credibility with the participants.

However in some research the representation is either not the primary goal or what they seek to represent is deliberately subjective in character. Thus in some participatory approaches (see Chap. 11, Barreteau et al. 2013) it may be the primary goal to raise awareness of an issue, to intervene in or facilitate a social process like a negotiation or consensus process within a group of people. The modeller may not focus as much on whether the model captures an objective reality but rather on how stakeholders[9] understood the issues and processes of concern and how this might influence the outcomes. This does not mean that there will be *no* elements that are objective and/or representative in character – for example such models might have a well-validated hydrological component to them – but that the parts of the model that are the focus are checked against the opinions of those being modelled or those with an interest in the outcomes rather than any independent evidence.

Of course, this is a matter of degree – in a sense most social simulations are a mixture of objective aspects linked to observations and other aspects derived from theories, opinions, hypotheses and assumptions. In the participatory approaches the modeller seeks not to put their own ideas forward but rather take the, possibly more democratic, approach of being an expert facilitator expressing the stakeholders' opinions and knowledge. Whilst some researchers might reject such ideas as too "anecdotal" to be included in a formal model, it is not obvious that the stakeholders' ideas about the nature of the processes involved (for example, how the key players make decisions) are less reliable than the grander theories of academics. However researchers do have a professional obligation to be transparent and honest about their opinions, documenting assumptions to make them explicit and, *at least*, not state things that they think are false. Thus, although participatory approaches are not a world away from more traditional models of using simulation, they do have some different biases and characteristics.

### *26.2.3 Context and Social Simulation*

Human knowledge, but particularly human *social* knowledge is usually not context-free. That is there is a set of background assumptions, facts, relationships and meanings that may be necessary and generally known but not made explicit. These background features can all be associated with the context of the knowledge (Edmonds 1999). In a similar way, most social simulation happens *within* a particular context as given, thus for example the environment in which racial

[9] I.e. those who are part of, or can influence the social phenomenon in question.

segregation occurs might be obvious to all concerned. This context is sometimes indicated in papers but is often left implicit. The context of a simulation is associated with the uncountably many background assumptions that can be ignored, either because they are irrelevant or fixed in that context. Social simulation would probably be impossible if one was not able to assume a context whose associated assumptions need not be questioned for a given model (Edmonds 2010). Without such an effective restriction of scope every social simulation model would have to include *all* potential aspects of human behaviour and social interaction. Whilst such assumptions concerning the context are common to almost all fields of knowledge it is particularly powerful in the social sciences due to the fact that we unavoidably use our folk-knowledge[10] of social situations to make sense of the studied social phenomena. This process of (social) context identification is often automatic, so that we correctly identify the appropriate context without expending much conscious thought. For this reason the context is often left implicit, despite the fact that it can impact upon the use of simulation in understanding the phenomena. This leaves the decisions as what to implement as foreground, deliberate decisions.

Choosing a social context which is relatively identifiable and self-contained is an important factor if one is seeking to represent some evidence in a simulation. Being able to include all the important factors of some social process and obtain some evidence for their nature, allows the building of simulations that are not misleading in the sense of not missing out factors that might critically change the outcomes. Clearly the more restricted the context, the easier the representational task. However in this case one does not know whether what one learns from the simulation is applicable in other contexts. Using a simulation developed for one context and purpose for a different context and/or purpose might well lead to misleading conclusions (Edmonds and Hales 2005; Lucas 2010).

Those simulations which are more focused on exploring an idea will often seek to transcend context, in the hope that the models will have some degree of generality – these often deliberately ignore any particular context. Although these may seem general, their weakness can become apparent when its applicability is tested. Here the *ideas* they represent might give some useful insights, but may be misleading if taken as *the* defining feature of a specific case study. Clearly a simulation that is claimed to have general applicability needs to have been validated across the claimed scope before being relied upon by others. To date, *no* social simulation has been found to be generally applicable beyond theoretical and illustrational purposes (Lucas 2011).

[10] Folk-knowledge is the set of widely-held beliefs about popular psychological and social theories, this is sometimes used in a rather derogatory way even when the reliability of the academic alternatives is unclear.

## 26.3 A Plethora of Modelling Purposes with Examples

Given the different purposes for which simulation models are used (Epstein 2008), they will be considered in groups of those with similar goals. It is only relative to their goals that simulation efforts can be judged. Nowadays it is widely acknowledged that authors should state the purpose of their models clearly before how the model is constituted (Grimm et al. 2006). Firstly however it is worth reviewing two goals that are widely pursued in many other fields but have not been convincingly attained with respect to the simulation of human society.

The first of these goals is that of predicting what will definitely happen in unknown circumstances. In other words, social simulation cannot yet make accurate and precise predictions. The nearest social simulations come (to our knowledge) is predicting some outcomes in situations where the choices are very constricted, and the data available is comprehensive. The clearest case of this is the use of micro-simulation models to predict the final outcome of elections once about 30 % of the results are already known (Curtis and Frith 2008). Thus this is hardly new or unknown circumstances, and is not immune from surprises, since sometimes their predictions are wrong. This model is a micro-simulation model that relies on the balance between parties in each constituency and then translates the general switches between parties (and non-voters) to the undeclared results. Thus although it is a prediction, its very nature rules out counter-intuitive or surprising predictions and comes more into the category of extending known data rather than prediction. The gold standard for prediction is that of making predictions of outcomes that are unexpected but true.[11]

The second goal simulations do not achieve is to decisively *test* sociological hypotheses – in other words, they do not convincingly show that any particular idea about what we observe occurring in human societies can be relied upon or comprehensively ruled out. Here the distinction between modelling what we observe and modelling our ideas is important. A simulation that attempts to model what we observe is a contingent hypothesis that may always be wrong. However social simulations of evidence are always dependent on a raft of supportive assumptions – that the simulation fails to reproduce the desired outcomes may be due to a failure of any of its assumptions. Of course, if such a model is repeatedly tested against evidence and fails to be proved wrong we may come to rely upon it more (Popper 1963) but this success may be for other reasons (e.g. we simply have not tested it in sufficiently diverse conditions). Hypothesis testing using simulations is always relative to the totality of assumptions in the simulations and thus the gain in certainty is, *at best*, incremental and relative.[12] Thus the core assumptions of a field may be preserved by adjusting "auxiliary" aspects (Lakatos and Musgrave 1970).

---

[11] This is when prediction is actually useful, for if it only gives expected values one would not need the simulation.

[12] If a simulation is not directly related to evidence but is more a model of some ideas, then it might be simple enough to be able to test hypotheses but these hypotheses will then be about the abstract model and not about the target phenomena.

If a simulation is about ideas then a very restricted kind of test is possible: a counter example. If it has been assumed that factor *A* will lead to result *B*, then one might be able to show that this might not be the case in a plausible simulation. Indeed the simulation may show that to obtain result *B* from factor *A* an additional and implausible assumption *C* is necessary. This does *prove* that "it is not necessarily the case that *A* leads to *B*", but it may shift the burden of proof back onto those who have assumed *A* will lead to *B*. This very restricted test is only useful if the context of causation between *A* and *B* is appropriately identifiable. This case of using a simulation to establish counter-examples is considered below.

A particular case of seeking for counter-examples is that of testing for the "existence of a sufficient condition" for some particular results. For example, it may be possible to show that there is no need to add some particular hypothesis to see a phenomenon take place, as in economics where it can be shown that in many cases the assumption of perfect rationality for agents does not need to be made.[13]

One might be disappointed that simulation provides neither predictions nor proofs (in the stronger senses of those terms), but that does not stop them being useful in other ways, which the sections below illustrate.

In the following, we look at how simulations might contribute to the understanding of human societies in a number of different ways, with examples from the literature. Unfortunately many articles describing social simulation research do not make their goals explicit (as advocated by ODD, see Polhill et al. 2008 and Chap. 7, Grimm et al. 2013), therefore the categorisation below is that of the chapter's authors and not necessarily the category that the authors of the papers discussed would choose. Also it appears that some researchers have multiple purposes for their simulations or simply have not thought about their goals clearly.

## *26.3.1 Abstract Goals*

First we consider simulations that have more abstract goals i.e. these tend to be more about *ideas* and *theories* than observed evidence (as discussed above in Sect. 26.2.1).

### 26.3.1.1 Illustration of Ideas

Simulations can be good ways of making processes clear, because simulations are what simulations *do*. Thus, if one can unpack a simulation to show *how* the outcomes result from a setup or mechanism, then this can demonstrate an idea clearly and dramatically. Of course, if how the outcomes emerge from the setup in a

[13] This fact has lead some to argue that such assumptions of perfect rationality should be dropped and that it might be better to adopt a more naturalistic representation of human's cognition (Gode and Sunder 1993; Kirman 2011).

simulation is opaque and/or difficult to understand then this is not an effective technique. For this reason this tends to be done using relatively simple simulations that are specifically designed to bring out the focus idea.

An example is (Rouchier and Thoyer 2006) which models voting and lobbying in the EU decision making process. It does make fairly strong assumptions about how the voting strategies might operate, but it does not pretend to be a descriptive model. Instead it makes clear how the links between public opinion, lobbying groups and elected representatives might operate at the national scale as well as the EU one.

Another example is (Gode and Sunder 1993), a fairly simple demonstration that in some cases market institutions are so constraining that agents do not even need to be clever to achieve excellent results in this setting. They take the example of Continuous Double Auction (CDA), a two-sided progressive auction, which is the protocol that is most used in financial markets. At any moment, buyers can submit *bids* (offers to buy). Similarly, sellers can submit *asks* (offers to sell). Both buyers and sellers may propose an offer or accept the offer made by others. The main constraint is an improvement rule, imposed on new offers entering the market, which requires submitted bids/asks at a price higher/lower than the standing bids/asks. Each time an offer is satisfied for one of the participants, he or she announces the acceptance of the trade at the given price, and the transaction is completed. Once a transaction is completed the agents who have traded leave the market and the bid-offer process starts again following the same rule starting from any price. The result of Gode and Sunder's simulation is that even with completely stupid agents, who know nothing of the market and only follow two constraints: the bid-offer rule described above and not selling below or buying above their reservation price, the market converges and enables agents to get excellent profits. This paper shows how institutional constraints might act to ensure a reasonable allocation of goods when agents are very clear about the value of things they want to sell or buy, and that this does not *require* any other substantive rationality by the agents. Of course this result cannot necessarily be extended to any observed markets, which are most of the time complex, where the agents do have intelligence, where the value of items might be unclear and where there might be many other social and institutional mechanisms, but at least this result clarifies an idea about why protocols of this kind might be important.

The OpenABM project[14] has made significant progress in the development of a community of people using illustrative models to facilitate the communication of ideas. Working with others this group in particular promotes the educational value of agent-based models.

A particular case of using a simulation to illustrate an idea is that of using a simulation in teaching. Whilst demonstrating an idea to one's peers might lead one to choose a simulation that emphasises the idea's generality and power, in teaching one may well choose to simplify and highlight certain features of the idea that will

[14] http://www.openabm.org.

be important later on. This is a matter of degree, but tends to result in simulations of a slightly different kind.

For example, researchers at Oxford University Department of Computer Science have developed a web application to assist students (particularly non-programmers) in understanding the behaviour of systems of interactive agents (Kahn and Noble 2009). They model, for example, the dynamics of epidemics in schools and workplaces, and effect of vaccination or school closing/quarantine periods upon spread of disease in the population (Scherer and McLean 2002). The students can quickly and easily test different policies and other parameter combinations, or for more intensive sessions can work through a series of guided steps to build models from pre-existing modular components or 'micro-behaviours' – a process called 'composing'. The models can also be run, saved, and shared through a web browser in order to facilitate discussion and collaboration as well as ownership of the ideas and creative thinking.

#### 26.3.1.2 Establishing the Possibility of a Process

A simulation can be used to show *how* a mechanism might result in certain outcomes, and thus established that a proposed process is possible, demonstrated by enfolding micro-processes in the simulation. This established plausibility of the process is relative to the plausibility of the assumptions behind the simulation – clearly if the simulation is one that could not convincingly be related to any observed system then one would not have established that the process is possible in any encountered system, but only be a *theoretical* possibility. This does not require that the simulation is an accurate representation of any observed system since all that is required is that one could imagine that a version of the target process in the simulation *could* occur in a real system.

A classic example of this is Axelrod's (1984, 1997) work on the evolution of cooperation. Previous models in evolutionary biology had suggested that cooperative behaviour would not be selected within an evolutionary setting, as any group of co-operators would be vulnerable to a single non-cooperative invader or mutant. Axelrod's books described simulations in which a population of competing individuals evolved, playing repeated games against others. Some cooperative strategies, in particular 'tit-for-tat' (cooperate unless your partner did not last time) where shown to survive and flourish in many game set-ups. Although the simulations described were highly speculative and abstract, they did firmly establish that it was possible that cooperative strategies might evolve within an evolutionary setting, where selfish strategies had a short-term advantage.

One use for establishing the possibility of a process is as a *counter-example* to an existing assumption or accepted theory, if the process demonstrated contradicts the assumption. Thus the simulations of Axelrod above can be seen as a counter-example to the assumption that cooperative behaviour cannot survive in an evolutionary setting.

The particular case of the Schelling (1969, 1971) model can be classified in this trend. Through very simple simulations, which Schelling ran by hand at the time, he discovered that segregation could be attained at a group level although each individual agent had no strong preference for segregation. This paper was important, because it was one of the first examples of emergent phenomena applied to social issues. But the most important element was the positive result obtained with the model. Schelling used a very intuitive (though not necessarily realistic) way of describing the change of location of agents in a city where they are surrounded by neighbours, which can be of two different types: identical to themselves or different. Each agent decides if it is satisfied with its location by judging if the proportion of neighbours that are different is acceptable to it. If this is not the case it moves to a new location. Even when each agent accepts up to 65 % of agents different to itself in its neighbourhood, high levels of segregation in the global society of agents result. This is a counter-example to the assumption that segregation results from a high level of intolerance to those of different ethnic origins, since one can see from the simulation that the apparition of high levels of segregation in cities could be due to the movement of people at the edges of segregated areas who are in regions dominated by those of different ethnicities. Of course, what this does not show is that this is what actually causes segregation in cities, it merely undermines the assumption that it *must* be due to high levels of intolerance.

#### 26.3.1.3 Understanding the Properties of an Abstract Model

With some analytic mathematical models and very few, very simple simulation models, one might seek to *prove* some properties of that model, for example the parameter values under which a given end condition is reached. If this is not possible (the usual case), then one has two basic options: to simplify the original to obtain a model that is analytically tractable or to simulate it. If the simplifications that are necessary to obtain a tractable model are well-understood and plausible, then the simplified model might be trusted to approximate the original model (although it is always wise to check). If it is the case that to obtain an analytically tractable model one has to simplify *so much* that the relationship between the simplified and the original model is suspect (for example, by adding implausibly strong assumptions) then one cannot say that the simplified model is *about* the same things as the original model. At best the simplified model might be used as an analogy for what was being modelled – it cannot be relied upon to give correct answers about the original target of modelling. In this case, if one wants to actually model the original target of modelling, then simulation models are the only option. In this case one might wish to understand the simulation itself by systematically exploring its properties, such as doing parameter sweeps. In a sense this is a kind of pseudo-maths, trying to get a grasp of the general model properties when analytic proof is not feasible.

An example of such an exploration is (Biggs et al. 2009). This examined the regime-shifts using the fisheries food web model, in particular looking at the existence of turning points in a system with two attractors (piscivore and planktivore dominated regimes). Anthropogenic drivers were modelled as gradual changes in the amount of angling and shoreline development. Simulations were carried out to investigate the onset of regime shifts in fish populations, the possibilities to detect these changes and the effectiveness of management responses to avert the shift. In relation to angling it was found that shifts could be averted by reducing harvesting to zero at a relatively late stage (and well into the transition to alternate regime) whereas with development it required action to be taken substantially earlier, i.e. the lag time was substantially longer between taking action and the resultant shift. The behaviour of different indicators to anticipate regime shifts was examined. This is an example of a mathematical model with stochastic elements that is solved numerically by means of a simulation.

Such stylised models, although based on well-understood processes, are caricatures of real systems and have a number of simplifying assumptions. Nevertheless they may provide an insight that would be applicable to many types of real world issues. In contrast to this some seek to understand the properties of some fairly abstract models, aiming to uncover some structures and results that might be quite generally applicable. This is directly analogous to mathematics that seeks to establish some general structures, theorems and properties that might later be usefully applied as part of the extensive menu of tools that mathematics applies. In this case the usefulness of the exercise depends ultimately on the applicability of the results *in practice*. The criteria by which pure mathematics are judged can be seen as distinguishing those that are likely to be useful in the future: soundness, generality, and importance.

An example of where the study of an abstract class of mechanisms has been explored thoroughly to establish the general properties is the area of social influence, in particular the sub-case of opinion dynamics. It can be found in works that use physics methodologies (Galam 1997) or more artificial life approaches (Axelrod 1997). The topic in itself is extremely abstract and cannot be validated against data in any direct manner. The notion of culture or opinion that is studied in these models is considerably abstracted and so hard to accept for any sociologist (von Randow 2003). In this area the most studied mechanism is the creation of consensus or convergence of culture represented by a single real number or a binary string (Galam 1997; Deffuant et al. 2000; Axelrod 1997). Many variations and special cases of these classes of model have been undertaken, for a survey see (Lorenz 2007). Some of these studies have indeed used a combination of parameter sweeps of simulations and analytic approximations to give a comprehensive picture of the model behaviour (Deffuant and Weisbuch 2007). Other merely seem to point out possible variations of the model.

Sometimes the exploration of abstract properties of models can result in surprises, showing behaviour that was contrary to expectations, so this category can overlap with the one discussed in the next section (26.3.1.4).

#### 26.3.1.4 Exploration of the Safety of Assumptions in Existing Models

This is similar to the previous goal, but instead of trying to establish the behaviour of the model *as it is*, one might seek to explore what happens if any of the assumptions in the model are changed or weakened. Thus here one is seeking to explore a space of possibilities *around* the original model. The idea behind this is often that one has a hypothesis about a particular assumption that the model is based upon. For example one might suspect that one would get very different outcomes if one varied some mechanism in the model in (what might seem) a trivial manner. Another example is when one suspects that a certain assumption is unnecessary to the outcomes and can be safely dropped. Thus for this goal one is essentially comparing the behaviour of the original model with that of an altered, or extended model.

For example, Izquierdo and Izquierdo (2006) carried out a systematic analysis of the effect of making slight modifications to structural assumptions in the Prisoner's Dilemma game: in the population size, the mutation rate, the way that pairings were made, etc. all of which produced large changes in the emergent outcome – the frequency of strategies employed. The authors conclude that "the type of strategies that are likely to emerge and be sustained in evolutionary contexts is strongly dependent on assumptions that traditionally have been thought to be unimportant or secondary" (Izquierdo and Izquierdo 2006:181).

How cooperation emerges in a social setting was first fashioned into a game-theoretical problem by Axelrod (1984). The outcome was long thought to be dependent upon the defining questions such as which strategies are available, what are the pay-off values for each strategy, number of repetitions in a match, etc. whereas other structural assumptions, thought to be unimportant, were ignored. On further investigation, however, conclusions based on early work were shown to be rather less general than would be desired, and sometimes actually contradicted by later work.

A different case is explorations of the robustness of the simulation described in (Riolo et al. 2001). This showed the emergence of a cooperative group in an evolutionary setting similar to the Axelrod one mentioned above. Here each individual had a characteristic (modelled as a number between 0 and 1) and a tolerance in a similar range. Individuals are randomly paired and if the difference between the partner's and its own characteristic is less than or equal to their tolerance then they cooperate, otherwise do not. As a result a group of individuals with similar characteristics formed that effectively shared with each other. However later studies (Roberts and Sherratt 2002; Edmonds and Hales 2003) probed the robustness of the model in a number of ways, but crucially by altering the rule for cooperation from "cooperate if the difference between the partner's and its own characteristic is *less than or equal* to their tolerance" to "if the difference between the partner's and its own characteristic is *strictly less than* their tolerance", i.e. from "$\leq$" to "$<$". When this change is made the crucial result – the emergence of a cooperative group – disappeared. It turned out that the (Riolo et al. 2001) effect relied on the existence of a group of individuals with *exactly* the same characteristic with whom they had to cooperate, since the smallest tolerance possible was 0. When the existence of completely selfish individuals was made possible by this change, the cooperation disappeared.

#### 26.3.1.5 Exploring Counter-Factual Possibilities

We only observe a few of the possible configurations of the social phenomena around us. Thus it is natural to wonder what might happen if events or processes were other than as observed or known to be the case. This is the world of artificial societies, where possible worlds loosely related to the one observed are explored. Sometimes an analogy with artificial life is made, where alternative algorithmic versions of life in the broadest sense are specified and experimented with – not life-as-it-is but life-as-it-might-have-been.

An extreme example of this is Jim Doran's model of a society with knowledge of the future (Doran 1997) – this can be thought of as what a society might be like whose members' predictions of the future happen to be correct. Clearly this is a case that does not hold in human society.

Such explorations might not contribute much to the understanding of our society, but it may inform the design of distributed computational systems where the components have a need to flexibly organise themselves in a way analogous, *but not identical to*, how humans organise (see Chap. 21, Hales 2013).

### *26.3.2 Concrete Goals*

Here we consider some of the goals that are more at the concrete and descriptive end of the simulation spectrum. These tend to be *more* concerned to relate to evidence and also to be much more specific. In the subsections below the "plausibility" of assumptions, results and simulations is a frequent issue. The simulation of human societies has not yet reached the situation where there is enough evidence to obtain much more than simple plausibility. At this current stage of social simulation, getting close enough to be deemed a "plausible" model is difficult enough, and there is almost never data enough to justify a stronger claim. Thus claims of anything stronger should be treated with appropriate scepticism.

#### 26.3.2.1 Building Towards Realism

One common approach is to start with a fairly simple model that is easier to understand and then to add aspects and mechanisms that are thought to be significant aspects of an observed system. That is to build an additional level of realism to make the model more plausible or useful in some way (e.g. as a thought experiment). This is sometimes known as the TAPAS approach, i.e. to 'take a previous model and add something'. It is consistent with the engineering principle of "KISS" – keep it simple, stupid. In this approach one starts simply and adds more features or aspects one at a time if and only if the simple approach turns out to be inadequate for some purpose.

Thus Izquierdo (2008) starts with some standard models of the iterated prisoner-dilemma games and adds some more "realistic" features, such as case-based learning and reasoning. A key idea in this is to maintain rigorous understanding of the extended model, but take a step towards models that might eventually be validated against observed data from human interactions.

Whether one would, in fact, reach useable and valid models by this means is contested, with the alternative approach being to start with a complex model that reflects the evidence as well as possible and then seek for understanding and simplifications of this (Edmonds and Moss 2005).

To investigate the social aspects of socio-environmental systems, some highly complicated models often have to be used that include the relevant biophysical dynamics, coupled with social simulation. Rather than developing all such components of the simulation system "from scratch" (and because the biophysical parts are relatively universal), these systems have a modular architecture designed to be reusable. It may therefore be more accurate to refer to the software as a "toolkit" from which various sub-models can be configured depending on the desired purpose of a particular study. In the area of land-use simulation, PALM (Matthews 2006) is one such integrative model, and FEARLUS (Polhill et al. 2001, 2008) is part of another longstanding approach to socio-ecological modelling. With each iteration the toolkit obtains further refinement and new features – whilst the level of understanding of its user(s) increases. The social simulator is interested in what additional complexity the human interaction part brings, and to what extent it adds realism to the model's behaviour when compared with observed evidence.

#### 26.3.2.2 Extending Evidence to Extrapolate to Unobserved Cases

Data about social systems is often limited to measurements from a limited number of observed cases. Thus there may well be many cases within a spectrum of observed cases that one would like to estimate the outcome for. Of course, one could use simple statistical techniques such as linear interpolation or similar to do this, but such techniques depend upon assumptions concerning the regularity of the results with respect to small changes in the set-up, which may be implausible for some social systems. In this case one might simulate the system using plausible assumptions, and validate it against the known observations, then find the outcomes for set-ups that are different to those observed. For the results of this to be reliable the simulation needs to be well validated; for it to correctly indicate the observed cases; to not differ very much from the observed cases (in contrast to the case described in Sect. 26.3.2.1); and for any unvalidated assumptions to be of a mild and uncontroversial nature.

The plausibility of the results from such experiments depends upon the validity of the original measurements as well as the generality of the assumptions (which must be plausible for the unobserved as well as observed cases).

For example (Brown and Harding 2002) use a microsimulation model to extend regional socio-demographic (census) data to cases that are not directly observed (synthetic householder-level records for each spatial district). The extension is attempted with assumptions that are thought as deliberately cautious.

The "Sienna" programme (Snijders et al. 2010) fits a particular class of dynamic network model to "waves" of panel data. Simplifying a little, what happens is that the modeller specifies some basic assumptions (e.g. symmetry of network links) along with more than one set of panel data concerning the properties of the nodes at certain points in time. The algorithm then finds the dynamic network model that is consistent with the given specified constraints and that most closely fits the data. This is directly analogous to the process of fitting a line to a set of values using minimum total squared errors (or similar). What one gets out of this are some "surprise free" projections to network and node properties for times other than those given in the waves of panel data. This is not simulation in the same sense as other simulations mentioned here, since what is simulated is not a kind of process (that is given in the base specification of the family of models this technique uses) but rather a set of structures and values that fit given data in a statistical sense. When this technique is reliable and what its particular biases are, have not yet been established.

#### 26.3.2.3 Establishing the Consistency of a Process/Assumption with Evidence

Oftentimes a social process is not included in a study because it is not considered to be valid in the same way as a physical or biological principle might be. This is particularly true in historical examples where social processes are less in evidence. Going back to our second example of generative archaeology (Sect. 26.1.2), there are few archaeological findings that suggest a particular social structure and set of social processes, hence the need often for guesswork and the resulting coexistence of many competing theories. However, as previously discussed, this is an area in which social simulation can make an important contribution.

Perhaps the most well-known example is the Artificial Anasazi simulation model (Axtell et al. 2002). The objective was to see if a model could be constructed broadly consistent with available evidence – the number of households settled in part of the U.S. South West region over the period 800–1350. The performance of the model was impressive in its convergence upon the actual historical time-series after a calibration of several parameters (a 'fitting' process), which suggested new social explanations as to the apparent land abandonment after 1350 might be possible. Interestingly a later paper (Janssen 2009) demonstrates that the model fit is mainly explained by two parameters related only to the model's carrying capacity. The author argues that a more insightful basis might be to generalise the target domain, working initially from the less concrete goals, rather than fitting a particular case and focusing on one evident and quantifiable trend (such as population). If the evidence base is broadened to include more ethnographic knowledge this approach would resemble the pursuit of abstract goals as discussed in Sect. 26.4.1.4.

Data about real world social networks introduced at the design or validation stages can be a valuable way of checking the consistency of a model. For example Guimera et al. (2005) reconstruct the history of team collaborations in five different scientific and artistic fields and the development of corresponding collaboration networks. The authors develop and parameterise a probabilistic model of team selection. Using real data on team sizes, along with estimation of probabilistic parameters, to control the team assembly mechanism, the characteristics of the resulting networks (the degree distribution and the largest component) are compared with the real ones (independently for each of the five cases). The interest is in the transition of the collaboration network from "isolated schools" to an "invisible college" – the point at which the largest component of the network contains 50 % or more of the nodes (which is the case for all representative fields). All simulated network measurements are shown to be in close agreement with the real networks, which establishes the plausibility of the proposed team selection mechanism. However, being a probabilistic model it does not attribute any particular decision process to this mechanism that might be able to reveal new questions.

Another example is in (White 1999), which attempts to evaluate some statistical assumptions against data about marriage systems in different cultures using a "controlled simulation".

#### 26.3.2.4 Analysis of Influence Factors

In any complex system it is very difficult to estimate the importance of different factors on particular outcome measures or results. This is due to the "non linearity" in many social systems where a normally insignificant factor can trigger a system-wide change in behaviour. However, given a trusted simulation model of the system, one can perform experiments to determine the importance of each factor in the class of simulation set-ups that are run. Thus one does not have to determine the relative importance of factors on an a priori basis; one can simply run the experiments and measure the outcomes. Clearly this approach depends on having a reliable simulation model.

In (Saqalli et al. 2010) a simulation model of the development over several generations of a rural agrarian society is investigated to weigh the importance of several different model parameters on simulation results. In the simulation experiments reported, four parameters were assessed in relation to six state variables – with measurements taken at the end of the run. The model was based on a case-study of the Nigrien Sahel, typified as a low data situation where, in particular, little has been published on the social factors governing access to economic activities (including off-farm activities so often neglected as an important revenue generating source) or on intra-household dynamics (which the authors recognise as having a complex structure). The objective was to assess the robustness of results against variation in socio-economic and biophysical parameters to show that it is "constrained by the different parameters of its structure" (Saqalli et al. 2010: para. 3.6). This step provides the researcher with an improved understanding of the range of outcomes possible with the model and what might constitute a significant or meaningful difference when comparing outcomes. It is worth noting, however that the single parameter approach neglects any possible parametric interaction that could be identified from a pair-wise analysis of influence factors.

A very different example of this is (Yang et al. 2009), which studies the factors that influenced success in the system of Chinese civil service exams that existed in the Imperial era in mainland China. The simulation model used historical data from civil service records and some assumptions to assess the importance of factors such as class, wealth and family connections in terms of success at passing this exam (and hence obtaining a coveted civil service post). It is difficult to see how such indications about events that are otherwise lost in the past could be obtained, although this is open to the criticism of being unfalsifiable.

The disadvantages of this approach are that the assessment of influence is only as good as the simulation model, and it only samples particular sets of initial conditions – it does not rule out the case where very special values of parameters cause totally different outcomes (unless one happens to be lucky and sample these).

#### 26.3.2.5 Assessment of Policy Options

Recently more and more articles have appeared in the literature featuring ABMs that address policy-making in contemporary issues such as developmental sustainability and climate change adaptation. For example Berman et al. (2004) consider eight employment scenarios defined by different policies for tourism and government spending, as well as different climate futures, for an ABM case study of sustainability in the Arctic community of Old Crow. Scenarios were developed with the input of local residents: tourism being a policy option largely influenced by the autonomous community of Old Crow (stemming from their land rights), and attracting great local interest. In ABM, policy options are often addressed as a certain type of scenario (scenarios are discussed in Sect. 26.3.2.9), embedding the behaviour of actors within a few possible future contexts. The attraction of this approach is that the model could potentially be used as a decision support tool, in a form that is familiar to many analysts, to provide answers to very specific policy questions. The merit is that it can improve the reckoning of human and social factors and information into the issues at stake; the drawback is the multiplication of uncertainties, not least of which is that we do not convincingly know how social actors might adapt (even if the possible policy options are more concrete).

For example (Alam et al. 2007) investigates the outcomes indicated by a complex, and detailed model of a particular village in South Africa. This model in particular looks at many aspects of the situation, including: social network, family structure, sexual network, HIV/AIDS spread, death, birth, savings clubs, government grants and local employment prospects. It concludes with hypotheses about this particular case. This does not mean that these outcomes will actually occur, but this does provide a focus for future field research and may provide thought for policy makers.[15]

#### 26.3.2.6 Social Engineering: "Designing" Better Systems

Market design is the branch of economic research aiming to provide insights about which market protocol, i.e. interaction structure and information circulation rules, is the best to obtain certain characteristics of a market. Agent-based simulation seems to be a good method to test several such protocols and see their influence on economic performances, e.g. efficiency, fairness, power repartition (Marks 2007). Each protocol is already known for its advantages and disadvantages (e.g. Dutch auction is fast; Double Auction extracts the highest global profit). Since not every good aspect can be achieved with a single protocol, one has to choose the aim to attain (LeBaron 2006). Then, assuming agents act rationally, it is possible to compare protocols to see what difference it makes in prices or other indicators (e.g. Kirman and Moulet 2008). Many studies were designed to fit the context of electricity markets (very crucial since unpredicted shortages are a problem and prices can vary quickly) and are usually treated by a comparison of protocol (Nicolaisen et al. 2001). One can also note the use of "evolutionary mechanism design" (Phelps et al. 2002; March 2007) where strategies of three types of actors – sellers, buyers and auctioneers – are all submitted to evolution and selection and the actual organization of the market evolves while the context of production and demand is fixed. In today's economy more and more artificial agents really interact – either on bidding on consumers' sites or even in financial markets (Kephart and Greenwald 2002) – so there is some convergence between real markets with artificial markets and designed artificial systems which utilise market mechanisms. For a more detailed discussion of modelling and designing markets see Chap. 23 in this handbook (Rouchier 2013).

#### 26.3.2.7 Data Integration

A mundane and sometimes overlooked aspect of the scientific process is simple description. That is, recording what has been observed in a suitable form. Traditionally these forms have included the like of narratives, logs, videos, measurements, and pictures. However simulations can also be used as a sort of description, where the aim is not to express a *theory* about a mechanism, but rather to integrate as much of the relevant evidence about what is observed as possible about a particular target. Simulation has some advantages in such a process, since it can allow the integration of several different kinds and levels of evidence within one framework. To take some examples: aspects of narrative texts can be

[15] Although in this particular case it did not, since the model indicated outcomes that the policy makers preferred to ignore, being not compatible with the policy they had already fixed upon.

incorporated within the behavioural rules of an agent; the social network of sub communities be compared to those that result from the simulation, the time-series can be compared to the corresponding time-series derived from measurements on the simulation outcomes, and survey data compared to the equivalent answers at instances of the simulation runs. Such integration is far from easy, since some aspects are programmed directly (e.g. agent behaviour) whilst others have to be achieved in terms of the results (e.g. aggregate statistics about the outcomes). Achieving any particular set of outcomes in a social simulation is difficult due to the prevalence of unpredictable interactions and effects (i.e. emergence), so the achievement of a data-integration model is not an easy one. Such models are not entirely (or solely) a description since the structure of a simulation sometimes brings into question the consistency of the various parts of the evidence. Thus if it is difficult to square an account of how individuals behave with some of the outcomes, one may be forced to make some choices, including possibly adding in aspects that are not directly observed. This is alright as long as these are clearly documented and can provide fertile issues for future data collection efforts. However, such data-integration models do not aim for generality beyond the case study (or studies) focused on, and hence can avoid "high" theory to motivate simulation features where this is not supported by the evidence with respect to the target case. It is not that there is *no* theory in such simulations – any description or abstraction, however mild, will rely on *some* theory but the point is that in a descriptive simulation such theory is either well established, or relatively mundane.

Examples of simulations that intend to be descriptive in this sense include (Christensen and Sasaki 2008) which aims at producing a simulation of the evacuation from a particular building, with a view to a future evaluation of evacuation plans and facilities, in particular with regard to disabled people. It uses many particulars of the building structure, but makes assumptions (albeit of a plausible variety) about how people behave when evacuating. Likewise (Terán et al. 2007) aim to simulate land-use and users within a forest reserve with a view to producing a computational representation of this. As in similar simulations there is a mixture of assumptions that are backed by some evidence and some that are plausible guesses. This simulation is loosely validated against some data and shows results that confirm the results found in some other models. The ultimate use of this (and similar models) is not described.

Such simulations can take a long time to construct, involving many iterations of model development as well as being complicated and slow to run. The advantage of such models is that they are a precise and coherent representation of a set of evidence – in a sense an encapsulation of a particular case study.[16] This can be the basis for experiments and inspection that can lead to further abstraction steps, resulting in theories of the processes observed within the data-integration model being modelled in simpler models whose properties are easier to establish, but whose outcomes can be checked against targeted experiments on the data-integration model.

[16] To be precise: a possible encapsulation of a particular set of evidence on the case study.

#### 26.3.2.8 Finding New Questions and Areas of Ignorance, Hypothesis Suggestion

Another use of a simulation is as an aid to good observation. That is, suggesting issues and questions that should be sought in order to gain an adequate observational coverage. The simulation is developed as in the data-integration model above, including different aspects of the observational evidence that are available. It is often the case that it is only when one tries to simulate a process that some of the gaps in knowledge become clear. Thus building a simulation *as one is observing* can help direct the data-gathering research in order to complete an adequate computational description. In this sense it forms a similar role to simulation in some cognitive science (Newell 1990; Sun 2005).

For example, (Moss 1998) exhibits a simulation built on a mixture of bases: (a) an assumed but plausible cognitive architecture which captures how one might divide up a problem into sub-problems until they are doable, (b) some suggestions elicited from an expert from the domain and (c) plausible guesses for the remainder. This model attempted to examine behaviour in the face of crises (defined as when one unwanted event causes another in an out-of-control chain), in particular how the rotating of crisis-management teams and the information they pass on to the next team might impact upon their effectiveness at fighting the crisis. The results were not independently validated, but this is not the point of this simulation. As the author says:

> . . .results obtained with the North West Water model indicate a clear need for an investigation of appropriate organizational structures and procedures to deal with full-blown crises.

In contrast, (Younger 2005) is a very much more abstract model, which is only loosely built upon evidence, but has the same broad aim of suggesting hypotheses – in this case, hypotheses concerning the occurrence of violence and revenge within egalitarian societies. Clearly the plausibility of the hypotheses or questions suggested by a simulation will be greater when the simulation is more firmly rooted in evidence. However, hypotheses and questions can be worthwhile investigating whatever their source, and at least having a simulation grounds and defines the question in a precise way, making clear what it might explain and the sort of other issues and questions that might accompany it.

#### 26.3.2.9 Creation/Critique of Scenarios

Berman et al. (2004) present an example of scenarios being used to constrain models to produce simulations of the wider consequences of those scenarios (as measured by relevant socio-economic or environmental indicators or by their possible influence on human institutions) that can then be used to inform discussions with stakeholders and may ultimately produce a better understanding of such changes. Bharwani et al. (2005) use climate change scenarios to investigate adaptive decision making among villagers in the Limpopo district of South Africa, focusing on the use of seasonal forecast information in farming strategies. Data from the Hadley Centre climate model – HadAM3 – showing a 100-year drying trend with increasing potential evapotranspiration (PET), were used as model input (providing PET and precipitation values). Results show a degree of resilience to these changes is afforded when the forecast is correct 85 % of the time so that farmers establish increased trust in, and use of, seasonal forecasts. They are able to choose cropping strategies that are suited to climate change, though this behavioural shift may only occur over a very long time-frame.

Bharwani et al. (2005) introduce the use of scenarios into the methodology in a further and very interesting way: by postulating them as 'drivers' of actors' decision-making processes. In this ethnographic approach, the authors combine simplified scenarios across different domains (irrigation, forecast and market) asking respondents what they would do under each scenario, in a given context. This information was then used to produce the model rules for the agents' decision-making.

In either case, where conventional scenarios used in futures planning can seem rather terse and lacking in specifics – which may be a limitation to their subsequent use in policy discussion – simulation outputs that explore scenarios offer a great deal of detailed information "that would be difficult to imagine otherwise" (Berman et al. 2004: 410). Moreover, this can apply at different levels of analysis from trends in macro variables down to the impacts on different sectors and regions, as well as differentiated impacts for agents fitting any given 'profile' in which the analyst is interested. Perhaps greater care has to be taken, however, in the use of model-generated scenarios, to ensure that these are not taken as 'more accurate predictions' by virtue of being 'computed' stories rather than conventional 'imagined' stories.

Scenarios are often used in policy discussions, e.g. climate change. However they are usually somewhat vague and/or only described in qualitative terms. Simulations can be used to produce consistent scenarios or to produce models that instantiate aspects of given scenarios (Taylor et al. 2009).

#### 26.3.2.10 Intervention with Stakeholders

Instead of developing a simulation to represent some aspect of society, one can also try to use a simulation to intervene *in* society. That is: use a simulation to change some interaction between stakeholders, for example to facilitate collective decision making or mutual understanding. One well-known approach is the Companion Modelling approach which has been developed in the last decade (see Chap. 10; Barreteau et al. 2013).

An example is demonstrated by Etienne (2003). Here, a model is used in conjunction with a role-playing game to show chosen participants the issues that can arise when several users compete on a pastoral resource. The building of the model was an integration of multidisciplinary knowledge acquired on French

Mediterranean sylvopastoral systems into a model capable of representing the interactions between ecological dynamics and social behaviours. In order to help foresters and livestock farmers to better integrate these interactions into their planning work, a multi-agent system was designed to simulate different management strategies and to compare their impact on forest quality. This model was coupled with a role-playing game (RPG) initially developed as a didactic support to sylvopastoral training programmes and very soon, it proved useful in the negotiations and interactions between livestock farmers and foresters involved in the management of the same forest. The tool revealed itself flexible enough to make it possible to play with actively involved stakeholders such as the current users of the resource (local farmers and foresters), with potential regulators of the system (managers or administrators), technical experts (extensionists, technicians) or learners concerned with the topic (students, scientists).

This model is effectively an intervention between the livestock farmers and foresters by being a subtle mediating tool, allowing the stakeholders to play at decision making, to educate them in the possible effects of their decision-making and to thus encourage debate and introspection. This model has also been used for didactic purposes (Sect. 26.3.1.1) by getting agronomy students to play it.

## 26.4 Inputs and Results of Simulation Models

One method of assessing the use, and ultimately the success, of a simulation for understanding aspects of society is to tease out what has gone into making a simulation model, the *input*, and how the results from the simulation are interpreted and used, the *output*. These, the input and the output together form the mapping from the computer program and its calculation to and from the target of study. They are crucial parts of what characterises a simulation, even if they tend to be described in a less formal manner than the simulation code and behaviour.

### *26.4.1 Inputs*

What is put "in" to the design of a model tends to be more explicitly distinguished in papers than what comes "out". This might be because the "job" of a simulator is seen as a process of deciding what processes and structures will go in to a model and because the inputs are under the control of the simulator in a way which the results certainly are not, and hence can be displayed and talked about with greater confidence. However, all social simulations are based on a raft of different assumptions, settings and processes. These are somewhat separated out for analysis here.

#### 26.4.1.1 Evidence-Based Assumptions

If there is some evidence about the nature or extent of the processes that are being observed then this can be used to inform the set-up or structure of a simulation. For example, evidence from social psychology might be used to inform the specification of the behavioural rules of a set of agents in a simulation, or the narrative account of a participant used as the basis for programming a particular agent.[17] Of course, it is rare that such evidence constrains the possible settings and algorithms completely but rather that it partially constrains these or constrains them in conjunction with additional assumptions from another source. Clearly the more assumptions can be constrained by evidence (either directly or as the result of previous research) the better. The presence of other assumptions and inputs does not make a simulation useless, especially if documented, it is just that simulation results and usefulness are relative to the assumptions, so that if assumptions are included that are completely misguided and critically affect the results, then this would seriously limit their use with respect to the observed world.

#### 26.4.1.2 Indirectly Inferred Settings

In situations where there are some parameters that are unknown, and where there is a relative abundance of time-series data, one can attempt to infer the values of these by seeing which parameter values result in the model giving the best fit to a segment of the time-series data. This is a sort of evidence-based setting, but it often seems to be used when the parameters concerned do not have any discernable meaning in terms of the target of modelling. Thus when there is a tradition of using a certain kind of decision or learning algorithm in an agent, then this might be "fitted" to an initial segment of the data (so called "in sample" data) *even when it is unlikely*[18] *that the algorithm corresponds to how the target agents think.* Thus the credibility of this technique is dependent on the reliability of the other assumptions in the model, and the meaning of the parameters being fitted. If the parameter was a scaling parameter, then this might well be a sensible way to proceed.

[17] This can either be done directly as a translation of an interview text into programmed rules or used to check that such programming is correct by comparing the resulting behaviour of an agent against what happens when the simulation is run. Thus there is not an absolutely clear distinction between verification and validation from evidence. In a sense this second method is verification since the programming is rejected until correct but, on the other hand, this is part of the production of a simulation, which may only be completed later for its validation as a whole.

[18] Unlikely in view of the psychological or sociological evidence about the target subjects.

#### 26.4.1.3 Documented Theoretical Assumptions

Clearly, researchers do not invent all the details and algorithms of their model from the ground up, but are doing their research with knowledge of certain approaches and algorithms and within a community of other research, with established techniques and traditions. Thus many parts of a simulation model will be based on those of other models, or algorithms from other fields. Thus many models in Economics will use a decision algorithm based on constrained comparisons of predicted utility and other models might import techniques from the fields of Artificial Intelligence or Evolutionary Computation. It seems impossible to completely avoid all such theoretical assumptions; however there are distinctions to be made in terms of the *strength* of the assumptions, the likely *biases* behind such assumptions and the degree to which they are *evidence-based*.

"Strong" assumptions are those that are surprising or seem to specify conditions that are rarely observed. Thus an assumption that an agent has *in effect* a perfect model of the economy in its head is a very strong assumption, since even experts find it difficult to understand the economy as a whole. Strong assumptions are often introduced to allow for analytically solvable models to be specified and used, for example the assumption of perfect information in game theory. Whilst analytically tractable models were necessary when there was no other avenue for the precise modelling of many kinds of phenomena, the advent of cheap computing power and accessible simulation platforms means that often more appropriate methods are now available, with analytic models possibly being used to check or understand the reference simulation model, rather than being the focus. Clearly, other things being equal, weaker assumptions are preferable to strong ones – the stronger an assumption the more evidence is needed to justify its use. In any case all such assumptions should be as fully documented as possible.

#### 26.4.1.4 Explored Conditions

In much simulation work there will be a focus hypothesis or set of hypotheses that are being investigated. In these cases it is usual to try the simulation using that hypothesis and then compare the results to those coming from a version of the simulation with a different hypothesis implemented. This provides evidence about the possible effects of that hypothesis on the outcomes, allowing comparison with evidence and possible subsequent inference as to which is more likely to be the case. The clearest case of this is testing the significance of the inclusion of a hypothesis against that of a *"null" model*[19] to see if the properties of the results that are deemed significant indeed result from the hypothesis or from other aspects of the model. Thus a simulation of a stock-market might compare the results

[19] A "null" model is a model version where the claimed causal mechanism is eliminated to see if the resultant "effect" would have arisen as the result of background (e.g. random) mechanisms anyway.

obtained with intelligent agents that notice patterns in pricing and try and exploit these to agents that buy and sell at random. Unfortunately it is sometimes the case that a simulation is presented purporting to show the significance of a hypothesis without indicating what the comparison case is.

#### 26.4.1.5 Randomness and Other Essentially Arbitrary Assumptions

A simulation modeller is often faced with deciding how to design a part of a simulation model for which there is neither evidence nor any tradition of modelling to guide them. In such a case one might simply make that aspect random. For example, where it is unknown how a kind of choice is made in the modelled situation it might be implemented as a random choice in a simulation model of that situation.[20] This is usually done in conjunction with a "Monte Carlo" approach which runs the simulation a number of times and averages the resulting different sets of outcomes. Presumably this is done under the assumption that the introduced randomness will be averaged out leaving only the effects of the other design settings – however this assumption is rarely proved but often simply remains a hope. Of course, if it can be shown that the value of the particular input does not influence those aspects of the results that are deemed significant by a series of simulation experiments (or otherwise) then a random input or process might well be acceptable. However, in this case, a constant value might be simpler and have the same effect.[21]

We suspect that many uses of randomness in simulations are in the nature of a programming "stub" – that is, a stand-in that the programmer intends (or intended) to expand to a more plausible algorithm at a later date. Whilst this is perfectly acceptable during model development and to some extent inevitable given that researchers always have time constraints, such stubs are likely targets for criticism by other researchers. At the very least some exploration of them to assess the extent to which they affect those aspects of the results deemed significant is advisable.

Randomness can be considered as a special case of a broader class of assumptions: those that are added into the model simply to get it to run, and for no theoretical or evidence-related reason. Hopefully these are honestly declared rather than "dressed up" under some other category, although often these are excused under the broad umbrella of "simplicity".[22]

---

[20] Another option is to exhaustively try all the possibilities in a series of simulations or by using techniques such as constraint logic programming but these are technically difficult and require a lot of computational power.

[21] There are possible reasons why a constant value might not work, for example when the input provides some mechanism of symmetry-breaking.

[22] There is nothing wrong with assumptions that had to be made due to constraints on resources, such as time, expertise or computing power, but it is simply disingenuous to pretend that this is sanctioned by a higher "virtue".

#### 26.4.1.6 Undocumented Assumptions

It is not feasible to document all of the assumptions in a model. Firstly, this might take too much space in a single paper[23] and secondly, many might be previously established and well known to those in a particular field of work. However it is also likely that researchers are simply not aware of all the assumptions inherent in their simulation models, due to the limitation of human cognition.[24] Clearly, it is part of the job of other researchers to point out undocumented assumptions where these can be shown to be significant.[25]

### *26.4.2 Outputs*

A similar set of distinctions can be made about what comes out of a simulation, the results. There is not an obligation to describe all the outputs from a simulation, but rather one tends to get a sample of results, which typically is composed of: sample results, sensitivity analyses, evidence of validation, and the outcomes from experiments designed to test a hypothesis. However not all the details of the results are considered as equally significant – we now consider each of these in order of increasing significance.

*Firstly*, there are those aspects of the results that are considered as artefacts of the model, for example the randomness that might have been input to the model.

*Secondly*, there are those features that might be considered to reflect some of the model structure and the processes that result from them. These features may not be one of those parts of the simulation that reflect what is being modelled, but may be caused by theoretical or arbitrary assumptions that were put in. These features of the results may well not be so much of a surprise to the modeller.

*Thirdly*, there are those features of the results that are interpreted as indicating something about what is being modelled, for example they may suggest a hypothesis about those phenomena. That is they indicate a possibility that may be inherent in what is being modelled or that is possibly inherent in the target of modelling. This may well go beyond what can be directly validated in the model but, for example, track counter-factual possibilities concerning what *might* have occurred.

---

[23] However this is a poor excuse given that a technical paper which is relatively complete can easily be archived and referenced along with a journal article or report.

[24] Alternatively it may be because the simulation designers had not thought about what they were doing.

[25] It is trivial to point out that a simulation has missed out some assumption or other, but this is not very useful. It is far more useful to point out *how* and *why* an assumption might be important and for *which* purposes.

*Lastly*, there are those features that would be positively expected of the phenomena being modelled. That is, if it were *not* present then it would be taken as evidence that there was something amiss with the model. In other words, it is a necessity of the phenomena. It is against this category of results that models are validated.

It is not easy to distinguish these different categories of significance in terms of the results, since causation within a model can be very complicated, being a result of many of the model aspects interacting together. It is also usually the case that the modeller has hypotheses (or assumptions) about what aspects of the results are significant in which ways. This is crucially useful information to impart to a reader interested in the results. However, this is often left implicit.

One might justifiably criticise many social simulations in terms of the lack of empirical grounding of both inputs and outputs. Many social simulations have only the weakest connection with anything observed: the inputs are largely assumption-based, and indeed often highly artificial; the outputs only relating in the broadest way to any data and then only in terms of a few aspects of the possible outputs (i.e. only a few selected aspects are deemed significant to what is observed and then in the loosest, "hand waving", manner). It may well be that simulating human society is just very, *very* difficult, and one suspects that it is simply easier to stick to considering abstract ideas.

## 26.5 An Assessment of Simulations for Understanding Social Phenomena

As discussed in previous sections, complex social phenomena containing multiple interacting actors can be computationally analysed using ABM. In this sense it is worth noticing that evidence-driven modelling approaches tend to guide modellers towards more up-to-date data and better understanding of social phenomena than theory-driven approaches. To avoid using highly speculative ("strong") assumptions whilst modelling, it is essential to have very detailed knowledge of the phenomena in question. That comprehensive grasp, backed with evidence, is helpful to identify relevant model parameters, estimate configuration values and evaluate simulation results.

Perhaps the biggest open challenge for modellers is to work out how results from social simulation models can be useful beyond theory and hypothetical illustrations. To harness the potential explanatory prowess of ABM in the social sciences, it seems necessary to safeguard models with scrutinised findings from large amounts of data about the social phenomenon. Nevertheless this is currently very difficult as access to such datasets is often hampered either due to (Lucas 2011):

- Being non-existent, thus requiring funding to collect and process all relevant information;
- Being unavailable because of privacy agreements, such as in non-disclosure agreements;
- Or, if at hand, being incomplete or outdated.

Validating outputs that have no comparable evidence is an eminent issue, and this seems only clarified by comparing simulation results with further data. In (Lucas 2010, 2011) there is a discussion of a survey carried out in 2009[26] with 12 leading academic researchers – each of them having managed mid to long-term (3 to 5 years) social simulation projects in Europe and the United States – regarding how their endeavours were modelled, applied and whether these have been useful beyond theory. The questionnaire consisted of the following questions aimed at eliciting views:

1. How were fieldwork findings used to guide the simulation development?
2. What were the contributions of simulation results to stakeholders?
3. Were simulation results regarded by them as useful as fieldwork findings?
4. What could have improved the chances of providing these via simulations?

Three businesses offering simulations, which take into account social behaviour, were also approached, but all refused stating non-disclosure agreements. All interviewees mentioned that their models targeted the scientific community and could only, at best, provide plausible results regarding scenarios and that, albeit coherently justified, nothing obtained in simulations could be regarded as directly useful for policy-making purposes. All 12 cited that gathering detailed data about the actual phenomena by interviewing stakeholders and reviewing existent literature helped to better understand their context and served as a good guide during the modelling process. More than half (7) said stakeholders and policy-makers were not interested in simulations per se, and that only real success cases (even those with only anecdotal evidence) are what are taken into account in decisions. On the other hand few (3) mentioned that simulations, despite their shortcomings, attracted significant interest from stakeholders and policy-makers. In their view the modelling process is a time-consuming task and occasionally even regarded it as unproductive. This contrasts with positive experience with fieldwork, which – despite also demanding a lot of time – a majority (10) confirmed was useful in acquiring relevant new knowledge to practitioners. Engaging with stakeholders and policy-makers was interpreted by all (12) as indispensable to improve their understanding of the actual social phenomena. Yet maintaining efficient interaction and managing practitioners' interest over many months of work engagement was generally deemed as strenuous and difficult. This finding is partially supported by other larger projects regarding the effective collaboration between policy-makers

[26] Data was compiled until July 2009, on time for discussion at the 6th European Social Simulation Association Conference in September, and then updated for the Artificial Intelligence & Society Journal.

and researchers such as (Young and Mendizabal 2009). Some (4) cited that social simulation models perhaps could be integrated in tools for aiding mediation of group decisions. Lacking confidence about model results is another aspect raised by all (12) interviewees, along with difficulties of coding or interpreting qualitative data appropriately, plus communicating the model itself and its results intelligibly to a non-technical audience. Some (6) interviewees said that they had no intention to influence the social phenomenon in question, but only to model it plausibly.

Commissioned fieldwork, to date, has greater chances of being timely useful in this sense, as resultant reports can provide very specific and up-to-date information that is easily understood by stakeholders. The implication of such new information is usually more quickly recognised than results obtained in simulation models, as these usually deal with intricate processes over longer time scales. Claims that social simulations could support, or guide, decision and policy-making seem only possible with aid of in-house experienced modellers working in close liaison with stakeholders. That is necessary as, meeting social simulation aims and objectives (beyond theory) is still an experimental process of many trials and errors, which requires good technical judgement to know what should be done when simulation results have gone beyond the existing evidence (what is factually known).

A survey of agent-based modelling practices in the literature (Heath et al. 2009) revealed that there were more applications in social sciences (24 %) than in public policy (8 %) among the 279 articles published between 1998 and 2008. Of the 68 social science applications, a majority (66 %) used ABM as a hypothesis generator for systems that were assumed to be less well-known, and the remaining articles (34 %) used ABM as a mediator in order to represent a system that was moderately understood and gain insights on the system's characteristics and behaviour. In contrast, only a small portion of the 23 public policy applications (4 %) used ABM as a hypothesis generator, and most articles (96 %) used ABM as a mediator.

## 26.6 Conclusion

Simulation has undoubtedly helped to improve our understanding of human society, although in a number of different and usually indirect ways. It is fair to say that, so far at least, this has served to improve our understanding of some societal processes and our ideas *about* society rather than directly in terms of being able to strongly predict aspects of society or conclusively test hypotheses about society.

Simulation is not a replacement for other ways of understanding society[27]; it is simply a flexible way of precisely modelling it in a way that can represent some of the dynamic and complex aspects of it. It can be especially productive in conjunction with other approaches. For example, analytic models can be used to check the outputs and properties of a simulation model and help us understand the model and,

[27] At least, not in any of the cases we have yet come across.

conversely, a simulation used to probe and check some of the simplifications and assumptions used in an analytic model. In participatory models, social science techniques of engagement and elicitation can be used to inform the construction of agent-based social simulations as well as the simulations suggesting what might be usefully investigated in terms of the collection of new data.

Clearly social simulation has some way to go in terms of the maturity of its method and the reporting and use of simulation models. There are still a number of areas in which the methodology needs substantial improvement and standardising. There are also significant unresolved issues, such as how to decide what level of detail to include, and to what extent one should rely on prior theory.

We predict that simulation will be even more significant in helping us understand human society in the future, *in particular where it is used in close conjunction with other relevant approaches*.

## Further Reading

The best general introduction to social simulation is (Gilbert and Troitzsch 2005) which covers general issues and gives code examples. For a wider range of views on social simulation the published papers from the US National Academy of Sciences colloquium on "Adaptive Agents, Intelligence, and Emergent Human Organization: Capturing Complexity through Agent-Based Modeling" (PNAS 2002) give a good cross-section of the different approaches people take to this area. It is difficult to point to further good sources as this topic is so diverse but the Journal of Artificial Societies and Social Simulation has many accessible papers.

## References

Alam SJ, Meyer R, Ziervogel G, Moss S (2007) The impact of HIV/AIDS in the context of socioeconomic stressors: an evidence-driven approach. J Artif Soc Soc Simul 10(4), http://jasss.soc.surrey.ac.uk/10/4/7.html

Axelrod R (1984) The evolution of cooperation. Basic Books, New York

Axelrod R (1997) The complexity of cooperation. Princeton University Press, Princeton

Axtell R, Axelrod R, Epstein JM, Cohen MD (1996) Aligning simulation models: a case study and results. Comput Math Organ Theor 1(2):123–141

Axtell RL et al (2002) Population growth and collapse in a multi-agent model of the Kayenta Anasazi in Long House Valley. Proc Natl Acad Sci 99(3):7275–7279

Barreteau O et al (2013) Participatory approaches. Chapter 10 in this volume

Berman M, Nicolson C, Kofinas G, Tetlichi J, Martin S (2004) Adaptation and sustainability in a small arctic community: results of an agent-based simulation model. Arctic 57(4):401–414

Bharwani S et al (2005) Multi-agent modelling of climate outlooks and food security on a community garden scheme in Limpopo, South Africa. Phil Trans R Soc B 360 (1463):2183–2194

Biggs R, Carpenter SR, Brock WA (2009) Turning back from the brink: detecting an impending regime shift in time to avert it. Proc Natl Acad Sci (PNAS) 106:826–831

Brown L, Harding A (2002) Social modelling and public policy: application of microsimulation modelling in Australia. J Artif Soc Soc Simul 5(4), http://jasss.soc.surrey.ac.uk/5/4/6.html

Cartwright N (1993) How the laws of physics lie. Oxford University Press, Oxford

Christensen K, Sasaki Y (2008) Agent-based emergency evacuation simulation with individuals with disabilities in the population. J Artif Soc Soc Simul 11(3), http://jasss.soc.surrey.ac.uk/11/3/9.html

Clifford J (1986) Writing culture: the poetics and politics of ethnography. University of California Press, Berkeley

Cole HSD, Freeman C, Jahoda M, Pavitt KL (eds) (1973) Models of doom: a critique of the Limits to Growth. Universe Books, New York

Curtis J, Frith D (2008) Exit polling in a cold climate: the BBC–ITV experience in Britain in 2005. J R Stat Soc A Stat Soc 171(3):509–539

David N (2013) Validating simulations. Chapter 8 in this volume

Deffuant G, Weisbuch G (2007) Probability distribution dynamics explaining agent model convergence to extremism. In: Edmonds B, Hernandez C, Troitzsch KG (eds) Social simulation: technologies, advances and new discoveries. IGI Publishing, Hershey, pp 43–60

Deffuant G, Neau D, Amblard F, Weisbuch G (2000) Mixing beliefs among interacting agents. Adv Complex Syst 3(1):87–98

Doran JE (1997) Foreknowledge in artificial societies. In: Conte R, Hegselmann R, Tierna P (eds) Simulating social phenomena. Lecture notes in economics and mathematical systems, 456. Springer, Berlin, pp 457–469

Dunbar RIM (1998) The social brain hypothesis. Evol Anthropol 6(5):178–190

Edmonds B (1999) The pragmatic roots of context. In: Bouquet P, Serafini L, Brezillon P, Benerecetti M, Castellani F (eds) Modelling and using context. Second international and interdisciplinary conference, CONTEXT'99, Trento, Italy, 9–11 Sept 1999, Proceedings. Lecture notes in artificial intelligence, 1688. Springer, Berlin, pp 119–132, http://cfpm.org/cpmrep52.html

Edmonds B (2001) The use of models – making MABS actually work. In: Moss S, Davidsson P (eds) Multi agent based simulation. Lecture notes in artificial intelligence, 1979. Springer, Berlin pp 15–32

Edmonds B (2007) The practical modelling of context-dependent causal processes – a recasting of Robert Rosen's thought. Chem Biodivers 4(1):2386–2395

Edmonds B (2010) Context and social simulation. Paper presented at the IV edition of epistemological perspectives on simulation (EPOS2010), Hamburg, 23–25 June 2010, http://cfpm.org/cpmrep210.html

Edmonds B, Hales D (2003) Replication, replication and replication – some hard lessons from model alignment. J Artif Soc Soc Simul 6(4), http://jasss.soc.surrey.ac.uk/6/4/11.html

Edmonds B, Hales D (2005) Computational simulation as theoretical experiment. J Math Sociol 29(3):209–232

Edmonds B, Moss S (2005) From KISS to KIDS – an 'anti-simplistic' modelling approach. In: Davidsson P et al (eds) Multi agent based simulation 2004. Lecture notes in artificial intelligence, 3415. Springer, Berlin, pp 130–144

Epstein J (2008) Why model? J Artif Soc Soc Simul 11(4). http://jasss.soc.surrey.ac.uk/11/4/12.html

Etienne M (2003) SYLVOPAST: a multiple target role-playing game to assess negotiation processes in sylvopastoral management planning. J Artif Soc Soc Simul 6(2), http://jasss.soc.surrey.ac.uk/6/2/5.html

Galam S (1997) Rational group decision making: a random field Ising model at T =0. Physica A 238:66–80

Galán JM et al (2009) Errors and artefacts in agent-based modelling. J Artif Soc Soc Simul 12(1), http://jasss.soc.surrey.ac.uk/12/1/1.html

Gilbert N, Troitzsch K (2005) Simulation for the social scientist, 2nd edn. Open University Press, New York

Gode DK, Sunder S (1993) Allocative efficiency of markets with zero intelligence traders: markets as a partial substitute for individual rationality. J Polit Econ 110:119–137

Grimm V et al (2005) Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310(5750):987–991

Grimm V et al (2006) A standard protocol for describing individual-based and agent-based models. Ecol Model 198(1–2):115–126

Grimm V, Polhill JG, Touza J (2013) Documenting social simulation models: the ODD protocol as a standard. Chapter 7 in this volume

Guba EG, Lincoln YS (1994) Competing paradigms in qualitative research. In: Denzin NK, Lincoln YS (eds) Handbook of qualitative research. Sage, London, pp 105–117

Guimera R, Uzzi B, Spiro J, Amaral LA (2005) Team assembly mechanisms determine collaboration network structure and team performance. Science 308(5722):697–702

Hales D (2013) Distributed computer systems. Chapter 21 in this volume

Heath B, Hill R, Ciarallo F (2009) A survey of agent-based modeling practices (January 1998 to July 2008). J Artif Soc Soc Simul 12(9), http://jasss.soc.surrey.ac.uk/12/4/9.html

Izquierdo LR (2008) Advancing learning and evolutionary game theory with an application to social dilemmas. PhD thesis, Manchester Metropolitan University, http://cfpm.org/theses/luisizquierdo/

Izquierdo SS, Izquierdo LR (2006) On the structural robustness of evolutionary models of cooperation. In: Corchado E, Yin H, Botti VJ, Fyfe C (eds) Intelligent data engineering and automated learning – IDEAL 2006, 7th international conference, Burgos, Spain, 20–23 Sept 2006, Proceedings. Lecture notes in computer science 4224. Springer, Berlin, pp 172–182

Janssen MA (2009) Understanding artificial Anasazi. J Artif Soc Soc Simul 12(4), http://jasss.soc.surrey.ac.uk/12/4/13.html

Kahn K, Noble H (2009) The modelling4All project – a web-based modelling tool embedded in Web 2.0. In: Dalle O et al (eds) Proceedings of SIMUTools '09, 2nd international conference on simulation tools and techniques, Rome, 2–6 Mar 2009, ICST, Brussels, Article 50

Kephart JO, Greenwald AR (2002) Shopbot economics. Auton Agents Multi-Agent Syst 5 (3):255–287

Kirman A (2011) Complex economics: individual and collective rationality. Routledge, London

Kirman A, Moulet S (2008) Impact de l'organisation du marché: Comparaison de la négociation de gré à gré et des enchères descendants. Working papers, halshs-00349034, HAL, Centre pour la communication scientifique directe, http://halshs.archives-ouvertes.fr/docs/00/34/90/34/PDF/DT2008-56.pdf

Klein G (1998) Sources of power: how people make decisions. MIT Press, Cambridge

Kohler T (2009) Generative archaeology: how even really simple models can help in understanding the past. Invited talk at 6th conference of the European social simulation association, University of Surrey, Guildford

Kohler TA, Gumerman GJ, Reynolds RG (2005) Simulating ancient societies. Sci Am 293 (1):76–82

Kohler TA, Varien MD, Wright A, Kuckelman KA (2008) Mesa Verde migrations: new archaeological research and computer simulation suggest why ancestral Puebloans deserted the northern Southwest United States. Am Sci 96:146–153

Kuhn T (1962) The structure of scientific revolutions. University of Chicago Press, Chicago

Lakatos I, Musgrave A (eds) (1970) Criticism and the growth of knowledge. Cambridge University Press, Cambridge

LeBaron B (2006) Agent-based computational finance. In: Tesfatsion L, Judd KL (eds) Handbook of computational economics, vol 2. North-Holland, pp 1187–1232

Lincoln YS, Guba EG (1985) Naturalistic inquiry. Sage, Beverly Hills

Lorenz J (2007) Continuous opinion dynamics under bounded confidence: a survey. Int J Mod Phys C 18(12):1819–1838

Lucas P (2013) Conventional social behaviour amongst microfinance clients. PhD thesis, Centre for Policy Modelling, Manchester Metropolitan University

Lucas P (2011) Usefulness of simulating social phenomena: evidence. AI Soc 26(4):355–362

Malthus T (1798) An essay on the principle of population. Johnson, London, Transcript available online at http://socserv2.mcmaster.ca/~econ/ugcm/3ll3/malthus/popu.txt

Marks RE (2007) Validating simulation models: a general framework and four applied examples. Comput Econ 30(3):265–290

Matthews RB (2006) The people and landscape model (PALM): towards full integration of human decision-making and biophysical simulation models. Ecol Model 194:329–343

Meadows DH, Meadows D, Randers J, Behrens WWIII (1972) The Limits to Growth: a report for the Club of Rome's project on the predicament of mankind. Universe Books, New York

Moss S (1998) Critical incident management: an empirically derived computational model. J Artif Soc Soc Simul 1(4), http://jasss.soc.surrey.ac.uk/1/4/1.html

Moss S (1999) Relevance, realism and rigour: a third way for social and economic research. Report no. CPM-99-56. Centre for Policy Modelling, Manchester Metropolitan University, Manchester, http://cfpm.org/cpmrep56.html

Newell A (1990) Unified theories of cognition. Harvard University Press, Cambridge

Nicolaisen J, Petrov V, Tesfatsion L (2001) Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. IEEE Trans Evol Comput 5(5):504–523

Phelps S, McBurney P, Parsons S, Sklar E (2002) Co-evolutionary auction mechanism design: a preliminary report. In: Padget J, Shehory O, Parkes D, Sadeh N, Walsh WE (eds) Agent-mediated electronic commerce IV, designing mechanisms and systems. Lecture notes in computer science, 2531. Springer, Berlin, pp 193–213

PNAS (2002) Colloquium papers. Proceedings of the National Academy of Sciences, 99(s3), http://www.pnas.org/content/99/suppl.3.toc#ColloquiumPaper

Polhill JG, Gotts NM, Law ANR (2001) Imitative versus non-imitative strategies in a land use simulation. Cybernet Syst 32(1–2):285–307

Polhill JG, Parker D, Brown D, Grimm V (2008) Using the ODD protocol for describing three agent-based social simulation models of land-use change. J Artif Soc Soc Simul 11(2), http://jasss.soc.surrey.ac.uk/11/2/3.html

Popper K (1963) Conjectures and refutations: the growth of scientific knowledge. Routledge, London

Riolo RL, Cohen MD, Axelrod R (2001) Evolution of cooperation without reciprocity. Nature 411:441–443

Roberts G, Sherratt TN (2002) Does similarity breed cooperation? Nature 418:499–500

Rouchier J (2001) Est-il possible d'utiliser une définition positive de la confiance dans les interactions entre agents? Paper presented at Colloque Interactions, Toulouse, May 2001

Rouchier J (2013) Agent-based simulation as a useful tool for the study of markets. Chapter 23 in this volume

Rouchier J, Thoyer S (2006) Votes and lobbying in the European decision-making process: application to the European regulation on GMO release. J Artif Soc Soc Simul 9(3), http://jasss.soc.surrey.ac.uk/9/3/1.html

Rouchier J, Bousquet F, Requier-Desjardins M, Antona M, Econ J (2001) A multi-agent model for describing transhumance in North Cameroon: comparison of different rationality to develop a routine. J Econ Dyn Control 25(3–4):527–559

Saqalli M, Bielders CL, Gerard B, Defourny P (2010) Simulating rural environmentally and socio-economically constrained multi-activity and multi-decision societies in a low-data context: a challenge through empirical agent-based modeling. J Artif Soc Soc Simul 13(2), http://jasss.soc.surrey.ac.uk/13/2/1.html

Schelling T (1969) Models of segregation. Am Econ Rev 59(2):488–493

Schelling T (1971) Dynamic models of segregation. J Math Sociol 1:143–186

Scherer A, McLean A (2002) Mathematical models of vaccination. Br Med Bull 62(1):187–199

Snijders TAB, Steglich CEG, van de Bunt GG (2010) Introduction to actor-based models for network dynamics. Soc Networks 32:44–60

Sun R (2005) Theoretical status of computational cognitive modelling. Technical report, Cognitive Science Department, Rensselaer Polytechnic Institute, Troy, New York

Taylor R, Takama T, Taylor A, Ziervogel G (2009) Participatory validation of agent-based models to support policy dialogue. In: Gilbert N, Edmonds B (eds) Proceedings of the 6th European social simulation association conference, University of Surrey, Guildford

Terán O, Alvarez J, Ablan M, Jaimes M (2007) Characterising emergence of landowners in a forest reserve. J Artif Soc Soc Simul 10(3), http://jasss.soc.surrey.ac.uk/10/3/6.html

Vermeulen PJ, de Jongh DCJ (1976) Parameter sensitivity of the 'Limits to Growth' world model. Appl Math Model 1(1):29–32

von Randow G (2003) When the centre becomes radical. J Artif Soc Soc Simul 6(1), http://jasss.soc.surrey.ac.uk/6/1/5.html

White DR (1999) Controlled simulation of marriage systems. J Artif Soc Soc Simul 2(3), http://jasss.soc.surrey.ac.uk/2/3/5.html

Yang C, Kurahashi S, Kurahashi K, Ono I, Terano T (2009) Agent-based simulation on women's role in a family line on civil service examination in Chinese history. J Artif Soc Soc Simul 12 (2):5, http://jasss.soc.surrey.ac.uk/12/2/5.html

Young J, Mendizabal E (2009) Helping researchers become policy entrepreneurs – how to develop engagement strategies for evidence-based policy-making. Briefing paper, Overseas Development Institute, London

Younger S (2005) Violence and revenge in Egalitarian societies. J Artif Soc Soc Simul 8(4), http://jasss.soc.surrey.ac.uk/8/4/11.html

# Index




B. Edmonds and R. Meyer (eds.), *Simulating Social Complexity*, Understanding Complex Systems, DOI 10.1007/978-3-540-93813-2,