Texts in Computer Science

Claudio Cioffi-Revilla

Introduction to Computational Social Science

Principles and Applications

Springer

Texts in Computer Science

Editors
David Gries
Fred B. Schneider

For further volumes:
www.springer.com/series/3191

Claudio Cioffi-Revilla

Introduction to Computational Social Science

Principles and Applications

Springer

Claudio Cioffi-Revilla
George Mason University
Fairfax, VA, USA

Series Editors
David Gries
Department of Computer Science
Cornell University
Ithaca, NY, USA

Fred B. Schneider
Department of Computer Science
Cornell University
Ithaca, NY, USA

ISSN 1868-0941 ISSN 1868-095X (electronic)
Texts in Computer Science
ISBN 978-1-4471-5660-4 ISBN 978-1-4471-5661-1 (eBook)
DOI 10.1007/978-1-4471-5661-1
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2013957559

Printed on acid-free paper

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

To my Lady Jean, L.G.C.H.S.,
on our XL anniversary

Preface

This textbook provides an introduction to Computational Social Science (CSS), an emerging field at the intersection of traditional social science disciplines, computer science, environmental science, and engineering sciences. CSS is inspired by 20th-century pioneers such as Herbert A. Simon, who saw essentially a new way of doing social science enabled by computational science and technology. Scientist and visionary Peter J. Denning once said that "the science of the 21st century will be computational," so this book is proof of that idea in social science domains.

As a textbook, this is intended as a systematic introductory survey to familiarize the reader with the overall landscape of CSS, including its main concepts, principles, applications, and areas of research. CSS investigates social complexity at all levels of analysis—cognitive, individual, group, societal, and global—through the medium of computation, as we will examine in greater detail in Chap. 1. This book is not intended as an advanced, specialized monograph to develop deep expertise.

The need for this book arose from the lack of unified treatment of the various areas of theory and research in CSS. As a consequence, those of us involved in teaching this new subject have been constrained to use a disparate library of readings without a single, unified framework. This book aims to be both comprehensive (include all major areas of CSS) and scientifically integrated by an overarching framework inspired by the paradigm of complex adaptive systems, as developed by Simon and his contemporaries in what may now be called the Founders's Generation (described in Chap. 1).

This project originated from the course on Introduction to CSS that has been taught at George Mason University for the past ten years. It is the core course in CSS, required of all students entering our graduate program in the Department of Computational Social Science. Initially, I taught the course, then other colleagues joined. Approximately ten students have taken the course each year, mostly from the CSS program, but also from other departments across the social sciences, computer science, environmental science, and engineering sciences.

This book is intended for two types of readers, which reflect the diverse student communities who have taken this course over the years. Some students will use it as a one-time, comprehensive exposure to the field of CSS. Other students might use it as foundation for further study through more advanced, specialized work in one or more of the areas surveyed here. This book should also be helpful to students preparing for their doctoral examination in CSS, as a review of basic ideas and a way to integrate knowledge.

The background assumed of the reader consists of some familiarity with one or more of the social sciences at a level equivalent to undergraduate study, basic knowledge of programming in any language (nowadays Python has become quite popular and is an excellent language for learning about computation), and some ability to follow mathematical modeling using logic, elementary probability, and basic calculus. Higher mathematics are unnecessary for introducing CSS.

The plan of the book is as follows: Chapter 1 provides an introduction, focusing primarily on the meaning of complex adaptive systems in social domains, including the significance of Herbert A. Simon's seminal theory and the paradigm it provides for CSS. This initial chapter also explains the main areas of CSS covered in this textbook, which are taken up in Chaps. 3 to 10. Chapter 2 provides a review of basic ideas in computing from a social science perspective, or computation as a paradigm for developing social science; it is *not* intended as a substitute for formal instruction on computation and programing for social scientists.

The following chapters cover major areas of CSS, corresponding to four distinct methodological approaches, as summarized in Sect. 1.6:

- Automated information extraction (Chap. 3)
- Social networks (Chap. 4)
- Social complexity:
  - Origins and measurement (Chap. 5)
  - Laws (Chap. 6)
  - Theories (Chap. 7)
- Social simulation:
  - Methodology (Chap. 8)
  - Variable-based models (Chap. 9)
  - Object-based (Chap. 10)

Each chapter contains a brief opening section introducing and motivating the chapter. This is followed by a section summarizing some of the history of CSS in the chapter's area, based on significant milestones. The purpose of these historical chronologies associated with each chapter's theme is to make the reader aware of significant scientific roots of the field of CSS, including its braided development with related disciplines; it does not provide a systematic history. Each chapter also includes a list of Recommended Readings, primarily intended as a guide for deepening understanding of each chapter, not as exhaustive bibliographies.

The style of the textbook attempts to strike a balance between an informal, reader-friendly, narrative tone, and a more formal tone that is necessary for highlighting rigorous concepts and results. Concept formation is a major emphasis, as is the statement of laws and principles from theory and research in quantitative social science, especially formal theory and empirically validated models. Along these lines, an effort is made, beginning in Chap. 2, to provide CSS with systematic, scientific, graphic notation that has been so sadly lacking in traditional social science.

This is done by adopting the Unified Modeling Language (UML) as a viable system for describing social complexity through graphic models that have powerful analytical meaning, as well as having direct correspondence with computation and code. Mathematical notation used in this book is standard and aims at maintaining consistency across chapters.

Finally, in terms of possible uses of this textbook, instructors may consider the following options. The ten chapters of this textbook are normally more than sufficient for a one-semester course, because some chapters will require more than one week to work through. Chapter 1 is best covered in a single session. Chapter 2 can easily be covered in two sessions, by dedicating the second session to UML. Chapters 4, 5, 6, 7, 9, and 10 can also each be covered in two sessions, by dividing the material into the main sections composing each chapter. Hence, another option is to use this textbook for a two-semester sequence, as is done in many other fields. This extended format would also permit more use of Recommended Readings, supplemented by additional bibliography, and spending more time analyzing examples to deepen understanding of concepts and principles. Readers are strongly encouraged to use the list of Recommended Readings to study the classic works, which are highlighted in the historical section at the beginning of each chapter.

This book has benefited from significant feedback from students, so I welcome future suggestions for corrections and improvements. I hope you, the reader, enjoy learning from this book at least as much as I have enjoyed writing it.

Washington, DC — Claudio Cioffi-Revilla
September 2013

Acknowledgements

During the past four decades I have benefited from scientific discussions with mentors, colleagues, and students who have influenced my research and teaching in Computational Social Science. Much of my original interest in the field came from discussions with Herbert ("Herb") A. Simon and members of the Triple-I Seminar on Complex Systems during the 1980s, including Elinor ("Lin") and Vince Ostrom and Harvey Starr from Indiana University, Dina A. Zinnes, Dick Merritt, Bob Muncaster, Jim Kuklinsky, and Mike Krassa from the Merriam Lab at the University of Illinois at Urbana-Champaign, and Bob Boynton from the University of Iowa. Discussions with Karl Deutsch, Ed Azar, Andy Scott, Harold Guetzkow, Bruce Russett, Hayward Alker, Raoul Narroll, Steve Wolfram, Larry Smarr, Benoit Mandeldrot, Ray Dacey, Martin Shubik, Dwain Meford, Jim Rosenau, Pierre Allan, Giorgio Natalicchi, Sam Kotz, and Kurt Johnson from the earlier phase of my academic career are still memorable. Craig Murphy, Doug Nelson, Chuck Taber, Kelly Kadera, Terry Clark, and Paul Pudiate were among my earliest students. When I moved to the University of Colorado at Boulder I learned a great deal from John Rundle and colleagues at the Colorado Center for Chaos and Complexity (C4), especially V. J. Gupta and Liz Bradley.

This textbook grows out of the interdisciplinary Program in Computational Social Science at George Mason University, which I founded in 2002 through joint teamwork with numerous students, faculty, staff, and administrators from the Mason campus. At the risk of unintentionally omitting someone, I wish to thank the many who have helped me in myriad ways: Giorgio Ascoli, Rob Axtell, Peter Balint, Jacquie Barker, Ernie Barreto, Andrea Bartoli, Jeff Bassett, Sheryl Beach, Jim Beall, Pete Becker, Tony Bigbee, Christina Bishop, Kim and Sharon Bloomquist, Gary Bogle, Annetta Burger, Joey Carls, Randy Casstevens, Gabriel Catalin Balan, Debbie Boehm-Davis, Dan Carr, Jack Censer, Guido Cervone, Kai-Kong Chan, Barbara Cohen, Marc Coletti, Jim Conant, Tim Conlan, Chenna Cotla, Julie Christensen, Andrew Crooks, Paul Cummings, David Davis, Ken De Jong, Dan Druckman, Bob Dudley, Debbie V. Duong, Kim Eby, Allan Falconer, Win Farrell, Tatiana Filatova, Kim Ford, Jennifer Fortney, Aaron Frank, Brendon Fuhs, Jim Gentle, Aldona Gozikowski, Omar Guerrero, Cathy Gallagher, Jack Goldstone, Jon

Gould, John Grefenstette, Beth Grohnke, Greg Guagnano, Renate Guilford, Tim Gulden, Ates Hailegiorgis, Joey Harrison, Melissa Hayes, Kingsley Haynes, Dee Holisky, Bill Honeychurch, Dan Houser, Chris Hyungsik Shin, Bob Jonas, Chris Jones, Mark Katz, Bill Kennedy, Matt Koehler, Maction Komwa, Dorothy Kondal, Raj Kulkarni, Mike Laskofski, Maciej Latek, Randy Latimer, Kate Leonard, Alex Levis, Collette Lawson, Ann Ludwick, Cynthia Lum, José Manuel Magallanes, Julie Mahler, Michelle Marks, David Masad, Steve Mastrofski, Kevin McCabe, Mike McDonald, Hugh McFarlane, Danny Menascé, Alan Merten, Tish Moreno, Michael Naor, Johnny Nelson, Jim Olds, Leslie Painter, Liviu Panait, Dawn Parker, Ann Palkovich, Sean Paus, Nicolas Payette, Carolyn Payne, Kathleen Pérez-López, Bianica Pint, Margaret Polski, Paul Posner, Steve Prior, Denise Quinto, Pris Regan, Cindy Roberts, Suzanne Robbins, Pedro Romero, Tom Rosati, Dave Rossell, Mark Rouleau (our first Ph.D. in CSS), Cathy Rudder, John Sacco, Mickey Satija, Tim Sauer, Laurie Schintler, Paul Schopf, Linda Schwartztein, Jagadish Shukla, Steve Scott, James Snead, Paul So, Arun Sood, Peter Stearns, Roger Stough, Jennifer Sturgis, Keith Sullivan, Burak Tanyu, Rhonda Troutman, Max Tsvetovat, Karen Underwood, Dick Wagner, Nigel Waters, Shandra Watson, Jane Wendelin, Steve Wilcox, Debbie Williams, Sarah Wise, David Wong, Chun-Yi Yang, Carol Zeeve, and Matt Zingraff. Their cumulative efforts have enabled Mason's Graduate Program in Computational Social Science (over a dozen courses in CSS, a Certificate, a Masters in Interdisciplinary Studies/CSS, and a Ph.D. degree), the Department of Computational Social Science, and the generative unit for CSS at Mason: the Center for Social Complexity (CSC).

I have learned much from co-authoring publications, developing new courses and grant proposals, discussing theory and research, and organizing events with a large and stimulating community of colleagues from around the world, including: Tef Abate, Petra Ahrweiler, Guillermo Algaze, Luís Antunes, Aruna Apte, George Atkinson, Fulvio Attinà, Scott Atran, Brent Auble, Tom Baerwald, Bill Bainbridge, Steve Bankes, Mike Batty, Ana Lucia Bazzan, Russ Bernard, Brian Berry, Ravi Bahvnani, Dmitri Bondarenko, Nathan Bos, Peter Brecke, Stuart Bremer, Cathy Cameron, Kathleen Carley, Cristiano Castelfranchi, John Casti, Lars-Erik Cederman, Fahmida Chowdhury, Alfred Cioffi, Wayne Clough, Helder Coelho, Louise Comfort, Rosaria Conte, Chet Cooper, Linda Cordell, Angela Corolla, Nuno David, Guillaume Deffaunt, Hiroshi Deguchi, Christophe Deissenberg, Jerry Dobson, David Dornish, Jim Doran, Massimo Drei, Julie Dugdale, Bruce Edmonds, Giorgio Einaudi, Carol and Mel Ember, Josh Epstein, Mike Fischer, Bill Fitzhugh, Bruno Frohlich, José Manuel Galán, Jianbo Gao, Michele Gelfand, Nigel Gilbert, Gary Goertz, Rebecca Goolsby, Nick Gotts, Ariel Greenberg, Steve Guerin, Alessandro Guidi, George Gumerman, Myron Gutmann, David Hales, Dirk Helbig, Matt Hoffmann, Barry Hughes, Luís Izquierdo,Wander Jager, Eric Jones, Steve Kaisler, Anna Kerttula, Dennis King, Alan Kirman, Jürgen Klüver, Tim Kohler, Nick Kradin, Arie Kruglanski, Larry Kuznar, Steve Lansing, Efraim Laor, Randy Latimer, Steve Lekson, Nicola Lettieri, Mark Lichbach, David Lightfoot, Fred Liljeros, Corey Lofdahl, Urs Luterbacher, Thomas Lux, Patty Mabry, Charles Macal, Ed MacKerrow, Michael Macy, Greg Madey, Artemy Malkov, Joyce Marcus, Jack

Meszaros, Manny Midlarsky, Jeff Millstein, Byong Won Min, Harold Morowitz, Scott Moss, Akira Namatame, Dana Nau, Martin Neumann, Michael North, Andrzej Nowak, Sean O'Brien, Paul Ormerod, John Padgett, Mario Paolucci, Domenico Parisi, Peter Peregrine, Peter Perla, Gary Polhill, Brian Pollins, Denise Pumain, Rodolfo Ragionieri, Bill Rand, Dwight Read, Colin Renfrew, Bob Reynolds, Fred Roberts, J. Daniel Rogers, Juliette Rouchier, Dieter Ruloff, Jerry and Paula Sabloff, John Salerno, David Sallach, Lena Sanders, Todd Sandler, Antonio Sanfilippo, Dez Saunders-Newton, Vittorio Scarano, Steve Schlosser, Phil Schrodt, Lee Schwartz, Frank Schweitzer, Payson Sheets, Andrew Sherratt, Carl Simon, Ian Skoggard, Ricard Solé, Jim Spohrer, Detlef Sprinz, Flaminio Squazzoni, Gene Stanley, John Sterman, Christina Stoica, Rick Stoll, Gary Strong, Lee Schwartz, David Sylvan, Rein Taagepera, Keiki Takadama, John Tangney, Takao Terano, Pietro Terna, Rita Teutonico, Jim Thomas, Qing Tian, Klaus Troitzsch, Peter Turchin, Alex Vespignani, Mitch Waldrop, David Warburton, Paul Werbos, Jon Wilkenfeld, and Peyton Young.

A graceful invitation from my colleague and friend, Shu-Heng Chen, to deliver the 2011 Herbert A. Simon Lecture Series in Computational Social Science at National Chengchi University in Taipei, Taiwan, provided a unique opportunity to organize my ideas for this textbook. A preview of this textbook was provided in March 2012 at the invitation of the Center for the Study of Complex Systems at the University of Michigan. I am grateful to Robert Axelrod, John Holland, Scott Page, Rick Riolo, and their students for sharing their insights and suggestions.

Some of the examples or modeling applications discussed in this book originated or came into focus through discussions with members of the government policy and analytical community. None of them bears any responsibility for my interpretations or inferences.

Students from the Fall 2012 session of CSS 600—Introduction to CSS—helped as I finalized the outline of this textbook. I am especially grateful to Gary Bogle, Tom Briggs, Annetta Burger, Paul Cummings, José Manuel Magallanes, and Dan Pryce. Several chapters of this textbook were also used while lecturing at the Lipari International Summer School in Computational Social Science, now in its 5th year. I am grateful to students and invited faculty, including David Beaver, Kathleen Carley, Alfredo Ferro, Giovanni Giuffrida, Carlo Pennisi, Alessandro Pluchino, Kalev Leetaru, Roy Lindelauf, Huan Liu, Roel Popping, Raghu Ramakrishnan, Marc Smith, Philip Schrodt, V.S. Subrahmanian, Alberto Trobia, and Calogero Zarba.

I am especially grateful for input on various chapters received from Dan Rogers, Linda Cordell, Sean Luke, Nazli Choucri, Bill Kennedy, Siggy Scott, and Joey Harrison. I also wish to thank Jean N. Cioffi, Dorothy Kondal, and Nancy Turgeon for careful editing and assistance with proofreading.

Support received from the US National Science Foundation and the Office of Naval Research, as well as from the Center for Social Complexity and the Provost's Office at George Mason University, especially intellectual support received from Provost Peter Stearns, is gratefully acknowledged.

I also wish to thank Wayne Wheeler and Simon Rees, my editors at Springer, for their encouragement and patience. They are among the most professional, persevering, and pleasant editors I have worked with.

Images used in this textbook were produced by me or obtained from NASA, PublicDomainPictures.net, Wikipedia, Malteser International, and faculty webpages without copyrights.

Contents

Acronyms

ABM	Agent-based model
ACE	Agent-based computational economics
ACM	Association for Computing Machinery
AI	Artificial intelligence
AND	Boolean conjunctive operator
BDI	Beliefs, desires, intentions
CA	Cellular automaton or automata
CAMEO	Conflict and Mediation Event Observations
CAS	Complex adaptive system
CASOS	Center for Computational Analysis of Social and Organizational Systems, Carnegie Mellon University
CCDF	Complementary cumulative density function (also c.c.d.f.)
CDF	Cumulative density function (also c.d.f.)
CIDCM	Center for International Development and Conflict Management, University of Maryland
CIKM	Conference on Information and Knowledge Management of the ACM
CMU	Carnegie Mellon University
COA	Course of action
COPDAB	Conflict and Peace Data Bank
CPU	Central processing unit
CSC	Center for Social Complexity, George Mason University
CSS	Computational Social Science
CSSSA	Computational Social Science Society of the Americas
CSSN	Computer-supported social networks
DARPA	Defense Advanced Research Projects Agency
DDR3 SDRAM	Double-data-rate three synchronous dynamic random access memory
DYNAMO	DYNAmic MOdels
EC	Evolutionary computation
ECPR	European Consortium for Political Research

ECML-PKDD	European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases
EOS	Evolution of Organized Society project, University of Essex
EPA	Evaluation, potency, activity. Dimensions of Osgood's semantic space
ERG	Exponential random graph
EU	European Union
FEARLUS	Framework for the Evaluation and Assessment of Regional Land Use Scenarios
FIFO	First-in-first-out
FILO	First-in-last-out
FORTRAN	FORmula TRANslation
GB	Gigabyte
GCM	General Circulation Model
GDELT	Global Data on Events, Location, and Tone
GeoMASON	Geospatial MASON
GHz	Gigahertz
GIS	Geographic Information System
GPU	Graphic processing unit
GUI	Graphic user interface
HMM	Hidden Markov model
HPC	High-performance computing
HRAF	Human Relations Area Files, Yale University
ICPSR	Interuniversity Consortium for Political and Social Research
ICR	Institute for Communications Research, University of Illinois at Urbana-Champaign
IEEE	Institute of Electrical and Electronic Engineers
INSNA	International Network for Social Network Analysis
I/O	Input-output
IPCC	Intergovernmental Panel on Climate Change
ISIMADE	International Symposium on Intelligent Multimedia and Distance Education
ISS	International Space Station
JVM	Java virtual machine
KWIC	Keywords in context
KWOC	Keywords out of context
kya	Thousands of years ago
LEO	Low Earth orbit
LIFO	Last-in-first-out
LILO	Last-in-last-out
LISP	LISt Processing
LOC	Lines of code
LRD	Long-range dependence
LUCC	Land-Use and Cover Change
MAS	Multi-agent system or systems

MASON	Multi-Agent Simulator of Networks or Neighborhoods
MC	Monte Carlo
MDS	Multi-dimensional scaling
MDIVVA	Motivate-design-implement-verify-validate-analyze
MIT	Massachusetts Institute of Technology
MINUIT	Numerical minimization computer program
MLE	Maximum likelihood estimate, estimator, or estimation
M2M	Model-to-model
NAACSOS	North American Association for Computational Social and Organizational Sciences
NASA	National Aeronautics and Space Administration
NATO	North Atlantic Treaty Organization
NER	Named entity recognition
NIST	National Institute of Standards and Technology
NRR	Normal relations range
NSF	National Science Foundation
NVAC	National Visualization Analytics Center, PNNL
OCR	Optical character recognition
OR	Boolean disjunctive operator
ORA	Entity extraction algorithm by CASOS
OMG	Object Management Group
ONR	Office of Naval Research
OO	Object-oriented
OOM	Object-oriented model or modeling
OOP	Object-oriented program or programming
PDF	Probability density function (also p.d.f.)
PNAS	Proceedings of the National Academy of Sciences of the USA
PNNL	Pacific Northwest National Laboratory, Department of Energy
PPNB	Pre-Pottery Neolithic B period
PRNG	pseudo-random number generator
RAM	Random access memory
RNG	Random number generator
SAS	Statistical Analysis System
SD	System dynamics
SDC	Size, development, and capability
SEQAND	Boolean sequential conjunctive operator
SES	Socioeconomic status
SIAM	Society for Industrial and Applied Mathematics
SIGKDD	Special Interest Group on Knowledge Discovery and Data Mining of the ACM
SIMPEST	Simulation of Political, Economic, Social, and Technological Systems
SIMPLE	Simulation of Industrial Management Problems with Lots of Equations
SIMPOP	SIMulation of POPulation project, University of Paris-Sorbonne

SNA	Social network analysis
SOCPAC	A FORTRAN IV program for structural analysis of sociometric data
SPSS	Statistical Package for the Social Sciences
SSRC	Social Science Research Council
SSRN	Social Science Research Network
STELLA	System dynamics simulation system
TABARI	Textual Analysis by Augmented Replacement Instructions
TBJ	Truth, beauty, and justice
TRIAL	Technique for Retrieval of Information and Abstracts of Literature
UAV	Unmanned autonomous vehicle
UCINET	University of California-Irvine social network analysis software
UCLA	University of California-Los Angeles
UML	Unified Modeling Language
UN	United Nations
URL	Uniform resource locator
US	United States
USSR	Union of Soviet Socialist Republics
VENSIM	System dynamics simulation system
WWW	World-Wide Web
XOR	Boolean exclusive disjunctive operator

List of Figures

List of Tables

1 Introduction

The goal of this chapter is to present foundational concepts and some operational definitions in the field of Computational Social Science (CSS for short) by introducing the main assumptions, features, and research areas. A key feature of CSS is its interdisciplinary nature. Computational modeling enables researchers to leverage and integrate knowledge from many different disciplines, not just the social sciences. This chapter also provides an overview of the whole textbook by providing a "peek" into each chapter. The purpose is not to enter into many details at this stage, but to provide a preview of some of the main ideas examined in subsequent chapters.

One of the key challenges in the field of Computational Social Science is that several relatively subtle or complicated ideas need to be introduced simultaneously. Social complexity, complex adaptive systems, computational models, and similar terms are introduced in this chapter, and later elaborated upon in greater depth. What we need for now are some initial concepts so that we may get started in establishing foundations. There is no attempt in this chapter to provide an exhaustive treatment of each and every term that is introduced.

1.1 What Is Computational Social Science?

The origin of social science—in the *pre*-computational age—can be traced back to Greek scholars, such as Aristotle, who conducted the first systematic investigations into the nature of social systems, governance, and the similarities and differences among monarchies, democracies, and aristocracies. In fact, Aristotle is often considered the first social science practitioner of comparative social research. Modern social science, however, is usually dated to the 17th century, when prominent French social scientists such as Auguste Comte first envisioned a natural science of social systems, complete with statistical and mathematical foundations and methods to enhance traditional historical and earlier philosophical approaches. Since then, the social sciences have developed a vast body of knowledge for understanding human and social behavior in its many forms (Bernard 2012). This is how modern anthro-

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_1,

pology, economics, political science, psychology, and sociology—the so-called Big Five (Bernard 2012; Horowitz 2006; Steuer 2003)—were born four centuries ago.

The new field of **Computational Social Science** can be defined as the interdisciplinary investigation of the social universe on many scales, ranging from individual actors to the largest groupings, through the medium of computation. This working definition is somewhat long and will be refined later as we examine many topics involved in the practice of CSS and the variety of computational approaches that are necessary for understanding social complexity. For example, the " many scales" of social groupings involve a great variety of organizational, temporal, and spatial dimensions, sometimes simultaneously. In addition, computation or computational approaches refer to numerous computer-based instruments, as well as substantive concepts and theories, ranging from information extraction algorithms to computer simulation models. Many more will be invented, given the expansive character of computational tools. In short, CSS involves a vast field of exciting scientific research at the intersection of all social science disciplines, applied computer science, and related disciplines. Later in this chapter we will examine some analogues in other fields of knowledge.

Another useful clarification to keep in mind is that CSS is not limited to Big Data, or to social network analysis, or to social simulation models.[1] That would be a misconception. Nor is CSS defined as any one of these relatively narrower areas. It comprises all of these, as well as other areas of scientific inquiry, as we will preview later in this chapter.

1.2 A Computational Paradigm of Society

Paradigms are significant in science because they define a perspective by orienting inquiry. A paradigm is not really meant to be a theory, at least not in the strict sense of the term. What a paradigm does is provide a particularly useful perspective, a comprehensive worldview (*Weltanschauung*). Computational social science is based on an **information-processing paradigm** of society. This means, most obviously, that information plays a vital role in understanding how social systems and processes operate. In particular, information-processing plays a fundamental role in explaining and understanding social complexity, which is a subtle and deep concept to grasp in CSS as well as in more traditional social science.

The information-processing paradigm of CSS has dual aspects: substantive and methodological. From the *substantive* point of view, this means that CSS uses information-processing as a key ingredient for explaining and understanding how society and human beings within it operate to produce emergent complex systems. As a consequence, this also means that social complexity cannot be understood

[1] Big Data refers to large quantities of social raw data that have recently become available through media such as mobile phone calls, text messaging, and other "social media," remote sensing, video, and audio. Chapter 3 examines CSS approaches relevant to Big Data.

without highlighting human and social processing of information as a fundamental phenomenon. From a *methodological* point of view, the information-processing paradigm points toward computing as a fundamental instrumental approach for modeling and understanding social complexity. This does not mean that other approaches, such as historical, statistical, or mathematical, become irrelevant. On the contrary, computational methods necessarily rely on these earlier approaches—and other methodologies, such as field methods, remote sensing, or visualization analytics—in order to add value in terms of improving our explanations and understanding of social complexity. In subsequent chapters we shall examine many examples pertaining to these ideas. For now, the best way to understand the information-processing paradigm of CSS is simply to view it as a powerful scientific perspective that enables new and deep insights into the nature of the social universe.

1.3 CSS as an Instrument-Enabled Science

CSS is by no means alone in being an *instrument-enabled scientific discipline*. Consider *astronomy*, a science that was largely speculative and slow in developing before the invention of the optical telescope in the early 1600s. What Galileo Galilei and his contemporaries discovered through the use of telescopes enabled astronomy to become a real science in the modern sense. In particular, the optical telescope enabled astronomers to see and seek to explain and understand vast areas of the universe that had been previously unknown: remote moons, planetary rings, sun spots, among the most spectacular discoveries. Centuries later, the radio telescope and infrared sensors each enabled subsequent revolutions in astronomy.

Or, consider *microbiology*, prior to the invention of the microscope in the late 1600s. Medical science was mostly a descriptive discipline filled with untested theories and mysterious diseases that remained unexplained by science. The microscope enabled biologists and other natural scientists, such as Anton von Leeuwenhoek and Louis Pasteur, to observe and explore minuscule universes that were entirely unknown. Later it was discovered that the majority of living species are actually microorganisms. Centuries later, another kind of microscope, the electron microscope, enabled biologists and other scientists to see even smaller scales of life and beyond, down to the molecular and atomic levels. Nano-science was also born as an instrument-enabled field, which also includes an engineering component, as does biology in the form of bioengineering.

Linguistics is a human science that experienced a similar phenomenon, through the application of mathematics. Prior to mathematical and computational linguistics the study of human languages was more like a humanistic discipline, where various interpretations and traditions contended side by side without each generation knowing much more than the previous, since the main tradition was to offer new perspectives on the same phenomena—not exploring and attempting to understand entirely new phenomena. Mathematical and computational linguistics propelled the discipline into the modern science that it is today.

Much the same can be said of *physics*. Greek and medieval scientists viewed the physical universe as consisting of substances with mysterious "essential" properties, such as a heavy object belonging at rest—a state caused by its essence. Physics became a modern, serious science through the application of mathematical instruments, especially the infinitesimal calculus of Newton and Leibniz, in addition to the empirical method. The empirical approach alone would have been insufficient, since theory was enabled by mathematical structures responsible for the main thrust of the hypothetic-deductive method.

What all of these and numerous other cases share in common in the long and well-documented history of science is quite simple: in every culture, science is always enabled and revolutionized by instruments, not just by new concepts, theories, or data. Instruments are the main tools that science uses to create new science. As computers have revolutionized all fields of science since the invention of digital computing machines in the 1950s, and many humanities disciplines in recent years (from the fine arts to history), so the social sciences have been transformed by computing. Moreover, such transformations are irreversible, as has been the case for other instruments in other fields. CSS is in great company; it is not alone in being an instrument-enabled science.

1.4 Examples of CSS Investigations: Pure Scientific Research vs. Applied Policy Analysis

Another stimulating characteristic of CSS is that it encompasses both pure science and policy analysis (applied science). It is not a purely theoretical science such as, for instance, mathematical economics, rational mechanics, or number theory.[2] This means that CSS seeks fundamental understanding of the social universe for its own sake, as well as for improving the world in which we live. In fact, as we discuss later in this chapter, CSS has a lot to do with improvement of the human condition, with building civilization. These are obviously large claims, but they are not different from those found in other scientific disciplines that attempt to better understand the world both for its own sake and to improve it. It is a misconception to think that pure/basic science and applied/engineering science are somehow opposed or incompatible pursuits. Again, the history of science is replete with synergies at the intersection of pure and applied knowledge. Examples of pure scientific research in CSS include:

1. Investigating the theoretical sensitivity of racial segregation patterns in societies of heterogeneous agents.
2. Modeling how leaderless collective action can emerge in a community of mobile agents with radially distributed, robot-like vision and autonomous decision-making.

[2] Number theory actually has very concrete application in cryptology, a highly applied field in national security and internet commerce.

3. Understanding how crowds may behave in a crisis when interacting with first responders and their respective support systems.
4. A project on the impact of natural extreme hazards of a generic variety to assess risk and the potential for causing catastrophes and plan for mitigation.

A parallel set of applied policy examples would read more or less as follows:

1. A high-fidelity agent-based model of New York City neighborhoods to mitigate racial segregation without relying exclusively on laws.
2. Modeling how the Arab spring may have originated, based on an empirically calibrated social network model of countries in the Middle East and North Africa.
3. Understanding how the population of New Orleans responded when Hurricane Katrina hit the city and first responders and their respective support systems were activated.
4. A geospatially referenced agent-based model of the Eastern coast of the United States to prepare for seasonal hurricanes and changing weather patterns caused by climate change.

The use of proper nouns is often (not always!) a give-away in applied policy analysis. However, there is more to applied CSS than the use of proper nouns. In particular, high-quality applied CSS must add value to other policy analysis approaches—it must provide insights or knowledge significantly and demonstrably beyond that which can be provided by other analytical tools. Another distinctive feature of applied CSS analysis is that it contributes to a better understanding of situations that are too complex to analyze by other methods, even when prediction or forecasting is not involved. For example, a good use of applied CSS might be the use of computer simulations to better understand and prepare for unintended consequences—or what are called *negative externalities*—of policies.

The pure-applied synergy in science is also present in CSS in another respect: this has to do with pure research that occasionally generates applications for improving policies, and, conversely, a so-called *wicked problem* in the policy arena inspiring fundamental research questions in pure research. Examples of the former kind of synergy (basic science improving policy) would include:

- Better understanding how crowds of panicky individuals "flow" in an emergency in order to improve building design and evacuation procedures.
- Comparing formal properties of organizational structures to improve the workplace.
- Inventing a new algorithm to improve security of communication in complex infrastructure systems and their management interface with humans.
- Deeper understanding of the formal properties of distributions to design better queuing systems, such as those used by air traffic controllers and similarly complex systems.

Conversely, examples of the latter kind (policy needs informing basic research) would include:

- Developing the social theory of communication in racially mixed communities out of the policy need to create a high-fidelity model of a refugee camp.
- Deepening our understanding of complex network structures based on the need to model transnational organized crime in trafficking of persons.

- Improving a theory of origins of civilization while attempting to improve anti-looting laws and regulations that govern world heritage archaeological sites.
- Working on formulating and testing a new theory of learning in individuals and collectives of agents while trying to revise public policy in health care and education.

The synergies highlighted by these examples are not contrived or invented for pedagogical purposes. They are real in the sense that they have either already occurred, or are likely to occur in the not-so-distant future. In other words, they are not purely notional examples. Moreover, such synergies are likely to grow as the field develops through more mature stages—as has happened in many other areas of science.

The powerful and fascinating synergy between science and policy notwithstanding, it is also fair to say—indeed, be emphasized—that basic scientific research and applied policy analysis are different activities along numerous dimensions, such that they generate different professions:

Expectations: Basic science is expected to produce new knowledge and understanding, whereas applied policy analysis is more results-oriented in a practical sense. People built bridges across rivers centuries (perhaps millennia) before the fundamental laws of mechanics were discovered.

Training: Scientists and practitioners train in different concepts, tools, and methodologies, even when they may share training in some common disciplines, such as in the use of simple statistics.

Incentives: Pure scientists and policy analysts have different incentives, such as academic rewards for the former and promotions to higher organizational roles for the latter.

Facilities: Pure science is best conducted in labs and research centers; think tanks are specialized venues for conducting policy analysis. Both kinds of venues can be academic, private, or governmental; what matters is the main mission and associated support infrastructure.

Publicity: Pure scientific research is most frequently highly publicized, especially when it touches on public issues, such as climate change, health, communication, the economy, or national security. Moreover, open sources are more typical of academic CSS research, except when researchers impose a temporary embargo in order to publish first. Applied policy research is often less public, especially when it concerns sensitive information pertinent to public issues, or when private consulting firms protect intellectual property by requiring and enforcing nondisclosure agreements.

Some features that are common to both pure and applied research in CSS include the need for terminological clarity (not the "Tower of Babel" decried by Giovanni Sartori), systematic concept formation, respect for evidence, rigorous thinking, and thorough documentation. Also, in both areas one can find excellent, mediocre, and outright awful work—"the good, the bad, and the ugly," as in the proverbial phrase.

Throughout this textbook we will encounter cases of both pure CSS research as well as applied policy applications. Similarities and differences between the two are significant and instructive on the role of each and the synergy between the two orientations or activities.

1.5 Society as a Complex Adaptive System

Society is often said to be complex. What does that mean? In this section we examine this idea for the first time, developing deeper understanding in subsequent chapters.

1.5.1 What Is a CAS in CSS?

At the very beginning of this chapter we mentioned complex adaptive systems as being one of the key, fundamental ideas in the foundations of CSS. For now, we can define a **complex adaptive system** as one that changes its state, including its social structure and processes, in response to changing conditions. Later, especially in Chaps. 5–7, we will develop more rigorous definitions. A cybernetic system is an instance of a rudimentary CAS, whereas a system of government, an ecosystem, an international regulatory agency (such as World Bank or the International Monetary Fund), or a complex organization (such as NASA or the Intergovernmental Panel on Climate Change, IPCC), are more complete examples.[3] An essential aspect of this initial definition is to note that a complex adaptive system operates through **phase transitions** (significantly different states and dynamics) in the operating regime of the system in order to maintain overall performance in the face of changing environmental conditions or evolving goals or changes in resources.

A family is a social organization that can be viewed as a complex adaptive system, one based on kinship relations that undergo numerous changes throughout the life cycle of individuals who are members of the family, when viewed as a human grouping. Everyone in the family ages, and some mature successfully into old age, experiencing many different situations, acquiring new knowledge, in the face of numerous opportunities and challenges. In spite of many changes, the overall system of kin-based relations in some families can endure for decades; in other cases that is not the case and the system breaks down. Adaptation in the history of a given family manifests itself in numerous ways: children grow up and must adapt to going to school; parents might change jobs or occupations, having to adapt to labor market conditions or to changing priorities; social mobility also requires adaptation, perhaps to new norms or new locations; making and losing friends also requires adaptation. Adaptation is common and frequent in many social systems because internal components and relations are willing and able, even required, to change in order for the open systems to endure, sometimes improving or prospering.

Adaptation in social systems is best seen as a *multi-stage process*, not as a single event. As such, several occurrences are required for adaptation to operate successfully. We may view this as consisting of several events, which later we will refine in

[3]The example of a cybernetic system as a CAS is not by chance. In fact, the Greek etymology of the term government, or $\gamma\upsilon\beta\epsilon\rho\nu\eta'\tau\eta\varsigma$ (kybernētēs), means the rudder or steering mechanism in a ship. It's the same in Italian (*governo*), Spanish (*gobierno*), French (*government*), and in other languages.

more formal ways. First, the system, or the actors within the system, must be aware that there is a *need* to adapt—to undertake adaptive behavior. Second, there must be an *intent* to adapt, which is separate from the recognized need to adapt. Third, there must be *capacity* to adapt, since adaptation costs in terms of resources, be they tangible or intangible. Finally, adaptive behavior must be *implemented* in some form, which may involve executing plans or overcoming various kinds of difficulties and challenges. A key idea to understand regarding adaptation in social systems is that it is never automatic or deterministic, at least in the most interesting or nontrivial situations. Whether a person, a family, a group, an economy, an entire society, a whole nation, or even a global society adapts to change, such a process always consists of several stages.

A particularly noteworthy aspect of complex adaptive systems from a *computational perspective* is the key role played by information-processing:

1. Information is necessary for assessing the need for a complex system to require adaptation.
2. The activity of determining resources also requires information.
3. Information flows in the form of interpersonal and inter-group communication when adaptation is decided on, prepared for, implemented, or subsequently monitored for its effects on restoring a viable state for the system.

This is obviously a sparse and simple summary of the role of information in CAS, which serves to highlight the usefulness of the information-processing paradigm discussed earlier. Information-processing is pervasive and critical in complex adaptive systems; it is not a phenomenon of secondary importance. An interesting aspect of information in CAS is that it has many other interesting properties, as well as insightful connections to other essential ideas in CSS, such as complexity, computability, and sustainability, as we will examine later.

1.5.2 Tripartite Ontology of Natural, Human, and Artificial Systems

Another important distinction in CSS is among natural, human, and artificial systems—an ontological or categorical distinction that is different or does not exist at all, at least not to the same degree, in other fields of knowledge. The first computational social scientist to introduce this idea of a tripartite classification of entities was Herbert A. Simon, who used it as foundation for his theory of artifacts and social complexity through the process of adaptation. We will examine this soon, but the tripartite distinction is needed now. Complex adaptive systems of interest in CSS often combine all three categories of systems, so understanding the composition of each, as well as their similarities and differences, is important before entering more theoretical territory.

1. A **natural system** consists of biophysical entities and dynamics that exist in nature, mainly or completely independent of humans and their artifacts. Common examples are wilderness landscapes, animals other than humans, regional ecosys-

tems, and the biochemistry of life, including the biology of the human brain as a natural organ (*not* just mental phenomena).[4]

2. A **human system** is an individual person, complete with thoughts and body. Decision-makers, actors, agents, people, and similar terms denote human systems. The complexity-theoretic perspective highlights the human ability to create artifacts.
3. An **artificial system** is one conceived, designed, built, and maintained by humans. Artificial systems consist of engineered or social structures that act as adaptive buffers between humans and nature.

These initial conceptual definitions serve as building blocks that for now are sufficient for our initial purpose of establishing foundations. We shall return to these ideas to develop a better understanding of their properties and interrelationships.

1.5.3 Simon's Theory of Artifacts: Explaining Basic Social Complexity

Laws describe; theories explain. Having presented and discussed the first conceptual building blocks, now our main task is to move forward by providing an initial statement of Herbert A. Simon's theory of artifacts for providing an initial explanation of social complexity. Simon presented most of these ideas in his classic monograph, *The Sciences of the Artificial*, which first appeared in 1969, followed by a third and last edition in 1996.

From the previous ideas, it is important to note that artifacts exist because they have a function: they serve as adaptive buffers between humans and nature. This is the essence of **Simon's theory of artifacts and social complexity**. Humans encounter challenging and often complex environments, relative to their own simple abilities or capacities. In order to adapt to these circumstances, and not be overwhelmed by or succumb to them, humans pursue the strategy of building artifacts that enable their goals.

- Roads were first invented for moving armies and other military and political personnel from one location to another. They were also used for commercial and communications purposes. Without a proper road it is either very difficult or impossible to achieve such goals.
- Bureaucratic systems, and in some cases writing (e.g., Mesopotamia, China), were first created for maintaining records related to the governance and economy of a city. This enabled the first urban populations to attain the goals of becoming established and developed.

[4]The wording here is intentionally and necessarily cautious and precise. The paradigm being presented here separates humans from the rest of nature, based on the human ability to build artifacts, some of which are used to build other artifacts, especially intelligent, autonomous artifacts, using mental, cognitive, and information-processing abilities that are far more complex than those found in any other natural living organism. Ants might build colonies, corals build reefs, bees build hives, beavers build dams, but none of these or other examples of "animal-made artifacts" compares to human artifacts.

- The first large aqueducts, built by the Romans, required careful planning, engineering, and maintenance in order to provide water for large urban populations located at great distances from the sources (springs, rivers, lakes, or reservoirs).
- The International Space Station (ISS) is an engineering structure of unprecedented complexity, operating in the challenging environment of space, managed by a ground crew in coordination with the station's crew.

As already suggested by the previous examples, the artifacts that humans have been building for thousands of years, across all societies, can be **tangible** (engineered, i.e., physical) or **intangible** (organizational, i.e., social), as required by the goals being sought. Some adaptive strategies require tangible, engineered artifacts, such as dwellings, bridges, roads, and various kinds of physical infrastructure systems. At other times, an adaptive strategy may require planning for and creating an organization, such as a governing board or committee, that is to say, a social system of a given size and complexity to enable attainment of the goal being pursued.[5]

A fascinating aspect of this tightly coupled *synergy* between tangible and intangible, or engineered and organizational artificial systems, is that they often require each other—as in a *symbiotic relationship between humans and their artifacts*, where the latter enable human attainment of desired goals. This feature of social complexity is supported by historical and contemporary observation. To build a road or a bridge it is also necessary to create teams of workers supervised by managers, who depend on supply chains for the provision of building materials and other necessities: the tangible artifact (bridge) cannot be built without the intangible one (organization). Modern cities provide another excellent example of the same symbiotic relationship between engineered and social artifacts. The complex infrastructure that supports the life of humans in cities (as opposed to cave dwellers) requires numerous, specialized buildings and artificial systems—especially when cities are built in mostly inhospitable environments. This was also true of the earliest cities, which were supported by an organizational bureaucracy of managers, city workers, and other social components, working in tandem as a coupled socio-technological system to support urban life. For example, the capital of the USA, Washington, is built on a swamp, as is the Italian city of Venice. Both are enabled by physical and organizational infrastructure.

In sum, what does Simon's theory *explain?* It explains *why* artifacts exist, *why* humans build artifacts, and the fact that artifacts are *adaptive strategic responses* for solving the many challenges faced by humans in societies everywhere since the dawn of civilization.[6]

---

[5]This idea prompted Simon to suggest—in *The Sciences of the Artificial*—that social scientists, lawyers, and engineers should undergo university-level training of a similar kind, perhaps under a common College of the Artificial Sciences.

[6]Herbert A. Simon's work in the social sciences is widely known for its contributions to the study of organizations and bureaucracy. In computer science his work is equally well known for contributions to artificial intelligence and related areas. His theory of social complexity grew out of an interdisciplinary interest across these domains.

1.5.4 Civilization, Complexity, and Quality of Life: Role of Artificial Systems

Simon's theory of artifacts and adaptation goes a long way toward explaining the genesis and development of social complexity. It also explains important aspects of the same patterns that endure to this very day and will likely continue into the future. Humans everywhere pursue goals that are often sought in challenging environments, so in order to accomplish those goals they build artifacts—both engineered and social systems that are tangible and intangible, respectively.

However, thus far the story is incomplete, because sometimes humans seek goals that are *not* necessarily linked to challenging environments. For example, they may already live in a city that is quite viable, but they simply wish to live in a better way, such as enjoying better services and amenities, living longer or more comfortably, or enjoying culture and the fine arts. An additional, essential ingredient for developing a more complete theory of social complexity, one that explains a broader range of social complexity, is based on the empirical observation that *humans everywhere prefer to live a better life*. This is also a purpose of government: "The care of human life and happiness, and not their destruction, is the first and only legitimate object of good government" (Thomas Jefferson, American President, 1809).

A significant variation on the very same theme would be, for example, to wish that their descendants or friends enjoy a higher quality of life. The pursuit of a higher quality of life is a goal for many humans, which may occur independent of or in combination with taming a given environment. The strategic adaptive response is the same or isomorphic: artificial systems are conceived, planned, built, and maintained in the form of physical or social constructs. Complexity in all these forms increases in each case. Therefore, *both* challenging environments *and* human aspirations—and quite frequently the interaction of both—cause social complexity in a generative sense.

Sometimes complex systems come and go in a transient way; at other times they become permanent artifacts that can endure for very long periods of human history. Systems of government, infrastructure systems, monetary systems, and cultural norms provide examples of long-term artifacts that have increased in complexity over the millennia. Civilization is the result of this process, from the theoretical perspective of CSS. The dawn of civilization in all parts of the world where humans have created and developed social complexity is marked by the earliest engineered and organizational artifacts. Contemporary civilization in the 21st century is no different from the earliest civilizations, as seen from this universal theoretical perspective. Societies in the earliest days of Mesopotamia, China, South America, and Mesoamerica built the first irrigation canals, structures for communal worship, villages, towns and cities, the earliest infrastructure systems and systems of government and bureaucracies that supported them. All these artificial systems and many others that have since been invented persist to this day, and spacefaring civilization—if we manage to launch and mature it—will demonstrate comparable patterns in the evolution of social complexity.

Information-processing, goal-seeking behavior, adaptation, artifacts—engineered as well as organizational—and the resulting social complexity that they cause are the main ingredients of this interdisciplinary theory. Its purpose is to explain how and why natural, human, and artificial systems interact in the creation of history. The theory is *causal*, in a strict scientific sense, because it proposes an empirically demonstrable process that links together—not in a superficial correlational way devoid of causation—the elements thus far presented in this chapter and examined in greater detail across areas of CSS.

1.6 Main Areas of CSS: An Overview

Computational social science is an interdisciplinary field composed of areas of concentration in terms of clusters of concepts, principles, theories, and research methods. Each area is important for its own sake, because each represents fertile terrain for conducting scientific inquiry, as basic science as well as policy analysis. In addition, these areas can build on each other and be used synergistically, as when network models of social complexity are used in simulation studies, or through many other possible combinations of scientific interest.

The chapters of this book are dedicated to each of these areas, which we will now survey by way of introduction. The main purpose in this section is to provide an overview, not a detailed presentation of each area. By way of overview, it should be mentioned that these areas of CSS are also supported by statistical and mathematical approaches, and in some cases other methodologies as well, such as geospatial methods, visualization analytics, and other computational fields that are valuable for understanding social complexity.

1.6.1 Automated Social Information Extraction

CSS is an interdisciplinary field where data play numerous and significant roles, similar to those in other sciences. The area of automated information extraction refers to computational ideas and methodologies pertaining to the creation of scientifically useful social information based on raw data sources—all of which used to be done manually. Other names for this area of CSS might be computational content analysis, social data analytics, or socio-informatics, in a broad sense. For example, whereas in an earlier generation social scientists would gather data from sources such as census records, historical sources, radio broadcasts, or newspapers and other publications, today much of the work that takes place in order to generate social science research data is carried out by means of computational tools. As we will see, these tools consist of computational algorithms and related procedures for generating information on many kinds of social, behavioral, or economic patterns.

Social information extracted through automated computational procedures has dual use in CSS. For instance, sometimes it is used for its own sake, such as for analyzing the content of data sources in terms of affect, activity, or some other set of dimensions of interest to the researcher. An example of this would be a study to extract information concerning the political orientation of leaders or other governmental actors based on computational content analysis of speeches, testimony before legislative committees, or other public records.

Besides being used for analyzing the direct content of documents and other sources, information extraction algorithms can also be used to model networks and other structures present in raw data, but impossible to detect through manual procedures performed by humans. An example of this would be a model of organized crime organizations and their illegal activities, based on computational content analysis and text mining of court cases and other evidentiary legal documents that describe individuals, dates, locations, events, and attributes associated with criminal individuals. Another example would be automated information extraction applied to modeling correlations across networks, based on Internet news websites.

An extension of automated information extraction could also be used for building computer simulation models that require high fidelity calibration of parameters, such as models of opinion dynamics, international trade, regional conflicts, or humanitarian crises scenarios. The extraction of geospatial social data through computational algorithms represents a significant step forward in the development of CSS.

These and other examples illustrate how automated information extraction is sometimes seen as a foundational methodology in CSS: it can be used for developing models and theories in all of the other main areas of CSS, besides its intrinsic value.

1.6.2 Social Networks

Social network analysis is another major area of CSS, given the prominence of networks of many types in the study of social complexity. This area has become very popular in recent years, especially through the development of social media and Internet websites such as Facebook, Twitter, and numerous others. However, the analysis of networks in just about every domain across the social sciences—certainly in all the Big Five disciplines—predates computing by many years, so we should be examining the area of social network analysis from its historical roots. Social network analysis is the only area of CSS that has a well-documented history (Freeman 2004).

The advent of digital computing and CSS has transformed the study of social complexity through network analysis and modeling, expanding the frontiers of research at an unprecedented rate while advancing our understanding along many fronts in this area. There are numerous reasons for the exciting progress that this area is experiencing. For one, based on decades of pioneering research on networks, by the time computers became part of their methodological toolkit, social scientists had already developed a powerful set of concepts, statistical tools, and mathematical models and procedures, including formal theories, which enabled them to exploit computational approaches. Another reason for the explosion of progress on theory and research in this area of CSS is that computational tools, especially the most recent generation of computer hardware and software systems, now enable efficient processing of high-dimensionality data and large matrices necessary for understanding complex social networks.

Social network analysis has intrinsic value, and it also contributes to the other areas of CSS theory and research. We shall examine examples of these synergies, but before that it is necessary to gain familiarization with basic concepts, theories, and research methods in this area—*almost* as if it had no applications in other areas of CSS!

1.6.3 Social Complexity

In this introductory chapter we have already previewed some initial ideas for understanding social complexity, because this is such a defining, foundational theme for CSS. However, there is much more to understanding social complexity and its many exciting scientific and policy implications, besides the preliminary introduction that has been provided thus far. For example, research in the area also requires an understanding of **origins of social complexity** in regions where the earliest civilizations emerged, and their subsequent, long-range historical development. The study of origins of social complexity should be seen in much the same way as a science course in astronomy examines the *cosmology* of the physical universe, in terms of how the physical universe originated and how and why the earliest structures and systems emerged—the formation of stars, planets, moons, planetary systems, galaxies, and clusters of galaxies that span the cosmos. Traditionally—and perhaps not so surprisingly, given the standard (read: "turf-based") territorial disciplinary divisions of academic labor—most, albeit not all, of the study on origins of social complexity has been conducted by a relatively small community of archaeologists, mostly working in isolation from other social scientists. However, this is changing and CSS is playing an increasingly significant role in our scientific understanding of the origins of social complexity and civilizations.

In addition to understanding the origins of social complexity—just as astronomers are familiar with cosmology *and* contemporary theories and research for understanding the current universe—in this area of CSS it is also essential to develop a better understanding of interdisciplinary **concepts and theories of social complexity**. For example, whereas concepts such as information-processing, adaptation, and socio-technical artifacts provide some explanation of the phenomenon, CSS theory draws upon a broad array of other social science concepts, such as decision-making, coalition theories, collective action, and others. The **Canonical Theory** of social complexity provides a formal and empirically valid framework for describing, explaining, and understanding social complexity origins and development. Moreover, CSS investigation of social complexity also includes key concepts from **complexity science**, including the theory of **non-equilibrium distributions**, **power laws**, **information science**, and related ideas in contemporary science. This is another highly interdisciplinary area of CSS, bringing together quantitative and computational social scientists, as well as ideas and methods from other disciplines across the physical, geospatial, and environmental sciences.

1.6.4 Social Simulation Modeling

The CSS area of **social simulation modeling** can be characterized as foundational, multi- as well as inter-disciplinary, and diverse, meaning it is based on many different methodologies in modeling and simulation disciplines. The area is increasingly significant and mature for conducting both basic science and applied policy analysis. Like social network analysis, this area is sometimes confused with the totality of CSS, whereas it is only an area, not the whole field of CSS.

The simulation modeling tradition began in social science many decades ago, during the earliest days of digital computing. There are several different kinds of social simulation modeling frameworks, as we shall discuss. Regardless of the specific type, all social simulation models share a set of common characteristics. Every simulation model is always designed and built around a set of *research questions*, which may concern basic science or applied policy analysis, sometimes both. Research questions provide essential guidance for simulation models, just as in other models (for example, in formal mathematical models). Another characteristic shared by social simulation models is that they are developed through a set of *developmental stages*, not as a single methodological activity, especially in the case of complex modeling projects or those involving teams of investigators. Such stages include model verification and validation, among others. In addition, specific types of models often require additional stages in their development. It should be pointed out that each of the social simulation modeling traditions is sufficiently large to include specialized journals, conferences, and other institutional components in communities of practitioners that often number in the thousands of researchers.

The earliest kind of simulation models in CSS are the **system dynamics models**, which gained highly significant international notoriety through the global models of the Club of Rome in the 1960s and 1970s.[7] These social simulations built on the pioneering work of Jay Forrester and his group at MIT. From a computational perspective, these are equation-based models that employ systems of difference equations or systems of differential equations, as the situation and data might require. This class of models has been very significant for many decades—indeed, for half a century—because so many social systems and processes are properly amenable to representation in terms of stocks and flows, or levels and rates, respectively. Arms races, stockpile inventories in business enterprises, the dynamics of economic development, and numerous other domains of pure and applied analysis have been modeled through system dynamics simulations. A significant feature of theory and

[7] The Club of Rome is an international non-governmental organization founded in 1968 and dedicated to scientific analysis of the future and sustainable development.

research in system dynamics simulation models has been the availability of excellent software support systems, such as Forrester's DYNAMO, followed by the Stella system, and presently Vensim.

Another major tradition in social simulation models is represented by **queuing models**. As their name indicates, these models are used for social systems and processes where lines or queues of entities (such as customers, patients, guests, or other actors) are "serviced" by various kinds of stations or processing units. Banks, markets, transportation stations of all kinds, and similar systems that provide a variety of services are some examples. From a formal and computational point of view, these models are based on queuing theory, and various kinds of probability distributions are used to represent the arrival of entities at service stations, how long the service might take, and other statistical and probabilistic features of these processes. Hence, queuing models also belong to the class of **equation-based models**.

By contrast, the following kinds of social simulation models move towards the **object-based orientation** of modeling and simulation, rather than the equation-based paradigm. Of course, this is not to say that object-based models are devoid of equations; it simply means that the building blocks of this other class of models are object-like, as classes or entities. Their variables and equations are said to be "encapsulated" within the objects themselves.

The simplest kinds of object-based social simulation models are **cellular automata**, which generally consist of a grid or landscape of sites adjacent to one another, as in a checkerboard. The actual shape of the sites or cells can take on many different forms, square, hexagonal, or triangular cells being the most commonly used. The earliest work in cellular automata was pioneered by John von Neumann, who also invented game theory. The basic idea of social simulations based on cellular automata is to study emergent patterns based on purely local interactions that take place between neighboring cells on a given landscape. One of the most important and well-known applications of this kind of model has been the study of racial segregation in cities and neighborhoods, showing how segregation can emerge even among relatively unprejudiced neighbors.

Another major class of social simulation models is represented by **agent-based models**, often abbreviated as ABMs.[8] In this case the actors being simulated enjoy considerable autonomy, specifically decision-making autonomy, often including physical movement from one place to another, which is why they have had so much success in modeling social systems and processes having a geospatial dimension. Agent-based models can be **spatial** or **organizational**, or both combined, depending on what is being represented in the model. Spatial agent-based models can also use a variety of data for representing landscapes, such as GIS (Geographic Information Systems) or remote sensing data. Organization agent-based models are akin to dynamic social networks, where nodes represent agents and links represent various kinds of social relations that interact and evolve over time. These kinds of social simulation models have become increasingly significant for solving theoretical and research problems that require representation of heterogeneous actors and a

[8] The computer science terminology for these models is multi-agent systems, or MAS.

spectrum of interaction dynamics that are simply intractable through mathematical approaches that require closed-form solutions. They are also particularly appealing for investigation of emergent patterns indicative of complex adaptive systems. For example, a significant application of agent-based models is the study of complex crises and emergencies, given their ability to represent human communities in environments prone to natural, technological, or anthropogenic hazards. In another important application, as we shall see, agent-based models provide the first viable methodology for modeling entire societies, polities, and economies, as well as national, regional, and global scales of these social systems.

Finally, **evolutionary computation models** represent the class of social simulations based on notions and principles from Darwinian evolution, such as evolutionary algorithms. Although evolutionary computation models are still relatively new in CSS, they already have shown great promise. For example, they allow us to derive patterns of social dynamics that are not well understood, so long as the simulation model can be made to match empirical data. This use of evolutionary models in a "discovery mode" is characteristic of this particular kind of simulation.

Each of the preceding types of social simulation models can, at least in principle, include ideas and components from other areas of CSS, such as results from automated information extraction, social network analysis, complexity-theoretic ideas, and the like. Conversely, social simulation models can provide significant input and improvements pertinent to research in these other areas.

This brief survey of simulation models in CSS covers most of the areas that have been developed during recent decades. No doubt other social simulation methodologies will emerge in the future, either as outgrowths of current modeling approaches (as agent-based models originated from cellular automata models) or as novel inventions to analyze problems or investigate research questions that remain intractable by the current types of simulation models.

1.7 A Brief History of CSS

Each of the areas of CSS that we have introduced in this chapter has its own, more detailed, history, the main highlights of which are provided in each of the chapters to follow. The purpose in this section is to provide an overall, albeit brief, history of the entire field of CSS, beginning with its historical roots.

How, when, why, and who began the field of CSS as a systematic area of inquiry is similar in some respects to the history of other scientific fields. The historical origins of CSS are to be found in the Scientific Revolution that occurred in Europe during the late Renaissance and early Enlightenment periods. This was the epoch when the social sciences began to adapt universally held concepts and principles of positive scientific methodology (not just particular quantitative methods, such as statistics), specifically with regard to measurement of observations, systematic testing of hypotheses, and development of formal mathematical theories for explaining and understanding social phenomena. Human decision-making and voting behavior (i.e., the foundations of social choice theory) were among the earliest areas of inquiry. Statistics, initially intended to be the scientific discipline to study the state and improve policy analysis, was also invented during this period. Statistical and mathematical methods were introduced throughout the 18th and the 19th century by famous luminaries such as Denise Poisson, Adolphe Quételet, William Petty, Daniel Bernoulli, Pierre de Fermat, Jean Marie de Condorcet, Corrado Gini, and Vilfredo Pareto, among many others. The most important result of this formative period in the history of the social sciences was the adoption of a scientific culture concerning the quest for knowledge and understanding, a tradition that endures to this day.

For our purposes it is useful to mark the beginnings of CSS, in a strict sense, with the invention of digital computing during the closing days of World War II and the early days of the Cold War. This major milestone in the world history of science and technology affected the social sciences in two transformative ways, each of which is interesting in its own right. First, the modern digital computer enabled the emergence of CSS by providing the key instrument that would fuel and expand its research horizons in a way that would have seemed unimaginable just a few years earlier. For the first time social scientists were able to analyze vast amounts of data, test many novel scientific hypotheses, and explore the dimensions and structures of social space—from the human mind to the global system, with numerous levels of analysis in between. An early example of this was the invention of factor analysis—a powerful inductive, dimensionality-reduction methodology that led to many discoveries across the social sciences—by early CSS pioneers such as Charles Spearman, Rudolf Rummel, and L. Thurnstone. Among these was the discovery of the dimensionality of human cognitive spaces, as well as the structure of spaces wherein international interactions occur. Yet another example was the invention of the General Inquirer, a computational content analysis system that allowed social researchers for the first time to explore and test hypotheses concerning the content of an unprecedented volume of qualitative text data. Within the span of a single generation the volume of knowledge across the social sciences increased by many orders of magnitude thanks to the advent of the modern digital computer.

The second truly major, transformative way in which the modern digital computer affected the social sciences was as an inspiring metaphor that shed new light on classical and modern areas of investigation. Social scientists had known for some time the significance of communication and information-processing for understanding human and social dynamics. For example, the study of media and text data, as well as radio broadcasts and propaganda, had begun in earnest many decades before the advent of the computer. However, the digital computer inspired new concepts, hypotheses, principles, models, and theories about the vast array of systems and processes in the social universe. For instance, political scientists who became familiar with ideas from cybernetics and general systems theory (new fields pioneered by scientists such as W. Ross Ashby, Norbert Wiener, Ludwig von Bertalanfy, and Anatol Rapoport, among others) began viewing the structure and functioning of polities and other forms of political systems by highlighting the role of information-processing, goal-seeking behavior, social computing, and emergent phenomena. An example of this was the novel cybernetic theory of government formulated by Karl W. Deutsch and others, who played a leading role during the Behavioral Revolution of the 1960s. A polity, as we will see in subsequent chapters, can be described and understood as a complex adaptive system that carries out numerous, coordinated computations, such as voting and policymaking. Herbert A. Simon's theory of social complexity through adaptation and artifacts—published for the first time in the 1969 edition of *The Sciences of the Artificial*—was another result of the influence of digital computing machines. Harold Guetzkow developed innovative computer simulation approaches, as well as hybrid simulations (so-called man-machine simulations) that are still highly influential to this day. 1969 was also a seminal year in which Hayward Alker and Ron Brunner published the first paper on comparative simulation research.

All areas of CSS have experienced remarkable growth since the early days of the field. Progress in social theory and research, as well as remarkable advances in all areas of computing, particularly applied computational approaches and methodologies, have contributed to the current body of knowledge in CSS. Today CSS is also beginning to reap the benefits of interactions and synergies among its main areas, as they fertilize and stimulate each other each other in new and exciting ways. For example, the early history of social network analysis, or even automated information extraction, developed in relative isolation or autonomy—by endogenous development. Today, by contrast, these areas experience frequent overlays and mutually beneficial collaborations, as witnessed by the application of text-mining algorithms to populate social network models. Another example is the application of network models to improve the specification of social structures represented in agent-based models for the study of emergence in complex social systems. The history of CSS as an emergent field is still in its infancy. However, the field has already demonstrated significant capacity and promise for contributing to new understanding across all areas of social science theory and research.

1.8 Main Learning Objectives

This textbook has a set of main learning objectives intended to be pedagogically appropriate as an introduction to the field of CSS. As indicated in the preface, these objectives include learning basic concepts, models, theories, and methodologies used in CSS. These objectives are designed to serve two purposes: a basic exposure to the field of CSS, as well as building foundations for further study at more advanced levels.

The following scientific learning objectives are among the most important. Examples are provided as illustrations.

- Basic understanding of key CSS *concepts*, including all those highlighted in boldface and included in the Index, to a level where the reader can provide additional examples. Conceptual proficiency is fundamental, including concept formation in CSS.
- Familiarization with the *scope and content of each area* of CSS, grounded in elements of computing, including areas of automated information extraction, social networks, complexity-theoretic understanding of social systems and processes,

and various kinds of social simulations. Examples include complex adaptive systems, coupled systems, multi-scale processes, bifurcation, criticality, metastability, phase transitions, autonomous agents, verification, and validation.

- Understanding of *main theories* that are part of the CSS paradigm as causal explanatory frameworks that shed new light on the nature of human and social dynamics. Examples include Simon's Theory of Artifacts, the Canonical Theory of Social Complexity, the Theory of Social Networks, the Theory of Non-equilibrium Social Processes, and others.
- Ability to distinguish and analyze the different *levels of analysis* of social complexity using computational approaches, ranging from mental phenomena to decision-making, social groups and their interactions, to the global system.
- Ability to work with one or more of the *methodological tools* covered in one or more of the chapters. Examples include extracting entities from text data, computing social network indices, testing a power law hypothesis, and building a basic agent-based model in a programming language such as Python or a simulation toolkit such as Netlogo.
- Familiarization with the *main classes of entities, objects, and relations* that are most common in computational analyses of social complexity. Examples include various types of actors, associations, attributes, and methods.
- Proficiency in the *interdisciplinary integration of knowledge* in the context of social phenomena, including the synergistic nexus between social science and computational methodologies.
- Basic knowledge of the *history of each area* of CSS, including prominent pioneers, with an understanding of roots in early development of the social sciences and computer science, at least to the level detailed in the brief histories provided in each chapter.

This minimal set of learning objectives applies throughout chapters in this textbook, ideally independent of the content of each area of CSS. In addition, each chapter contains its own set of main learning objectives that are more specific to the scope and content of each area.

Motivated readers will benefit from further study of the supplementary reading materials provided at the end of each chapter under the heading of Recommended Readings. These are intended to provide more advanced foundations and knowledge that extends beyond the scope of this introductory textbook. The bibliography contains additional sources that interested readers will wish to look up, both early classic literature in CSS, as well as some of the most current and influential contributions.

Recommended Readings

H.R. Alker Jr., R.D. Brunner, Simulating international conflict: a comparison of three approaches. International Studies Quarterly **13**(1), 70–110 (1969)

H.R. Bernard, The science in social science. Proceedings of the National Academy of Science **109**(51), 20796–20799 (2012)

C. Cioffi-Revilla, Computational social science. *Wiley Interdisciplinary Reviews (WIREs): Computational Statistics*, paper no. 2. Available online (2010)

D. Collier, J. Gerring, *Concepts and Method in Social Science: The Tradition of Giovanni Sartori* (Routledge, New York, 2009)

R. Conte, G.N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffaunt, J. Kertesz, D. Helbig, Manifesto of computational social science. European Physical Journal Special Topics **214**, 325–346 (2012)

F. Fernandez-Armesto, *Civilizations: Culture, Ambition, and the Transformation of Nature* (Simon & Schuster, New York, 2001)

A.M. Greenberg, W.G. Kennedy, N.D. Bos (eds.), *Social Computing, Behavioral-Cultural Modeling and Prediction* (Springer, Berlin, 2012)

J.H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975)

I.L. Horowitz, Big Five and Little Five: measuring revolutions in social science. Society **43**(3), 9–12 (2006)

M. Kline, *Mathematics and the Search for Knowledge* (Oxford University Press, Oxford, 1985)

J.H. Miller, E. Page Scott, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life* (Princeton University Press, Princeton, 2007)

H.A. Simon, *The Sciences of the Artificial*, 3rd edn. (MIT Press, Cambridge, 1996)

L. Spinney, History as science. Nature **488**, 24–26 (2012)

M. Steuer, *The Scientific Study of Society* (Kluwer Academic, Dordrecht, 2003)

C. Williford, C. Henry, A. Friedlander (eds.), *One Culture: Computationally Intensive Research in the Humanities and Social Sciences—A Report on the Experiences of First Respondents to the Digging into Data Challenge* (Council on Library and Information Resources, Washington, 2012)

2 Computation and Social Science

2.1 Introduction and Motivation

Computation is a formal discipline used by scientists—in the social, physical, and biological disciplines—to uncover new insights and advance the frontiers of knowledge. It also informs the **Computational Paradigm of Social Science** introduced in Chap. 1. Social processes are *algorithmic*, and social systems are supported by *algorithms*, in the sense defined in this chapter. What are the elements of computation with the greatest significance for CSS? How is computation used to better understand social systems and processes? What are the core concepts and principles of social computation? Problem-solving, design, and programming are core elements of computation and the computational approach to social science. Similar activities are also foundational to understanding social systems.

The role of computation in CSS is comparable to that of mathematics in physics: it is used as a language to formalize theory and empirical research to express, study, and develop our understanding of social complexity in ways that are not accessible through other means. By contrast, pure computer scientists use computation to study computing, just as pure mathematicians use mathematics to study mathematics. This instrumental or utilitarian motivation does not prevent computational social scientists from developing deep interest in computation; there is much a computational social scientist can learn from the pattern of thinking of a computer scientist, a musician, a mathematician, or an historian. However, CSS is more like applied computer science or applied mathematics:[1] the formal approach (mathematical languages or programming languages) is used to gain substantive, domain-based knowledge about social complexity in all its rich forms.

This chapter uses the Python programming language for illustrative purposes, though not for providing tutorials. The notational graphic system known as the Unified Modeling Language (UML) is used for representing and better understanding

[1] For example, applied computer scientists work on areas such as robotics, data analysis, and optimization, to name some of the major areas of research in computer science.

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_2,

social systems and processes—including those with significant theoretical or real-world complexity. Importantly, UML is also used in subsequent chapters to describe social systems and processes, such as decision-making by actors, polities and their institutions, socio-environmental dynamics, and other entities of social science research interest.[2]

2.2 History and First Pioneers

Computation has a long, interesting history in social science. Computational Social Science (CSS) began with the first applications of computation during the early 1960s, with pioneers such as Harold Guetzkow (1963), Herbert A. Simon (1969), Karl W. Deutsch (1963), John C. Loehlin (1968), and Samuel J. Messick (1963), roughly a decade after von Neumann's (1951) pioneering Theory of Automata. That was during the age of punched tape, 80-column IBM cards, and long hours spent at the university's computer center awaiting output results, often in vain due to some syntactical glitch in the program, which often caused another day's worth of work. In spite of such early difficulties, the advent of computation in social science came at an auspicious time, because theoretical and methodological advances were taking place along numerous frontiers across disciplines. Field Theory (Lewin 1952), Functionalist Theory (Radcliffe-Brown 1952), Conflict Theory (Richardson 1952a, 1952b), the Theory of Groups (Simon 1952), Political Systems Theory (Easton 1953), as well as Decision-making Theory (Allais 1953), among others, required new formalisms that could treat conceptual and theoretical complexity of human and social dynamics, beyond what could be accomplished through systems of mathematical equations solved in closed form.

Each of the social sciences (Anthropology, Economics, Political Science, Social Psychology, and Sociology) and related fields (Geography, History, Communication, Linguistics, Management Science) witnessed the introduction of computation into its own frontiers of theory and research within a few years. However, formal training in computation did not begin until decades later through high-level software packages for statistical applications (SPSS, SAS, Stata), followed by true programming languages (S and R), as well as computational applications to content analysis, network models, and social simulations. Many of these computational contributions will be examined in subsequent chapters of this book.

Those were the origins of CSS, a fledging field that has evolved from pioneering roots that began with primitive algorithms running on archaic computers with (mostly) historical interest, to today's object-oriented models running on modern and more powerful computers that would have seemed like science fiction even to Isaac Asimov's psychohistorian Hari ("The Raven") Seldon in *Foundations*. What about the future? The future of CSS will be written in the language of advanced distributed computing, graphic processing units (GPU), quantum computing, and other information technologies still at the frontiers of computational science.

[2]The material in this chapter assumes a level of computer science knowledge comparable to Eric Grimson and John Guttag's famous MIT course (Grimson and Guttag 2008) or Guttag (2013).

2.3 Computers and Programs

2.3.1 Structure and Functioning of a Computer

All computers are information-processing systems: they compute, as the term indicates, based on a set of instructions called a **program**. Programs are written as a series of instructions, not unlike a recipe, in computer code. The code must be written so that it conforms to the format of the programing language, or syntax.

All computation can be seen as a **problem-solving system** consisting of subsystems of **hardware** and **software** components. While hardware provides the physical means for information processing (i.e., computing machines, or computers, in a narrow sense), software provides the algorithmic instructions that tell the hardware what to do (i.e., what to do with the information being processed) in some programming language. In computer science, software is also known as **code**, not to be confused with the same term as used in social science measurement and empirical research to represent the value of some variable (usually a nominal variable). Computationally speaking, code is distinct from **data**, which are processed by code.[3]

These initial ideas have resonance in social science, where information-processing systems are ubiquitous, significant, and highly consequential: individuals, groups, and institutions ranging from local neighborhoods to the global system of international organizations process information following procedures, engage in problem-solving, and use institutions (akin to hardware?) as well as established and adaptive systematic processes (software?) to address and solve problems pertaining to the full spectrum of societal issues. The mapping between computers and social systems is not exact, nor is it necessary for computation to be useful in social science, but it can be insightful in pointing out significant features of social complexity that extant social theories have neglected or simply been unable to explain. Metaphors are often useful in science, but for computation to be a powerful paradigm and methodology in CSS it is necessary to look deeper into its concepts and principles.

As illustrated in Fig. 2.1, in its most fundamental structure, a computer is a machine (hardware system) composed of five types of components, each designed to perform a specific function. There are two core components: a **central processing unit** (CPU) and **main memory**. The CPU carries out the most basic computations, such as arithmetic operations, comparisons, or Boolean true/false operations. Data and programs are stored in main memory (or RAM, random access memory), which has a tightly coupled interactive relationship with the CPU for performing computations (i.e., executing instructions).

Secondary memory (Fig. 2.1, lower center), in larger and typically slower form than main memory (e.g., a disk), is used to store information more permanently (as programs and data files) when a computer is turned off. When a computer is turned

[3]In social science, data and information denote different concepts. Data (the lower-level concept) normally refers to raw observations, such as field data or census data, whereas information (higher-level) is based on or is derived from data and provides a basis for knowledge. Data is the plural of *datum* (or fact, in Latin), so the correct phrases are "one *datum*" and "several *data*."

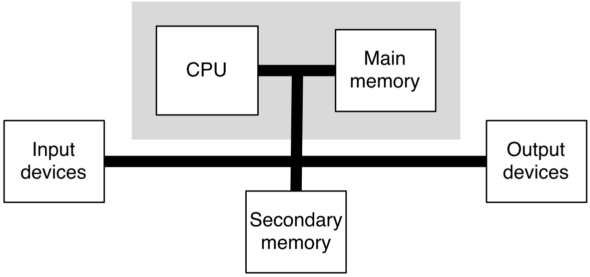


Fig. 2.1 A computer with its functional components (the five boxes) based on a bus architecture (fast-speed data connections)

on, secondary memory is accessed to retrieve data and programs that are executed by the CPU using main memory.

Input and **output devices** are for us humans to interact with the three components just described, as human-machine interfaces (Fig. 2.1, left and right). Input devices include keyboard, mouse, microphones, cameras, joysticks, and many kinds of sensors ranging from relatively simple (e.g., a thermostat fixed to a wall) to highly complex (biohazard sensors mounted on an unmanned autonomous vehicle or UAV). Output devices include printers, speakers, electromechanical devices (e.g., robots), and other devices. The earliest monitors were output devices, whereas today some monitors serve a dual function as input devices as well (e.g., a touch-sensitive video screen).

Fast data connections (called "internal buses") link core components (CPU and main memory) between themselves and, via other connections (also called "expansion buses"), with external components (I/O devices). The overall architecture of internal components with relations among them, versus external devices in the environment of a computer, bears resemblance to Herbert A. Simon's model of a complex adaptive system consisting of an inner system and an external environment—a paradigmatic model and theory that we will examine in much closer detail later, given its significance for explaining and understanding social complexity. This important approach is still mostly unknown in the social sciences, 40 years after Simon's pioneering work, and Simon is remembered mostly for his work on bureaucracy and incrementalism.

When a computer is turned on and a program (of any kind) is asked to run, the operating system handles what is called the *load-fetch-decode-execute cycle*, or **fetch-execute cycle** for short—and it is something a CSS researcher needs to know. Understanding this is helpful for deciding, for instance, whether a model can be implemented to run on a single processor, or whether some form of parallel processing is necessary. First, program instructions are loaded ("fetched") from secondary memory, where they reside (almost) permanently, onto the main memory (RAM). Second, the CPU accesses the first instruction from RAM, decodes that instruction, and executes it. When finished executing the first instruction, the same fetch-execute cycle is repeated as many times as there are instructions in the program. A well-written program will organize this cycling process in such a way as to take advantage of the fast cycling time of the CPU, subject to available RAM capacity. Knowledge of this cycling process is not necessary for most programing tasks, but becomes increasingly important with parallelization, especially GPU programming. Most multithreading is handled at a higher level of abstraction.

A significant feature of the fetch-execute cycle is that it consists of discrete events that are (i) critically necessary (i.e., conjunctive), (ii) sequential in a strict order (sequential conjunction), and (iii) each event takes time that cumulatively determines total cycle duration. While this is common knowledge for computer scientists, few social scientists have paid close attention at the deep properties and principles of such systems and processes of sequential conjunction for explaining and understanding social complexity. On time scales that are many orders of magnitude slower than computers, human cognition, decision-making by individuals and groups, policy-making, and numerous other social processes examined in this book—especially in Chaps. 6 and 7—share significant isomorphic features with fundamental patterns in computation, such as the sequential conjunction of the fetch-execute cycle and others.

DID YOU KNOW THAT ...? Comparing the time-scales of computers with that of individual humans and human institutions adds perspective to information-processing under different architectures of complexity. A MacBook Pro laptop computer has a 2.66 GHz Intel Core i7 CPU and four GB 1067 MHz DDR3 RAM chips. CPU speed is measured in cycles per second (or hertz), so this means that the CPU of the MacBook Pro laptop can execute $2{,}660{,}000{,}000 = 2.66 \times 10^9$ instructions per second. High speeds such as these allow a modern computer to execute many instructions in background mode while a relatively idle program, such as a word processor, is in use.

Suppose we compare an instruction execution by a CPU to a policy decision by a national legislative body. No one has yet estimated the number of decisions made *each year* by such institutions, but it is clearly many orders of magnitude slower. By contrast, human individual decision-making takes place on a scale of tens of milliseconds.

2.3.2 Compilers and Interpreters

A CPU understands only its own **machine language**, whereas most computer programs are written in a high-level language. In order for a computer to run a program written in a given high-level language (i.e., a program in other than low-level machine language), the program must first be either **compiled** or **interpreted**. The difference between these two processes is fundamental, subtle, consequential, and important for CSS researchers to understand. A **compiler** is a program that literally translates source code written in a high-level programming language (e.g., Fortran, C++, Pascal, Python) into machine code that is specific to and executed by the computer's CPU. Once compiled, a program can then be run many times without having to recompile the source code. Compiled code is machine-specific binary code, ready for execution by the CPU; it provides a complete translation of all instructions, line by line.

By contrast, other languages that are not compiled use an **interpreter** that is specific to the high-level language for communicating the program's instructions to a computer. An interpreter is a specialized, low-level program that enables hardware to execute the high-level software. A language requiring an interpreter must use its associated interpreter every time the program is executed; otherwise the CPU will not understand the program's instructions.

In sum, a compiler translates a program into machine code, line by line, to execute; an interpreter reads all the source code and directly communicates its instructions in machine code to the CPU without compiling a new program (as the compiler does). The main difference is similar to knowing a foreign language (compiling) versus translating one line at a time (interpreting). Comparing the two types of high-level languages, compiled programs run relatively faster but have drawbacks, whereas interpreted programs run somewhat slower but they can run interactively. The difference is important for a CSS researcher to understand, because it can mean choosing one programming language over another, depending on what model or algorithm is being implemented.[4]

2.4 Computer Languages

Social science uses mathematics as a language to formalize theory and investigate features of social complexity that are exclusively accessible through the medium of mathematical structures, such as sets, probability, game theory, or dynamical systems. The same is true when using computer languages in CSS. A **computer language** is a structured, formal grammar for communicating with and controlling what a computer does. Like all languages, including mathematical structures used by social scientists, computer languages consist of **syntax**, **semantics**, and **pragmatics**.[5] Syntax refers to the proper rules for writing instructions, the correct sentences of a properly written program. Semantics refers to the meaning of symbols; i.e., what various code elements stand for. Pragmatics refers to the primary purpose, function, or paradigmatic orientation of a language. Computer languages differ by intent, just like different symbolic systems or mathematical structures are created for various purposes (e.g., music notation or game theory).

Social science has used a significant array of mathematical structures over the past two hundred years, but formal instruction in mathematics has lagged behind statistics. Now, in addition to statistics and mathematics, social scientists require

[4]The case of Java is somewhat hybrid: Java is technically compiled into Java byte code, and then just-in-time compiled into machine code by the Java Virtual Machine (JVM)—which can be viewed as a byte code interpreter.

[5]Linguists would also add **genetics**, the origin of a specific language. For example, the Python programming language was created by Guido van Rossum in the late 1980s and has since evolved into version 3 (as of this writing), supported by a global community.

Table 2.1 Comparison of computer programming languages. Paradigm types are explained in the text. *Source:* Wikipedia, "Comparison of programming languages: General comparison"

Assembly language	Imperative
BASIC	Imperative, procedural
C	Imperative, procedural
C++	Imperative, object-oriented, procedural
Fortran	Imperative, object-oriented, procedural
Java	Imperative, object-oriented, reflective
Lisp	Imperative, functional
Mathematica	Imperative, functional, procedural
MATLAB	Imperative, object-oriented, procedural
Pascal	Imperative, procedural
Python	Aspect-oriented, functional, imperative, object-oriented, reflective
S and R	Functional, imperative, object-oriented, procedural

training in programming languages, which is essential for CSS theory and research on social complexity.

Every computer language has features that make it more or less effective in implementing human and social dynamics, just as is true for different modeling languages used in mathematical social science. Specifically, each programing language has its own syntax, semantics, and pragmatics, which results in features such as those listed in Table 2.1.

Python is a programming language with several desirable features for learning Computational Social Science: it is easy to learn *and* can be used to learn some of the best **computer programming habits**, such as consistent style, modularity, defensive coding, unit testing, and commenting.[6] A drawback of Python is that it can slow down considerably with increasing program complexity, such as when used for social simulations such as those examined later in this book. A recommended strategy is to learn how to program using Python, then learn a more advanced language, such as Java or C++.

Several other specifically technical features of Python include:

- **Object-orientation:** Python is a language that supports the object-orientation to programming (OOP), meaning that the basic building blocks of a Python program can represent social entities (e.g., actors, relations, groups), similar to the building blocks of many social theories. In turn, social entities (objects and associations) contain within them ("encapsulate") variables and dynamics that determine the state of the overall social entity or phenomenon being modeled. By contrast, earlier programming languages required direct modeling of variables

[6] By contrast, *bad* programming habits include lack of modularity, hazardous loops that can easily spin forever, "stringy" code, and comments that are unclear, unhelpful, quirky, or plain absent. Good coders avoid these and other bad habits and strive to develop an excellent, "tight" style, as discussed later in this chapter.

and equations, which is sometimes too difficult, cumbersome, or impractical for many social theories.[7]

- **Interpreted code:** Python code is interpreted, not compiled, so it can be run interactively. This is helpful for several purposes: developing a program as it grows from simple to more complicated; verification and debugging; running simulation experiments. Python code runs from a command line terminal or from a shell editor, as well as interactively or as an executable file.
- **Imperative style:** As an imperative language, a program written in Python can contain statements that change the state of the program. This means that a Python program can implement a series of commands or instructions that the computer can execute to change the state of social objects, constructs, or entities represented in the program.[8] Assignment statements, looping statements, and conditional branching are important features of imperative programing.
- **Function libraries:** As with other popular programing languages, Python supports the use of functions that are evaluated in terms of specific arguments. Given a function `f(x)` with argument `x`, the evaluation of `f` always returns the same result as long as `x` does not change. Functions are used to implement many kinds of social processes, such as utility functions in decision-making, interaction dynamics, and other behavioral features. Functions need not always be mathematical equations. For example, they can be table functions.

Python can be used for many scientific purposes in CSS, running in both *interactive* and *batch* modes. As a *calculator*, Python can be used to compute results, such as a probability value or an expected value, just like a hand calculator. More complicated functions are best analyzed in batch mode.

Example 2.1 (Interaction Between Human Communities) In human and social geography, the potential for many modes of human interactions between two communities (marriages, migrations, and phone calls, among others) is approximated by the so-called *gravity model:*

$$I \approx \frac{P_1 P_2}{D^\alpha}, \tag{2.1}$$

where I is the interaction potential, and P_1 and P_2 are the populations of the two communities separated by distance D. The exponent α denotes the difficulty involved in realizing interactions (costs, terrain, transportation opportunities, and similar), such that I decays rapidly with increasing α. Suppose two

[7] This is a significant advantage of OOP that will arise again in various chapters. The main idea of the object-orientation to programming is that basic social *entities* and *relations* are identified first; all the rest (variables, data, parameters, equations) comes later.

[8] By contrast, a so-called **declarative** style of programming emphasizes the desired result of a program, not the instructions necessary to produce results.

communities with 20,000 and 30,000 inhabitants are 120 miles away from each other. To appreciate the effect of difficulty α on the interaction potential I, we can compute the potential with $\alpha = 2$ (standard assumption) and 3 (greater difficulty), respectively:

```
>>> print((30000)*(20000)/(120**2))
41666.666666666664
>>> print((30000)*(20000)/(120**3))
347.2222222222223
```

We see immediately how a single unit difference in difficulty α (2 vs. 3) causes a drop in interaction potential of two orders of magnitude (10^4 vs. 10^2).

Example 2.2 (Terrorist Attacks) Terrorists face a daunting challenge when planning an attack, mainly because the probability of success in carrying out an attack (technically called a *compound event*, as we will examine later in greater detail) is contingent on many things going well: planning, recruitment of confederates (e.g., scouts, suppliers, operatives, etc.), training in weapons and tactics, proper target selection, execution, and overcoming target passive and active defenses, among other requisites. Assuming $N = 10$ critical requirements for a successful attack, each being solved with probability $q = 0.99$, we get:

```
>>> print(.99**10)
0.9043820750088044
```

Under somewhat more realistic (but still generous) assumptions, with a lower 0.90 probability of requirement-level success:

```
>>> print(.90**10)
0.3486784401000001
```

In fact, as demonstrated in subsequent chapters, the partial derivative $\partial(q^N)/\partial q$ is highly sensitive to the probability of individual task success q (more than to N). This explains why counterterrorism strategies aimed at hindering individual tasks are quite effective, without having to target every single stage of a potential attack process.

Python can be used as a simple calculator for exploring, analyzing, and learning more about social models as in these and other examples. Note that typing `print()` is not necessary, strictly speaking, but it is a good habit because when running a batch script it is always necessary to use print to output results.

Alternatively, and more interestingly from a research perspective, Python can be used for running *programs* for investigating a large spectrum of social models—from individual decision-making to global dynamics in the international system—that have been analyzed only through closed form solutions, without using the power of simulation and other CSS approaches. However, due to speed limitations mentioned earlier, running social simulations or network models in Python can be problematic in terms of speed, so a better methodological approach in some cases is to implement models in other, faster languages, such as Java or C++. This is why Java is a common language for multi-agent simulation systems or "toolkits," such as MASON and Repast, and why C++ is often used in parallel, distributed computing. For example, Repast-HPC is based on C++.

The following are other features or types of programming languages mentioned in Table 2.1:

- **Procedural programming:** This refers to the programming paradigm based on procedure calls (in high-level languages) or subroutines (low-level). Routines and methods are procedure calls containing some sequence of computations to be executed.
- **Reflective programming:** The ability of a programming language to read and alter the entire structure of the object at compile time is called reflection.

Why should a CSS researcher know about different features (or *paradigms*, as they are called in computer science) of programming languages? The reasons are similar to why a mathematical social scientist needs to know about what each formal language is capable of modeling. For example, classic dynamical systems can model deterministic interactions, whereas a Markov chain can model probabilistic change, game-theoretic models capture strategic interdependence, and so on for other mathematical languages. Reliance on the same mathematical structure every time (e.g., game theory, as an example), for every research problem, is unfortunately a somewhat common methodological pathology that leads to theoretical decline and a sort of inbreeding visible in some areas of social science research. Dimensional empirical features of social phenomena—such as discreteness-continuity, deterministic-stochastic, finite-infinite, contiguous-isolated, local-global, long-term vs. short-term, independence-interdependence, synchronic-diachronic, among others—should determine the choice of mathematical structure(s). Similarly, different programming languages provide different features, so they should be selected in accordance with the nature of the social phenomena to be modeled. The same is true of using programming languages in CSS, for the very same reason: not all problems can (or should!) be solved with the same scientific tool.

2.5 Operators, Statements, and Control Flow

The examples in the previous section used the **interactive mode** in Python, which works well for simple calculations or short code snippets that are brief and are used just once or a small number of times. When the calculations are more complex, when instructions need to be executed several times, or when the sequence of instructions is longer than just a few lines, it makes more sense to create a separate file containing a **program** consisting of statements. Then the program can be written, edited, and saved, just like any text file. The program is then executed any number of times by running (or *calling*) it from the command line or from Python's own shell (e.g., IDLE).

The following program illustrates a number of ideas concerning operators, statements, and control flow.

Example (Chaotic Process (Zelle 2010: 13)) This example is taken from a leading textbook on Python, illustrating the nature of chaotic processes. Write the following simple program in a text file (say, `chaos.py`) and run it from the Python shell.

```
>>> def main():
 print("This program illustrates a chaotic function")
 x = eval(input("Enter a number between 0 and 1: "))
 for i in range(10):
    x = 3.9 * x * (1 - x)
    print(x)

main()
```

When `main()` runs, it should return the following result:

```
This program illustrates a chaotic function
Enter a number between 0 and 1:
```

Next, enter a number between 0 and 1, and the program should return a sequence of 10 values. Change the range from 10 to N, call `main()` again, and now N values will be returned. The coefficient can also be changed to a value different from 3.9, which will generate a different chaotic series.

The example just discussed contains a number of points worth noting from a CSS perspective. First, it takes relatively little in terms of program sophistication to opt for a program, rather than using the interactive mode. Or we may wish to run a program with variations for conducting **computational experiments**. Most social models require some statements that warrant a program, even when the number of **lines of code (LOC)** is relatively small (i.e., less than a dozen), as in the example. Copying and pasting in interactive mode helps, but calling a program (e.g., as in `>>> import filename`) is even easier, and that is what most researchers would do.[9]

[9]Computer programs are artifacts—in the sense of Simon—which sometimes, in turn, provide support to other artifacts. An example of this is a spacecraft. As of early 2012 the International Space Station orbiting Earth—one of the world's most complex adaptive artifacts—was supported by

Second, the structure of a program is always a function of "The Question" (or set of questions) being asked in a given investigation. In this case, the question concerned the behavior of a chaotic process; specifically, which series would be generated by a given initial value, assuming a specific coefficient. A different program would be necessary to address closely related but different questions, such as:

- What happens when noise is introduced?
- What if the coefficient varies as a function of some other parameter affecting the process?
- What is the correlation between series of values generated by different initial conditions (or different coefficients)?
- How can we graph the process, as in a time-series plot, rather than observe a list of numbers?

None of these questions can be addressed by the same program, especially the last, which requires calling additional facilities, such as Python's `graphics` library. Each program is designed to address a specific question.

Third, note that each statement in a program is intended to control some aspect of the information being processed. In this case the program began by defining a new **function**, called `main`. *Knowing how to define new functions is a basic programming skill* and an easy task in Python. Next, the program states that something is to be printed exactly as specified by the `print` function. This is optional, but *good practice*, since it tells the user what is going on without having to look into details. The program then contains a core statement about evaluating another function, this time an `input` function in response to a query. Next, the program uses a series of related statements to control the computation of `x` by means of a **loop**: `for i in range(10):...`. Loops are essential **control flow** statements along with others, such as `if` and `while` statements.

2.6 Coding Style

Computer programming is a form of formal writing, so style matters and developing a good style for writing programs is important for a computational social scientist—just as it is for a computer scientist. General principles of **good coding style** apply to all programming, while specific principles or guidelines apply to particular programming languages, similar to mathematics in this respect.

The need for **general principles of good coding style** is motivated by many factors that operate in any field of modern science, including Computational Social Science:

---

computer programs with approximately 2.3 million LOC, a figure always increasing with growing project complexity until the ISS mission is completed. Unfortunately, however, LOC per se are not a good proxy measure for algorithmic or software complexity: high LOC may reflect mere lack of expertise, whereas low LOC may result from overly complicated implementations, instead of simpler, maintainable versions that would require more LOC.

- Code is a formal system of writing, so its syntax and semantics are governed by both technical and esthetic principles, not just the former. The same is true of mathematics: well-written mathematical papers are also based on technical and esthetic principles.
- Code is sometimes used by programmers long after it was first written by the original programmer(s). If it was not well-written to begin with, subsequent programmers (or even the initial programmer) may have a difficult time understanding it.
- Many multi-disciplinary projects in CSS contain researchers from diverse backgrounds (social science, computer science, environmental science, or other disciplines), which increases the communications requirements.

The following are important general principles of good coding style:

1. **Readability**: Always write code in such a way that others can easily read and understand it. Code should not be written using short variable names or function names, such as is common practice in mathematics. " `numberOfRefugees-intheCamp`" is good; "`N`" or even "`NORIC`" are not. Incomprehensible code is not a sign of genius; it is a sign of disrespect toward collaborators, current or future.
2. **Commenting**: Writing informative comments is an important way to implement readability. Uncommented code needs to deciphered, or it may be useless. The main consumer of comments is often the original programmer, since even a few days later it is easy to forget what a code segment was intended to do.
3. **Modularity**: Write in modules, such that the overall program is akin to a **nearly-decomposable system,** in the sense of Simon. Object-oriented design patterns can be useful when separating components that are not so obviously decomposable. Functions and their embedding property provide a viable strategy for modularization.
4. **Defensive coding:** Writing defensive code means to try to ensure that code does not malfunction, ending up doing something different from the intended purpose. An example would be being careful in avoiding loops that can cycle infinitely. This is achieved by careful coding and by inserting proper tests that will prevent infinite loops.

These basic principles of good coding style are intended not just for beginners; they are also practiced by good modelers and software engineers.

2.7 Abstraction, Representation, and Notation

How does science (any science) make fruitful inquiry feasible and tractable, given the complexity of the real world? The viability of doing science in any field depends on making the subject matter tractable in terms of research that is systematic, reproducible, and cumulative. Social, physical, and biological scientists render their substantive fields tractable through **simplifications** that are sufficient to ensure the growth of a viable science, but not so simple as to preclude deep understanding of phenomena. **Tractability** is therefore a sophisticated strategy of scientific inquiry that seeks to simultaneously maximize **parsimony** and **realism**—as in a *Pareto frontier*. Parsimony ensures causal explanations (theories) and empirical descriptions (laws) that contain a minimal number of factors deemed essential for explanation, understanding, and sometimes prediction. Realism ensures that the science remains empirically relevant and sufficiently rich in terms of capturing real-world features. Science seeks to make real-world complexity tractable.

Social complexity in the real world of people, their thoughts, decisions, social relations, and institutions, is intricate and far more complex than the simple world of two-body mechanics and equilibrium systems.[10] It consists of individual actors with bounded rationality, interactions that are often hard to predict (even when they are just dyadic), and the emergent social results generate networks, organizations, systems, and processes that challenge all areas of social science theory and research—transcending individual disciplines. To solve this challenge, social science has learned to rely on **abstractions**, **representations**, and specialized **notations** to advance our understanding of the social universe through concepts, theories, and models.

For hundreds of years, since the rise of modern social science in the Age of Enlightenment and the Scientific Revolution, social scientists have used statistical and mathematical representations based on abstractions of real human and social dynamics. All such models—and the social theories they involve—are formal linguistic inventions based on systems of specialized notations.[11] Just as social scientists have learned to use abstractions to formulate statistical and mathematical models of the social world in many domains, today computational social scientists use computer programs and computational models to abstract, represent, analyze, and understand human and social complexity. What do abstraction, representation, and notation require in CSS? How do they work in a coordinated way to produce viable code for modeling and analyzing complex social systems and processes?

Abstraction

In computer science, abstraction means hiding information. In CSS, abstracting from the world "reality"—whether directly experienced (observing a riot downtown) or indirectly learning about it (reading history)—is a process involving stimulus signals, perceptions, interpretation, and cognition. CSS relies on several **sources for abstracting** key entities, ideas, and processes from raw stimulus signals from the real world. These sources span a hierarchy in terms of their social scientific status. At the very top of the hierarchy are **social theories** with demonstrable validity

[10] A little-known fact among many social scientists is that the theory of mechanics in physics is built around the abstraction of single- and two-body problems. Already three-body problems are hugely difficult by comparison; and, most interesting, N-body problems defy mathematical solution in closed form.

[11] Interestingly, humanistic fields such as music and ballet also use systems of specialized notation, far beyond what is used in traditional social science. In music, Guido d'Arezzo [b. A.D. 991 (or 992), d. 1050] is considered the founder of the modern music staff; in ballet, Rudolf von Laban [b. 1879, d. 1958] invented the symbolic system known as "labanotation" (Morasso and Tagliasco 1986).

in terms of formal structure (internal validity) and empirical observation (external validity). Not all existing social theories meet these stringent requirements, although an increasing number of them do as research progresses. Examples of social theories that meet internal and external validity standards include Heider's Theory of Cognitive Balance in psychology, Ricardo's Theory of Comparative Advantage in economics, and Downs's Median Voter Theory in political science, among others. Social theories are abstractions that point to relevant social entities, variables, and dynamics that matter in understanding and explaining social phenomena.

A second source of abstraction consists of **social laws**. Examples of social laws include the Weber-Fechner Law in psychometrics, the Pareto Law in economics, and Duverger's Law in political science. Theories explain; laws describe (Stephen Toulmin 1967).[12] Some of the most scientifically usefully social laws can be stated mathematically, as in these examples. Social laws also contain relevant entities, variables, and functional relations for describing social phenomena.

A third source of abstraction consists of **observations** that can range from formal (e.g., ethnography, content analysis, automated information extraction, text mining, among others) to informal (historical narratives, media, and other sources about social phenomena). Observations of social phenomena can describe actors, their beliefs, social relations, and other features ranging from individual to collective.

Finally, a fourth source of abstraction consists of **computational algorithms** capable of emulating social phenomena, as in artificial intelligence (AI). Artificial (i.e., not really human) algorithms do not claim to be causal in the same sense as social theories. They "work," but without causal claims in the same sense as social theories. They are efficient, in the sense that they (sometimes) can closely replicate social phenomena. AI algorithms are typically (and intentionally) efficient and preferably simple; extreme parsimony in this case comes at the expense of realism. Examples of AI algorithms include Heatbugs (Swarm, NetLogo, MASON), Boids (Reynolds 1987), and Conway's (1970) Game of Life. In spite of their lack of social realism, AI algorithms can be useful sources for abstracting social entities, ideas, or processes because they can highlight features that either elude theories or are hard to observe. An example would be the agglomeration patterns generated in a Heatbugs model, as a function of varying parameters of "social" interaction among the set of agents, or the role of apparent "leadership" in a flock of boids.

Representation

Abstraction is a necessary early step in scientific inquiry, whether in the context of empirical observation, theoretical construction, or model-building. A second step requires representation of abstractions. In CSS this means representing abstracted social entities (e.g., actors, relations, institutions) in a way that a computer can understand sufficiently well to be able to execute a program about such entities.

[12] The late international relations theoretician Glenn H. Snyder [1924–2013] spoke often about this dichotomy, which he attributed to the philosopher of science, S. Toulmin.

Table 2.2 Main data types in Python

Type	Description	Examples
`str`	Alphanumeric text	United Nations, climate change, Leviathan
`list`	Mutable sequence	[7.4, ‘stress’, False]
`tuple`	Immutable sequence	(7.4, ‘stress’, False)
`set`	Group of unordered elements without duplicates	{7.4, ‘stress’, False}
`dict`	List of key-value pairs	{‘key1’: 2.57, 7: True}
`int`	Integer number	7
`float`	Floating point number	2.71828182845904523536
`bool`	Boolean binary values	True, False

Why does representation matter? The short answer is: because a computer can only understand sequences of the binary digits 0 and 1. In computer science, Donald E. Knuth is credited with playing an influential role in conceptually separating abstraction from representation (Shaw 2004: 68).

The more complete answer—to the question of why representation matters—warrants close attention. Earlier in Sect. 2.3.1, we distinguished between code (instructions) and data. In turn, data can be either numeric or alphanumeric, and numeric data can be either integer or real. Therefore, the information that needs to be represented to the computer (i.e., to both CPU and RAM) consists of four basic types: **real numbers**, **integer numbers** (positive or negative whole numbers, which include *ordinal variables*), **alphanumeric data** (including *nominal* or *categorical variables*), and **instructions**. Numbers, letters, and instructions are all represented in **bits** of information, consisting of sequences of the binary digits 0 and 1. More bits are necessary for representing more information.

Each programming language defines a set of data types as a semantic feature. The main data types defined in Python are summarized in Table 2.2. From a representation perspective, the Python interpreter translates each data type into binary code; i.e., every symbol in the syntax of a program (number, letter, or symbol) is represented as a sequence of the binary digits 0 and 1. The most commonly used data types are `str`, `int`, `float`, and `bool`.[13]

Representation can be seen as having two aspects. **Effective representation** refers to the choice of data types that helps answering the desired research questions(s). **Efficient representation,** on the other hand, refers to the choice of data types that minimize computational cost in terms of CPU cycles or RAM size. Achieving both effectiveness and efficiency is challenging.

[13] A boolean variable is called an “indicator variable” in probability and a “dummy variable” in social statistics and econometrics. (Dummy? As supposed to what? A strange phrase, don’t you think?)

Notation

Notation is necessary to express representations derived from abstraction. While in statistical and mathematical models, "notation" refers to equations and other formal structures (e.g., matrices, trees, graphs), in computational science the term refers to **programming languages** used to write software code. In 2004 it was observed that "hundreds of billions of lines of software code are currently in use, with many more billions added annually" (Aho and Larus 2004: 74). High-level programming languages (Python, Java, and many others) serve as bridges that span the "semantic gap" (Aho and Larus 2004: 75) between (a) the abstractions that we wish to investigate from the real world, and (b) binary notation understandable to computers. Without high-level programming languages a computational scientist would have no choice but to write software programs in binary code.

Several notational features of modern high-level programming languages (such as Python) are noteworthy:

Specificity: A programming language can be specifically dedicated to solving a narrow range of scientific problems, such as numerical computation, data visualization, or network dynamics.

Portability: A high-level programing language can be used to write code that executes in different computers, even those running different operating systems.

Reliability: Errors are difficult to avoid when writing low-level assembly language code, whereas they are more preventable with higher-level programming languages.

Optimization: While binary code executes at astonishing speeds (recall the earlier example of the MacBook Pro CPU cycling at many MHz), "a program written in a high-level language often runs faster" (Aho and Larus 2004: 75) because compiled code is highly optimized. Speed, memory, and energy are the most common goals of optimization.

Multiple approaches: High-level programming languages provide alternative and sometimes multiple approaches to programming, with emphasis on features such as imperative, declarative, and others.

Automated memory management: Information must be stored in main memory (recall Fig. 2.1), which is a major programming task when not automated. Automated memory management is a major useful feature of any high-level programming language.

Other features of modern high-level programming languages include procedures, patterns, constructs, advances in modularity, type checking, and other developments that are constantly being added to facilitate improvements in effectiveness of representation and efficiency of computation.

2.8 Objects, Classes, and Dynamics in Unified Modeling Language (UML)

A fascinating feature of social science is that the subjects of inquiry in the real-world social universe span a remarkable spectrum of ideas, entities, phenomena, systems, and processes, and have many ties to numerous other disciplines across the

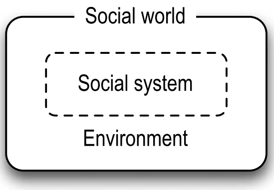


Fig. 2.2 A "social world" consists of a social system situated in its environment. This ontology is foundational for many social theories examined through formal and empirical analysis, including Simon's Theory of Artifacts, the Canonical Theory, and others based on the Complex Adaptive Systems Paradigm. Unfortunately, this graphic representation is useless although common throughout social science. Later in this section we introduce UML as a helpful graphic notation system for representing social worlds

sciences and humanities. The variety is so great that it is difficult to parse the entire landscape.[14] Not surprisingly, social science encompasses not one, but several disciplines (the Big Five: anthropology, economics, political science, social psychology, sociology) and related fields (communication, education, geography, history, law, linguistics, management), the totality of which is necessary to investigate the social world to understand it.

The vast landscape of all these disciplines includes an extraordinary variety of simple, complicated, and complex subject matter, much of which remains unknown and is poorly understood. So, these exciting scientific opportunities are innumerable! How does CSS handle such rich complexity to advance scientific understanding?

2.8.1 Ontology

Ontology refers to "what exists," or "the landscape of entities of interest," so to speak. It can be said that, from a high-level ontological perspective, the entire social world consists of **social systems** (which can be simple or complex; adaptive or not) and their **environments**, an idea introduced in Chap. 1 as a cornerstone of computational thinking about society and illustrated in Fig. 2.2. All entities in the social world (systems and environments) have a key **ontological** feature in common: they constitute **objects** and **classes** related by **associations** among them. An object belongs to a class, similar to the set-theoretic idea that an element is a member of a set. "Person" and "John Q. Smith," or "Country" and "Spain," are class and object, respectively, from an object-oriented (OO) computational perspective.[15]

The phrases "object-oriented modeling" (OOM) and "object-oriented programming" (OOP) denote, as the terms suggest, an approach to modeling (abstracting

[14]Winston Churchill (1948) said: "History is simply one damned thing after another."

[15]The idea of a tightly coupled relation between system and environment is also well-captured by the Spanish maxim, "*Yo soy yo y mi circunstancia*" (I am I and my circumstance), by José Ortega y Gasset (1914).

Fig. 2.3 Ontology across scales of human and social systems complexity: The family is the smallest kin-based social system (*upper left*). Teams of people provide assistance in humanitarian crises and disasters (*upper right*). Polities are complex social aggregates capable of producing historical milestones (*lower left*). Humans in space constitute complex, coupled, socio-technical systems operating in extreme environments (*lower right*)

and representing) that uses objects as the fundamental ontological entities. Note that the building blocks of computational methodology consist of social entities, not variables. (Variables come later, "encapsulated" in objects.)

Figure 2.3 illustrates people in four different social ontologies or "worlds." Let us consider each in some detail, from an "OO" perspective.

Upper left: **A family**. The first image shows a family consisting of a man, a woman, and a child as three distinct human entities that constitute a class we may call "people" or "family members." The basic association among them is defined by kinship. The environment is a professional photography studio that shows a white wall behind the family. From an OOM computational perspective, people and photo-studio are objects with attributes.[16]

Upper right: **Disaster victims**. The second image shows a team of humanitarian crisis workers and a victim being carried on a stretcher. The environment is a rural setting in the aftermath of a hurricane in Indonesia. The associations here are somewhat more complicated, involving collaboration among the aid

[16]For now, we don't care about the various features of entities. We'll explore that in the next section.

Table 2.3 Human entities and selected associations in socio-technical systems. Environments are named, not detailed

"World"	Classes	Objects	Associations	Environments
Family	Wife, daughter, husband	Sally J. Smith, Mary Smith, John Q. Smith	Mother-child, child-father, mother-father	Photo studio
Disaster situation	Aid workers, victim	J. Eno, T. Abij, unknown	Co-workers, assisted	Rural road
Leadership summit	Political leaders, aides	R. Reagan, M. Gorbachov, others	Speaker-audience	Urban location
Orbiting astronauts	Astronauts	J. Uko, K. Oli	Collaboration	Low Earth orbit

workers and assistance provided by aid workers to the victim. Here the objects consist of people, artifacts, and natural environment.

Lower left: **Leadership summit**. The third image shows a political gathering of heads of states and governments. Here the associations are even more complex, involving relations among people, polities, symbols, and historical events. The environment is urban, in 1980s Berlin, Germany. Still, the objects are the same: people and artifacts situated in some environment. In this case the environment is built (urban), not natural.

Lower right: **Orbiting astronauts.** The fourth image shows a contemporary space scene consisting of astronauts and a spacecraft (the International Space Station). This is arguably the most complex ontology of the four—a scene that would have been pure fiction just a few years ago. The environment is low Earth orbit (LEO) between 320 km (199 mi) and 400 km (249 mi) above the Earth's surface, orbiting at an average speed of 7,706.6 m/s (27,743.8 km/h, 17,239.2 mph). The objects are still people, artifacts (spacesuits, spacecraft), and nature ("empty" space and planet Earth).

The main purpose of OOM is to facilitate the abstraction of the most relevant set of classes, objects, and relations (associations among classes and objects) that we are interested in. After all, we can't represent the whole world, nor do we want or need to. We call the abstracted set a **model** or abstracted system, whereas the system in the real world is called the **referent system**, **focal system**, or **target system**.[17]

In spite of their diversity along numerous dimensions, from a computational perspective the four human situations or "social worlds" in Fig. 2.3 share a **common ontology in terms of entities and relations.** The entities, relations, and environments in Fig. 2.3 can be summarized as in Table 2.3 in terms of a socio-environmental perspective (Sect. 1.5.2). Note that this table is based on the process of *abstraction*, discussed earlier in Sect. 2.7. Obviously, each of the four social situations contains (much!) more detail than is abstracted in the table. But what *really*

[17]The three terms are synonymous. Target system is more common in simulation research, as we will see later. All three terms mean the same: the system-of-interest in the real, empirical world.

Table 2.4 Social, artifactual, and natural components of coupled systems

"World"	Social	Artifactual	Natural
Family	Family members	White wall in back	Indoors
Disaster situation	Relief workers and victims	Road, stretcher	Countryside Indonesia
Leadership summit	Leaders, staff	Monuments, flags, public address systems	Outdoors in Berlin
Orbiting astronauts	Astronauts	International Space Station	Near Earth orbit

matters is that three abstract categories alone (classes, objects, and associations) are universal across all social worlds. This fundamental ontology of CSS is consistent with classical social theory from ancient (Aristotle, Socrates, Plato) to modern (Parsons, Easton, Moore) and contemporary perspectives (including "constructivists").

The idea that objects of the same class share all common class-level features is called **inheritance** in object-oriented modeling. Thus, all wives are female, all husbands are male, all daughters have a mother, all disaster victims experience some level of stress, all political leaders govern through some base of support, all astronauts undergo many years of specialized training, and so on. Each object may also possess idiosyncratic features, but in order for it to belong to a class they must all share or "inherit" one or more features. Inheritance links classes and objects as a fundamental form of association.

Table 2.3 highlighted humans and associations among them, with only a coarse identification of the environments in which humans (social systems) are situated. A more complete abstraction, one based on the earlier socio-artifactual-natural perspective, is shown in Table 2.4. Now each type of "world" is decomposed (parsed) into three main components: the social (sub-)world is composed of the set of people and the set of social relations among them; the artificial component consists of built or engineered systems; and the natural component consists of the biophysical environment where the first two components (social and artificial) are embedded.

The buffering, adaptive, or interface character of artificial systems is highlighted by the ontological abstraction: Artifacts mediate between humans and nature, as the former adapt to the latter, following Simon's theory. In reference to Table 2.4 we see that:

1. The family is in a photographic studio room and only the white wall in the back is visible. Such an artificial room situation with highly controlled lighting conditions is necessary to ensure a high-quality portrait, as opposed to a more natural setting that cannot be as easily controlled.
2. The disaster situation is mitigated by the use of artifacts such as a re-opened road and medical equipment, in this case a special field stretcher. The uniforms of the relief workers are also functional artifacts, to protect the workers, to carry additional items, and to distinguish them from other members of the population.
3. The monuments, flags, and other stimulating symbols are used by leaders as artifacts to convey significance and power. Other artifacts consist of equipment for broadcasting and other communications infrastructure.

4. Astronauts use hugely complex artifacts such as spacesuits and the ISS to be able to function in the natural environment of orbital space, which would instantly kill them without such adaptive infrastructure.

In general: Artifact A is created for humans in social system S to perform in natural environment N. Symbolically, we might summarize this tripartite functional ontology as $A : S \leftrightarrows N$.

2.8.2 The Unified Modeling Language (UML)

Pictures and narratives, such as those used thus far, and other sources such as documents and data, can be informative, but they are usually insufficient for scientific purposes. They may tell us something about the focal world we are attempting to analyze, but are not very helpful in specifying the exact entities in terms of classes, objects, and their associations. Tables and other data can help, but can also be cumbersome for representing some features, such as complex relationships. The **Unified Modeling Language** (UML) is a standardized notational system for graphically representing complex systems consisting of classes, objects, associations among them, dynamic interactions, and other scientifically important features. Unlike most diagrams that appear in the social science literature, UML diagrams are rigorous, specialized graphics with specific scientific meaning—similar to a flowchart or a Gantt chart, where each symbol has specific meaning (semantics) and the arrangement of symbols is dictated by rules (syntax).

Although UML was created for representing systems of any kind, it is a valuable system for representing social systems and processes, given the lack of a standardized graphical notation system in the social sciences. There are different kinds of UML diagrams, because complex systems require alternative, complementary ways of modeling them—as is the case for any multi-faceted problem. There are three most useful UML diagrams for modeling social systems and processes: **class diagrams**, **sequence diagrams**, and **state diagrams**.[18] The first is used for representing *statics* while the other two represent *dynamics*.

Why, when, and how? UML was invented in the 1990s by a group of computer scientists and engineers that included James E. Rumbaugh, Grady Booch, and Ivar Jacobson. The Object Management Group (OMG) is the UML governance body that meets periodically to review and set standards. The original (and arguably still most prevalent) use of UML diagrams was to ensure that a diverse community of computer programmers and software engineers working on complicated code projects in large organizations (e.g., NASA, IBM, Boeing, Google) could work with a common understanding of a given programming project and collaborate effectively. Multidisciplinarity,

[18]The "state diagram" is also known as a "state machine diagram." We will use the simpler term "state diagram," without loss of meaning.

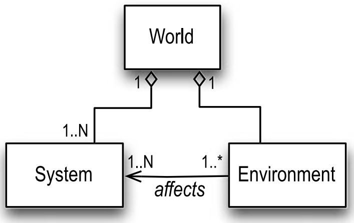


Fig. 2.4 UML class diagram of a basic world ontology consisting of a social system and its environment. Note that this graph is intended to represent the same as Fig. 2.2, but it conveys much more inforamtion

personnel turnover, multi-lingual requirements, and other complicating factors conspire against producing and maintaining excellent and sustainable code. UML diagrams help a modeler and programmer by providing graphic representations, of key aspects of a complex computational project, that are more inter-subjective than, for instance, narratives. The current UML standard is version 2.0, which is found at http://www.uml.org/.

2.8.2.1 Static Diagrams: UML Class Diagrams

A **class diagram** in UML is a graphic representation of the main entities and relations in a given social world or situation of interest. Figure 2.4 shows a simple class diagram of the general kind of social worlds we have been analyzing in the four instances discussed in this section (family, disaster, summit, astronauts): all four "worlds" consisted of a social system of some scale (small scale, as in the family, or large, as in the summit and space cases) and an environment of varying levels of complexity where the system was situated or embedded.

A UML class diagram consists of two main parts in terms of notation: *rectangles*, representing classes or objects, and *links* between them, representing associations, the labels and annotations of both are important. Rectangles are labeled by the name of each class or object (e.g., "world," "system," "environment"). Each association between entities (classes and objects) is also labeled by three elements: (1) an *arrowhead* symbol, (2) a descriptive *verb* describing the association (i.e., the role or function that one entity plays in terms of another), and (3) the *multiplicity* of the association, as defined below. In Fig. 2.4 the association between a system and its environment is denoted by the active but very general verb "affects;" the model does not include a reverse specification of anthropogenic effects (i.e., system feedback) on the environment, although in principle it could.

Dual graphic representations in UML. As with any graphic notation system in science, UML diagrams can be used to represent either the abstracted system (i.e., the model of reality) or a real-world system in greater detail than the abstracted model—for instance, as a reference of what is being omitted, if it is to added later. For example, there might be a UML diagram of a coali-

Table 2.5 Multiplicity values in UML class diagrams

Value	Meaning	Example	Mathematical notation
0..1	A range between no instances and one, meaning none or just one object	Number of prime ministers in a government	[0, 1]
1	One and only one instance	Each system has one environment	1
0..* or *	Range between 0 and unspecified many	Number of children in a family	$[0, +\infty]$
1..*	Range between 1 and unspecified many	Number of cities in a country	$[1, +\infty]$
0..N or N	Range between 0 and exactly N	Number of midlevel managers in a firm	$[0, N]$
1..N	Range between 1 and exactly N	Number of provinces in a polity	$[1, N]$

tion being modeled, as well as a more detailed UML diagram of a real-world cabinet system with details on support from the multi-party system. The *most common use* of a UML diagram is for representing a model in terms of its abstracted components, not the real world. However, nothing prevents the use of a UML diagram for describing a real target system of interest if that is helpful. This may happen for a number of reasons, as when we wish to highlight the difference between a target system and a simulation model of such a system. The difference between a model diagram (abstract) and realistic diagram (empirical) would highlight all those elements omitted by the abstraction.

The concept of **multiplicity** is fundamental in computational modeling, although it is often neglected or left mostly undefined or it is implicit in more traditional social science theory and research. Multiplicity refers to the precise number of instances of a class or object. The notation in Table 2.5 is standard for specifying the multiplicity of entities in UML diagrams. We will be using this notation throughout this textbook, so it is important to master it, although it may be omitted in a summary diagram. Note that the symbol ".." (two periods) is used in computational UML notation to signify a range of values, rather than the more traditional mathematical notation "..." (called ellipsis).

For example, in Fig. 2.4 there is one World entity (a class) consisting of one or more (up to N) System entities and one or many (up to an indefinite number, represented by the asterisk symbol "*") of Environment entities affecting the System. The multiplicity of World is implicitly one in this case, so the value of 1 is normally omitted because it is redundant (unnecessary). Later we will examine other examples.

Social Science dedicates a great deal of effort attempting to describe and understand *social relations* among various *entities* (actors, their beliefs, institutions, and

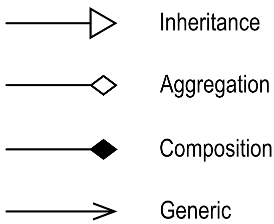


Fig. 2.5 Associations among classes or objects are drawn in UML using arrows with different arrowheads that denote different types of relations (e.g., social relations, socio-environmental interactions, or others). Unlike the informal and widespread use of arrows in many social science illustrations, the notation for social relations modeled with a class diagram is formal and strictly defined, making meanings inter-subjective and reliable from a conceptual and terminological perspective. Examples of each type of social relation are provided in the main text

their environments, among others). In UML the type of association that is assumed to exist between entities is denoted by the special form of the link's **arrowhead**. (As we will see, this is not arbitrary or esthetic, as in most traditional social diagrams! The form of an arrowhead has precise meaning in UML.) In the case of Fig. 2.4 the association between World and Environment is one of aggregation (hence the white diamond-head, as explained below), because the World class is being modeled or specified as consisting of two component classes: the social System of interest and the Environment in which such a system is situated, with the latter "affecting" the former. (For now we need not worry about the meaning of the term "affects"; the common meaning will suffice.)

The four most common **types of association** are called "inheritance," "aggregation," "composition," and "generic," which are distinct types of social relations denoted by the symbols illustrated in Fig. 2.5.

Earlier we encountered **inheritance** (empty arrowhead symbol) when discussing the association between classes and objects, in the sense that an object is an instance of a class, such that all objects belonging to the same class are said to share or "inherit" a common set of characteristics. The inheritance association is also called the "is a" relation. It is denoted by an arrow with a blank arrowhead. In Fig. 2.4 we saw an example of aggregation and a generic association relationship ("affects").

The following are examples of the inheritance association (one from each of the Big Five social science disciplines):

- **Politics: Political regimes**. Consider the class "Political Regimes." It contains classes such as "democracies" and "autocracies," both of which represent particular forms of political regimes but also share in common many features having to do with the relationship between society and government. Thus, both democracies and autocracies are said to inherit the properties of the class "Political Regimes."
- **Anthropology: Social complexity**. The classes "band," "tribe," "chiefdom," and "state," from anthropological archaeology represent ordinal forms of social complexity. All these forms inherit the features of a broader class that may be called "Polity." All polities—and therefore all chiefdoms, states, and empires—include a "society" (population, community) and a "system of government." In turn,

all systems of government share some common features, such as constitutional regime (defining the society-government relationship), bureaucratic structure, support mechanism, public finance (resource base), policy-making process, and other constituent or defining features.

- **Psychology: Cognitive balancing (Abelson 1959)**. The objects "Differentiation," "Bolstering," "Denial," and "Transcendence," are instances of the broader class of "Cognitive Balancing Mechanisms." All four mechanisms serve the purpose (have the function) of resolving or mitigating cognitive inconsistencies that arise in human complex belief systems.
- **Economics: Goods**. The classes "commodity goods" and "luxury goods" inherit the features of the broader class of "private goods." An instance of the private good-class, such as a 2012 Ferrari racing car, is an object, due to its concretely empirical specificity. All private goods share some common features, such as, for example, quantity produced, price, provenance, production method, and useful life, among others. In turn, "private goods" belong to the superclass of "economic goods," which also comprises "public goods," such as "clean air" and "public security."
- **Sociology: Organizations**. The class "organizations" comprises "private organizations" and "public organizations." Both types (whether private or public) inherit all the features of the former, such as mission, size, structural features, age, and domain of activity, among others. In addition to class-level features, both private and public organizations have other features, such as membership characteristics for private organizations or public finance for public organizations.
- **Aristotle's Classification of Governments**. Aristotle [384–322 B.C.] was the first comparative social scientist of whom we have a surviving record. The Aristotelian classification of governments distinguishes between normal and degenerate forms of government. The three normal forms are monarchies, aristocracies, and democracies, while the degenerate forms are tyrannies, oligarchies, and ochlocracies, respectively. Thus, an abusive monarchic ruler yields a tyranny; degenerative rule by an elite produces an oligarchy; and extreme democracy yields an ochlocracy (literally, "mob rule"). Representative government (e.g., as in a parliamentary system) is a regime that attempts to implement democracy to avoid ochlocracy (as occurred during the Reign of Terror, A.D. 1793–1794, in France). All six types inherit all the features of the class Government, with each type having additional characteristics.

These examples illustrate the inheritance association, which is represented by the empty arrowhead in Fig. 2.5. A UML class diagram of each example would include the main entities and the inheritance association link annotated with the multiplicity of each entity (class or object).

The next two types of association—called "aggregation" and "composition"—apply to *compound social entities.*[19] Committees, belief systems, organizations,

[19] A compound social entity C may be thought of in a similar way as a compound event in probability theory. Accordingly, C consists of several smaller parts or subsystems "smaller" than C, similar to the way in which a compound event is defined as a function of its conjunctive elementary events (sample points).

and whole polities, economies, and societies are prominent examples of compound social entities. CSS examines these compound entities by distinguishing between those that are structured by aggregation versus those that are structured by composition.

The second type of association is called **aggregation** (empty diamond arrowhead), which has the conceptual meaning of "consists of" in natural language. Aggregation is also called the "has a" relation. This is a loose type of collection, which in some cases may just be ad hoc (as opposed to the stronger form of membership rule implied by the composition association, discussed below). The following are examples of aggregation in compound social entities:

- A human belief system consists of concepts (represented as nodes) and associations among them (valued links).
- A family is a social aggregate consisting of parents and children.
- A society is comprised of individuals that share a set of commonly held attributes.
- An economy is composed of producers, consumers, and lenders.
- A coupled socio-techno-natural system consists of interacting social, artifactual, and biophysical components in interaction with one another.

A key feature of aggregation is that the members can survive without the aggregate, as in the examples above. Parents and children do not cease to be such when a divorce occurs. Concepts that are part of a belief system can endure after a belief system is no longer accepted. Producers, consumers, and lenders can endure even after an economy disintegrates.

Aggregation is denoted by the empty diamond arrowhead in a UML class diagram (Fig. 2.5), with the arrowhead pointing to the higher-order class (superclass).

The third type of association is called **composition** (solid diamond head symbol), which is a stronger form of aggregation. Composition is used instead of aggregation when member classes have a constituent relationship with respect to the superclass; i.e., when the set of member classes cannot exist without the superclass. Accordingly, composition can also be called the "is constituted by" relation, similar to "is a" and "has a" for inheritance and aggregation. Under composition the superclass compound is said to "own" the member classes of the compound entity, in the sense that if the superclass dies—or somehow is destroyed—so do the classes under it.

The following are examples of association by composition in compound social entities:

- A bureaucracy is an organization composed of bureaus or administrative units. The units exist by virtue of their contribution to the overall organization.
- The provinces, counties, and other administrative units of a country are associated to the larger country by composition.
- As a compound social entity or "body," a given committee with members playing various functional roles, as in the case of a ministerial cabinet, is linked to its members by composition. A cabinet minister does not exist without there being a cabinet. This is normally defined by a constitution.

- The institutions of international governmental organizations, such as the General Assembly and the Security Council of the United Nations Organization, or the Commission and Parliament of the European Union, are associated to the organization by composition, not just aggregation.

Many aggregate entities of common interest in social science are in fact compositions, not mere aggregations, because component classes are defined as a function of some superclass (compound social entity), such that parts are meaningless without the whole. When social scientists speak of "the importance of context," they often have in mind the composition association of compound social entities, rather than mere aggregation. Context can matter, precisely because some constituent social entities are fundamentally (constitutionally) dependent on larger compound entities only through association by composition.

The key *difference between aggregation and composition* is conceptually subtle, significant (theoretically and empirically), and unfortunately quite often left implicit in social science theory and research on compound entities of all kinds, which consist of actors, events, systems, and processes. The multiplicity of aggregation can assume any value (i.e., the natural numbers or positive integers $1..N$), whereas the multiplicity of composition is zero or one on the compound, higher-order class (the superclass). Testing this idea with examples is good for understanding the difference. Whether associations or relationships in a compound social entity are either by aggregation or by composition is something that should be decided and denoted accordingly in a well-specified UML class diagrams, or the unresolved ambiguity can result in confusion leading to modeling errors in implementation. Formally, composition spans a *tree*, whereas aggregation forms a *net* (Eriksson et al. 2004: 113).

Some compound social systems have **hybrid associations**, as shown in Fig. 2.6. An example is a polity P, which consists of a given society S and a system of government G for addressing issues I that affect members of S. Whereas G is "owned" by P (hence composition specifies the polity-government association), in the sense that it makes no sense to think of a governmental system except within the context of *some* polity, society S is an aggregate that has autonomy regardless of whether or not P exists (aggregation specifies the polity-society association), since S is an association among people in terms of identity and other features (whether members of some elite or the mass public). The class of issues I is also related to P and G by association, because issues affecting S can persist regardless of P and G. The compound social system P is therefore a hybrid of compositions (in G) and associations (in S and I).

Inheritance, aggregation, and composition have their own special ad hoc symbols because they are so common. Finally, a fourth type of association is called **generic** (plain arrow symbol), which is a category intended to represent any association in terms of a verb connecting any two entities. Generic association is symbolized by a simple arrowhead and the verb that best describes the association. For example, the association between Environment and System in Fig. 2.4 is represented by the simple arrow from E to S and the verb "affects" describing the association. Similarly, in Fig. 2.6 there are three generic associations represented: Public Issues affect Society, causing stress; Society places demands on Government to deal with issues;

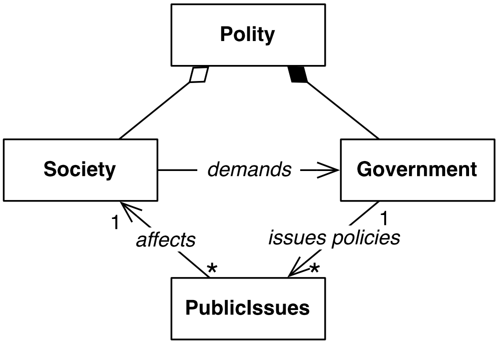


Fig. 2.6 UML class diagram of the standard model of a polity in political science. The diagram consists of four entities and three types of associations that denote different kinds of social relations, as explained in the main text. Diagrams such as these, and subsequent versions with more details, are valuable for communicating between social science modelers and computer programmers in charge of code implementation. Adapted from Cioffi-Revilla (2008)

and Government deals with issues by issuing (i.e., formulating and implementing) policies that mitigate stress on Society.

2.8.2.2 Dynamic Diagrams: UML Sequence and State Diagrams

In addition to the static diagrams introduced so far, the Unified Modeling Language also provides standardized graphics for representing dynamical aspects of social entities; i.e., social *processes*. Two of the most common dynamic diagrams are those called "sequence diagram" and the "state machine diagram." Other dynamic diagrams include "activity diagrams" and "communications diagrams" (Eriksson et al. 2004).

A **sequence diagram** portrays dynamic interactions that take place in a social process among entity components. Figure 2.7 shows a UML sequence diagram for the standard model of a polity represented earlier in Fig. 2.6. There are three main components in a sequence diagram: (1) a set of separate vertical "lanes," each representing the main interacting entities in the compound superclass (e.g., in this case PublicIssues, Society, and Government); (2) arrows indicating various activities among entities; and (3) a summary natural language chronology of main events of interest (left), which should say in plain English what the sequence diagram is intended to graphically represent.

Several features of the UML sequence diagram are noteworthy from the example in Fig. 2.7:

1. UML symbolic notation is standardized and systematically developed, not arbitrary. This enables researchers to communicate using a common set of universal symbols that have been agreed upon. By contrast, most traditional social science diagrams are drawn using ad hoc symbols often invented by an author and used by no one else.

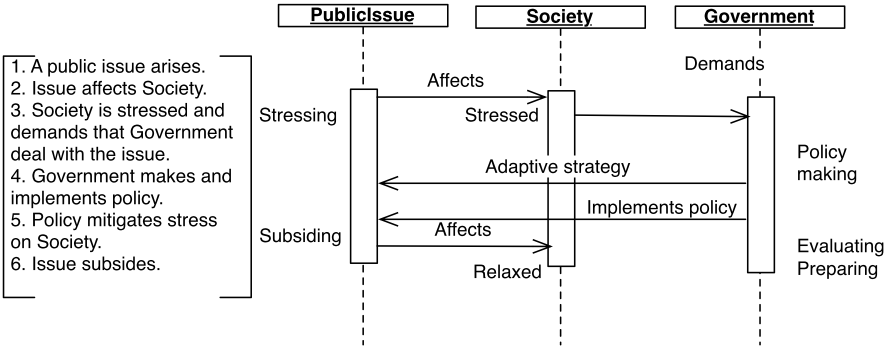


Fig. 2.7 UML sequence diagram of basic dynamic processes in a simple polity

2. A diagram like this cannot be drawn without fairly precise understanding of the social process being represented. At a minimum, a researcher needs to stipulate or hypothesize some parts of the process where theoretical explanation or empirical descriptions are missing in the basic relevant science.
3. The basic space-time ontology of the social process is discretized—in terms of classes and objects (social space) and events (time)—not continuous. This enables the specification of precise interactions and their sequence within an overall framework.
4. The diagram is ordered by time, flowing from top to bottom, as in an historical timetable or a flowchart. Thus, addition or deletion of events requires shifting down or shrinking everything downstream.[20]
5. The information dimensionality of the basic notation is simple, so much room is available for increasing the information content of the diagram by use of color, tones, patterns, and additional shapes.
6. A potentially significant drawback of the sequence diagram is its tendency to become too cluttered when more than a few objects or classes ("lanes") must be represented. Having to represent a process with many objects or classes almost guarantees an unreadable or messy diagram, so the sequence diagram does not scale well with respect to the cardinality of the ontology being modeled.

Another type of dynamic UML diagram is the **state diagram**, which represents transitions between macroscopic states of a system during its typical operating or life cycle. Figure 2.8 shows the state diagram for a polity, based on the standard model that we introduced previously. This diagram consists of three components: (1) a set of start-end states, represented by large black dots; (2) a set of labeled possible contingencies represented as transitions between states; and (3) a set of labeled states representing the condition of the system as a result of each transition.

[20] By contrast, archaeologists draw timelines from bottom to top, consistent with stratigraphic analysis, such that the oldest date is at the base.

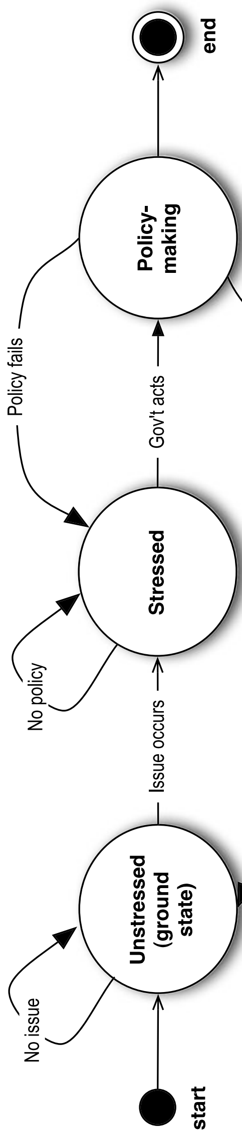


Fig. 2.8 UML state (or "state machine") diagram of macro system-level dynamics in a simple polity consisting of a Society stressed by issues and a Government that formulates policies to address public issues and lower or eliminate stress. A state diagrams provides a more dynamic model of a polity than a class diagram, but entities (classes, objects) are not represented. *Source:* This and other UML diagrams of a polity are adapted from Cioffi-Revilla (2008)

A UML state diagram depicts various states of the system and possible transitions between states. In this case the polity starts out in an unstressed state, which we may call a **ground state**, since Society S is not initially affected by any issues—everything is fine. When S is in an unstressed state, two things can happen: either some issue arises or it does not. If no issue arises, then the state of S remains unstressed (shown by the loop arrow labeled "No issue"). However, if an issue arises, then the state of S transitions to being stressed (by the issue). When S is stressed, two things can happen: either Government G pays attention and produces policy, or G fails and policy is not produced (the "No policy loop arrow"). When G produces policy, two things can happen: either policy fails, or it works. If policy works, then S is again unstressed. The process continues or ends after policy is produced.

This single-issue account of the standard polity and policy process is intentionally simplified (abstracted) for illustrative purposes. One simplifying assumption is that public issues affect S without any anticipation on behalf of G, which is sometimes (not always!) unrealistic. For example, in the case of many domestic policies, G often prepares by producing anticipatory mitigating policies. Another simplifying assumption is the direct, unmediated pressure of S on G, without intermediaries. In fact, interest groups and other intermediary groups (e.g., lobbies, unions) act between S and G, producing more transitions and additional intermediary states before policies are produced. Finally, another assumption in Fig. 2.8 is that the policy-making process is finite, rather than going on forever.

The UML state diagram is characterized by a set of features, as seen from Fig. 2.8:

1. The diagram is read from left to right, with various possible transitions and loops as the state of the system evolves to the end of each cycle.
2. The diagram is reminiscent of a Markov model representing the states of a system and possible transitions, minus the start and end states. Unlike a Markov model, however, transition probabilities are not generally represented.
3. The diagram is also reminiscent of a flowchart, but with exclusive emphasis on the state or condition of the system.
4. Classes or objects (entities) do not appear in a state diagram. Instead, this type of diagram focuses on the state or condition of the system, given the possible interactions among entities.
5. Each transition is specified by asking "what can happen next?" given some state.
6. The state diagram is formally a graph. Specifically, it is a directed graph. It can also be weighted, if transition probabilities (or other measures associated with the links between states) are known.

State diagrams such as the one in Fig. 2.8 are sometimes difficult to specify in a way that is sufficiently complete or precise, in part because the detailed dynamics of a social system and process may not be clearly understood. In that case, resort to other sources such as narratives or other diagrams may prove useful. For example, a classical flowchart may be helpful for uncovering the information needed for a state diagram.

When attempting to specify a detailed state diagram, it is always good practice to begin with a simple version with the fewest possible number of states and transitions.

Other UML diagrams include activity diagrams and use case diagrams. We will use UML diagrams throughout this book for two main purposes: increasing scientific clarity and enabling computational specificity. Both uses are new in social science.

2.8.3 Attributes

Now that we have covered the basics of classes, objects, and associations, we must take a closer look at them by focusing on two key computational aspects: their defining attributes and operations. We will do this assisted by some further UML notation created precisely for dealing with attributes and their operations, as in Fig. 2.9.

We will approach the parts of Fig. 2.9 in sequence, with the last part (e) being the most complete in terms of specification. To begin, note that the following notational details in Fig. 2.9 are standard and important, not arbitrary:

1. Class and object names are written in the center of the first compartment of each diagram, with initial capital letter (as in a proper noun), preferably in boldface (e.g, Class, Object, Polity, Switzerland).
2. The name of an object is underlined (Object, Province, County, City).
3. Attributes are written with the first letter in lowercase, followed by additional words without spacing (e.g., classAttribute1, classAttribute2, popSize, capitalCity, numberOfiPhones, inflationRate), always left-justified.
4. The data type of each attribute is written after the attributeName.
5. Operations are written in a similar way, followed by left and right parentheses.
6. The so-called "visibility" or "accessibility" of attributes and operations is denoted by plus and minus signs, representing their public or private status, respectively, as explained further below.

A feature, variable, or parameter that characterizes a social entity is called an **attribute** in CSS. Attributes are familiar concepts in Social Science, often under the name of "variables" or "parameters." The following are some illustrative attributes, based on earlier examples in this chapter. In the case of a coupled socio-techno-natural system, we may model the natural environment as consisting of ecosystems with biophysical attributes such as biomass distribution, climate variables, topography, and others. Similarly, social attributes are often used to characterize various actors and groups abstracted in a model, such as economic, political, and social variables. In the case of a polity, commonly specified attributes include population size, size of its economy, territorial extent, cultural indicators, military capabilities, and other numerous features. Each social object or class is always defined in terms of some set of attributes.

In Figs. 2.9(b)–(e) we saw how attributes are annotated in the second compartment of a UML class diagram. Figure 2.10 shows how this is done in a more complete UML class diagram, as part of each class or object, in this case using the Polity model discussed earlier. In this case we have chosen to abstract the following class attributes: the name, continent in which it is located, territorial size, and name of the polity's capital city; the population size and amount of resources of the society

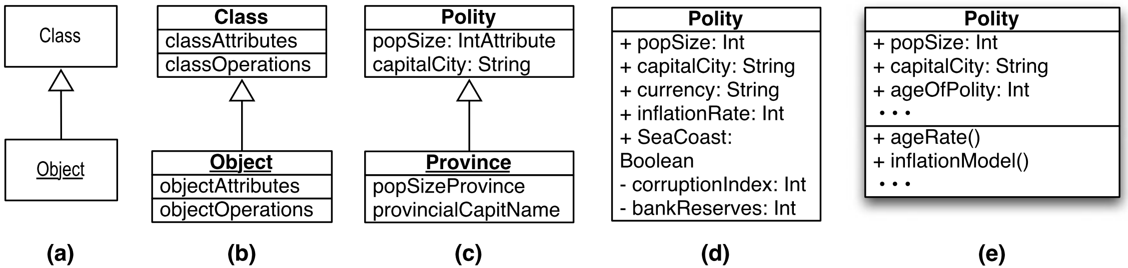


Fig. 2.9 UML class and object diagrams with various specifications of attributes and operations: (**a**) Class and object associated by inheritance, without specific attributes or operations, as in earlier class diagrams. (**b**) Class and object notation containing encapsulated attributes and operations shown by convention in the second and third compartments, respectively. (**c**) Example of class and object with some specific attributes. (**d**) Visibility of attributes denoted by public (*plus sign*) and private (*minus*) attribute notation. (**e**) Complete specification of a class with encapsulated attributes, operations, and visibilities

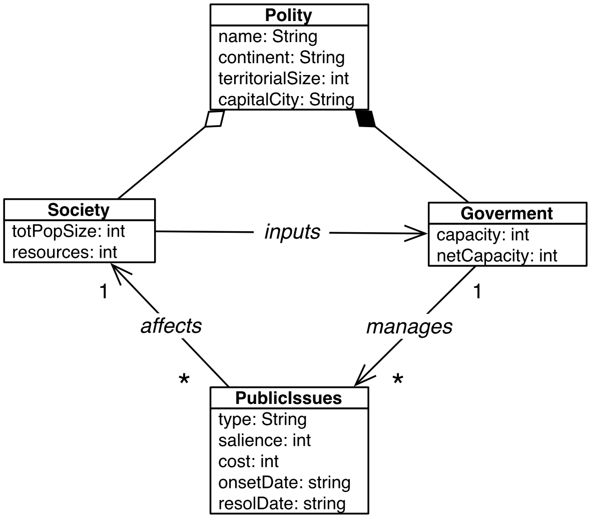


Fig. 2.10 UML class diagram of the standard polity model, with specified attributes (variables). Note that each attribute is denoted by a uniquely designated name and corresponding data type

of that polity; the government's gross capacity and net capacity for policy-making and implementation; and the type, salience, cost, and onset and resolution dates of public issues that arise in the polity.

Figures 2.9(d) and (e) also show the **visibility** or **accessibility** of each attribute by using plus and minus signs. This is a feature of attributes and operations that defines the status of information in relation to other classes. Specifically:

1. An attribute is **private** when it can be accessed only from its own class, denoted by the minus sign −.
2. An attribute is **public** when it can be used and viewed by any other class, denoted by the plus sign +.
3. An attribute is **protected** when it can be accessed only by its class or subclasses, denoted by the pound sign #.

The attribute of an object is called a **object variable**, to distinguish it from the class-level feature. This nomenclature is consistent with the earlier idea that an object belongs to a class. Similarly, the attribute of a class is also called a **class variable**.

2.8.4 Operations

We saw earlier (Fig. 2.9(b)) how attributes and operations define a class. An **operation** changes the value of one or more attributes, and consequently the state of objects and classes. At the object level, an operation is called a **method.** Operations and methods are implemented by functions in Python. A common example of an operation is a function that would specify how the population of a polity changes each year. Operations specify dynamics, whereas attributes define statics; both determine the state of classes and objects.

Figure 2.11 shows how operations are added to the third compartment of a classes's box to complete the model in greater detail, extending the earlier model in Fig. 2.10. This is the same familiar model of Polity, only now we have added some operations that tell us how attributes are supposed to change in each class. For example, in the Polity class, the attribute (or class variable in this case) called ageOfPolity will change as specified by a function called agingRate(), which is defined in the third compartment of Polity. This is presumably a simple function that returns an annual increment of 1.0. Similarly, the attribute called corruptionRate in the Government class is driven by corruptionChange(), which is a more complicated operation defined in the third compartment of Government. For example, corruptionChange() might be specified or modeled as a function of other attributes, such as levels of foreign investment, literacy, rule of law, and other variables (i.e., the known determinants or drivers of governmental corruption reported in the empirical literature) that are located in the same or other classes.

In any social system some associations are more important than others. For example, note the "manages" association between Government and PublicIssues in Fig. 2.11(left). This is a very significant relation between two major entities of a polity, which in this case abstracts the notion of a policy. It is through policies that governments address public issues. Thus, the seemingly simple association between Government and PublicIssues should be elevated to the higher status of having a class by itself, as **association class** named Policy. As shown in Fig. 2.11(right), an association class is denoted by the same class notation, joined to the association link by a dashed link. To decide whether a given association warrants the status of being modeled as an association class, rather than a mere association, the following heuristic questions are helpful:

1. Does the association in question have significant attributes that can be specified?
2. If so, what are they?
3. Moreover, do such attributes have operations that can be similarly specified?

If the answer is yes to questions 1 and 3, then the association in question is a candidate for promotion to association class status. For example, in the previous case it is certainly, true that a policy has attributes, such as type (economic, social, political, environmental, or other), effectiveness (degree to which it is likely to solve the issue), efficiency (cost/benefit), and other features. However, whether promotion to the status of association class, rather than mere association, is warranted is a different question, which depends on research questions and not just our ability to identify relevant attributes.

Both attributes and operations are said to be "encapsulated" within a class or object. "This process of packaging some data along with the set of operations that can be performed on the data is called **encapsulation**" (Zelle 2010: 418). Encapsulation is a powerful, defining feature of all OOM and OOP. In UML modeling terms this means that all attributes and operations must always appear contained within the second or third compartments, respectively, of *some* class or object entity—never unassociated, by themselves. More importantly in OOP, encapsulation means that classes and objects can interact without having to access inner components or computations that can be hidden within entities. All fully OOP languages implement

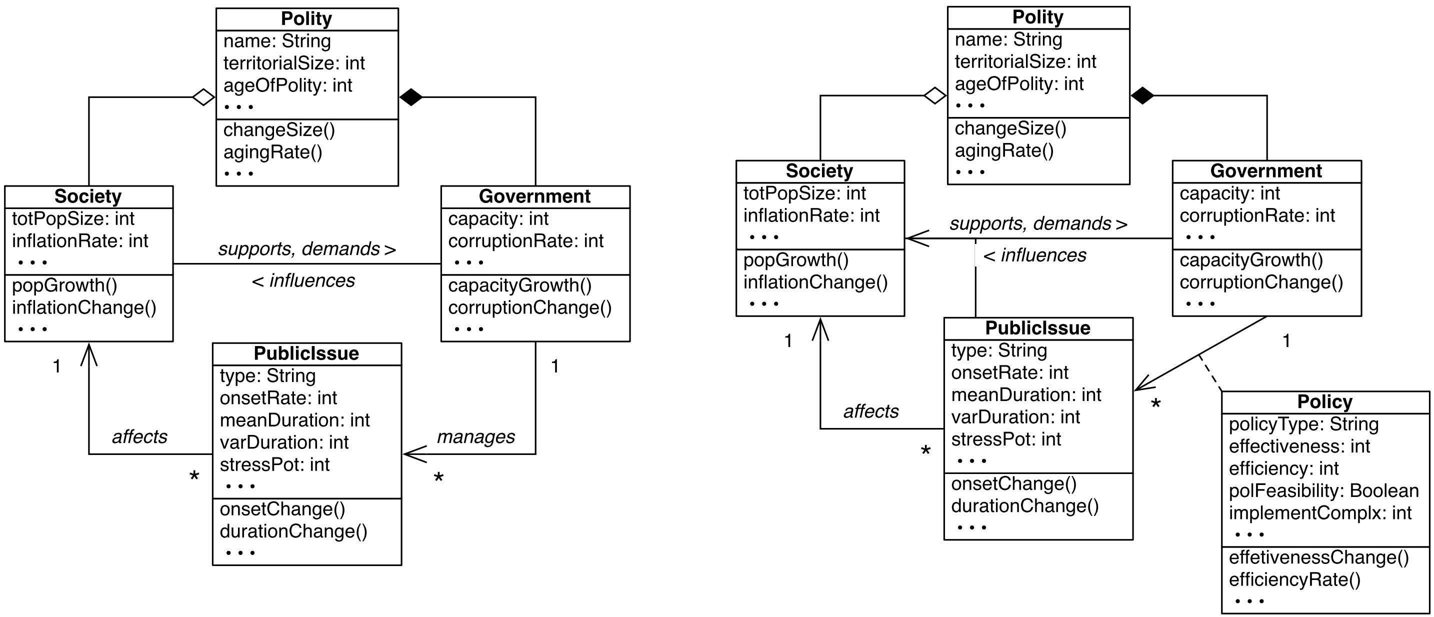


Fig. 2.11 UML class diagrams of a polity with class attributes and operations. The model on the left shows operations in the third vertical compartment of each class. The model on the right makes explicit the "manages" association between Government and PublicIssues, elevating the association to the higher status of a class by itself, named Policy

encapsulation, whereas most procedural languages do not. Hence, implementing social models comprised of entities that encapsulate attributes is best accomplished in an OOP language so as not to risk breaking encapsulation.

Python implements encapsulation as a convention, as part of proper programming style, and is not an absolute requirement of the language. By contrast, encapsulation is a required feature of abstraction in Java.

Encapsulation implies that variables and methods/operations are always defined as belonging to some object or class, never by themselves. Other common language phrases used to signify encapsulation are "in the context of," "with respect to," and "in relation to," among others. For example, the context for the variable inflation is an economy, the context for corruption is government (or business), voting behavior is associated with a polity, and so forth. From an OO perspective in Computational Social Science, variables or parameters make no sense by themselves, in isolation. Hence, they are always encapsulated within some class or object. *Understanding this idea also provides a powerful principle for turning a variable-based model into a potentially more powerful object-based model.*

2.9 Data Structures

Classes and objects represent one form of data that encapsulates a set of attributes/variables and operations/methods for changing the value of attributes/variables. However, data come in many forms—not a surprise in social science! We have already seen various value types for variables, such as **integer**, **string**, and **boolean**. The term **data structures** refers to the various ways in which data are organized for purposes of computation. Sometimes the same information is organized in different ways, so it will be structured differently, depending on computational need. When it comes to data structures, remember the famous design principle from architecture: "Form follows function" (Louis Sullivan, American architect, 1896).

The following are the most common data structures, listed in order of generalization:[21]

Tuple: A tuple is similar to a record structure, the main difference being that individual records need not be arranged as in the 2-dimensional structure typical of a spreadsheet. Elements of a tuple must all have the same type. Examples: calendar dates expressed by year, month, and day; N-dimensional Cartesian (or other coordinate system) n-tuple of coordinate values for a point $(x_1, x_2, x_3, \ldots, x_N)$; payoff values (u, v) in a 2×2 normal form game Γ, where u and v are the payoffs for each player. The elements of a tuple are ordered.

Array: An array has elements of the same type accessible by some index. Examples: all vectors and matrices; input-output table of sectors in an economy; adjacency matrix of a network. A vector is a one-dimensional array, whereas

[21] There are as many kinds of data structures as there are ways in which information can be organized. The US National Institute of Standards and Technology (NIST) provides a comprehensive, encyclopedic online survey (Black 2004).

a matrix is a 2-dimensional array. A vector is a datum with both scalar value and direction (whereas a scalar lacks direction). A "data cube" is a 3-dim array (e.g., countries $\times$ attributes $\times$ years). A *sparse array* is one where many entries are zero or missing, which may be better structured as a list.

List, or **sequence:** A list is a mutable tuple of variable length, with the first element called the *head* or *header*, and the ones that follow are called the *tail*. Examples: Cities ranked by population size, the head being the largest; conflicts or disasters ordered by severity, the head being the worst case; network nodes arranged by the number of links with other nodes (called *degree*), the head being the node with highest degree.

Queue: A list of items where the head is accessed first. Examples: legislative bills in a calendar for voting; items on a formal agenda; refugees arriving at a camp site; military units being deployed. Operations defined on a queue include addition (new value is added to the tail), deletion (from the head), as well as others. A queue is also called a FIFO (first-in-first-out) list, or pushup list. Queues are also a significant social process, so half of Chap. 9 is dedicated to them.

Stack: A stack is a data structure consisting of an ordered list of data such that the datum inserted last gets drawn first. Examples: location visited most recently; the most recent acquaintance; the most recent course taken by a student or taught by an instructor, from among a complete list of courses taken or taught, respectively. Chronological order of entry into the data structure is a key idea in a stack.

Bag: A bag is a set of values that can contain duplicates. Examples: the set of all countries that have experienced civil war during the past τ years; a list of individuals who have voted in the past N elections; the set of terrorist organizations that have launched suicidal bombing attacks since 9/11. The term *multi-set* is synonymous with bag.

Set: A collection of elements in no particular order with each element occurring only once. Examples: The set of cities in a given country; coalition members; candidates in an election; budget priorities; major powers in the international system (*polarity*); nodes in a network; legislative bill proposals in the "hopper." This is a general and powerful mathematical concept with broad applicability across the social sciences.

Hash table: Also known as a **dictionary**, a hash table is a data structure in which values and keys are assigned by a function, called the hash function. A hash table is an array of 2-tuples consisting of values and associated keys, such that there is a one-to-one mapping between values and keys (binary relation). The list of values is also called a hash table. Examples: a telephone directory; a list of voters and their voting precincts; administrative units (counties, provinces, states, countries) and abbreviations or codes; items and barcodes; geographic gazetteers; organizational charts; course catalogues. A hash table provides a fast way to lookup data.

Tree: A tree is a data structure consisting of a *root* element with *subtrees* branching out to terminal *nodes* called *leaves*. Nodes located between the root and leaves (i.e., "crotches," in common language) are called *internal nodes*. A taxonomy has the structure of a tree. Examples: classification of social entities; extensive form games; tree of phone calls for emergencies; hierarchal organization in business and public administration; star network; population settlement pattern (capital [root], provincial centers, town, villages, hamlets [leaves]). Tree-like data structures are ubiquitous in social systems and processes, but they are rarely analyzed as such.

Graph: A graph is a generalization or extension of a tree, in which nodes and links (also called arcs or edges) can be arranged in any way, as we discuss in detail in Chap. 4.

Note that data structures do not contain any code; they just contain data organized in various ways.

A **record** is like a composite data type rather than a true data structure, in a strict sense. It consists of information fields or members comprising a set. Examples: a person with contact information (address, telephone, email, Skype address); a polity profile (country name, capital city, total population, and other attributes); a bibliographic entry (author, title, place and date of publication); events data (actor, target, date, descriptive verb, other event attributes). A spreadsheet entry is often like a set of records, with columns representing various fields, as is a common in social science datasets.

All of these data structures can be used in the Python and Java programming languages (and many more). Python can handle lists of many types, including stacks, queues, matrices (a list of lists), tuples, and sets, among others. A set of operations (functions and methods) is defined for various data structures in each programming language.

2.10 Modules and Modularization

In all but the simplest cases, a computer program usually requires "parsing" into main components and subcomponents. This is because writing a long, "monolithic" program is impractical as soon as the program requires more than just a few lines of code (LOC). **Modularization** is not just a programming style; it matters greatly in terms of overall program performance.

One way to think of modularity is in terms of performance: how should a given computer program be written in order to maximize its speed? Intuitively, there may be many ways in which a computer program could be modularized. For example, computation and visualization could be separated; but so could various stages of execution, in sequential fashion, as derived from a flowchart. The way in which a given program should be modularized into parts is not necessarily obvious. David Parnas (1972), a famous computer scientist, introduced the influential **Principle of Decomposition by Information Hiding**. Given a program P, the **Parnas Principle** states that P should be structured in nearly-decomposable modules, such that each module encapsulates a nearly self-contained (encapsulated) cluster of instructions *and* the interface between modules is such that it minimizes "communication overhead."

Direct quote: "… one begins with a list of difficult design decisions or design decisions which are likely to change. Each module is then designed to hide such a decision from the others. Since, in most cases, design decisions transcend time of execution, modules will not correspond to steps in the processing. To achieve an efficient implementation we must abandon the assumption that a module is one or more subroutines, and instead allow subroutines and programs to be assembled collections of code from various modules" (Parnas 1972: Conclusions).

The following are significant advantages of Parnas-modularity:

- Modules are easier to understand.
- Independent programmers can work on different modules.
- The program can be more easily changed.
- Sensitive information may be more easily protected.

The overall structure of a modular program is that of a network composed of any number of communicating clusters, as in a cellular network (similar to the Horton or Tutte graphs), such that most of the communication takes place within clusters and minimal communication across them.[22]

2.11 Computability and Complexity

Consider the following questions:

1. A leader needs to form a coalition in order to ensure security against a powerful adversary. Given a set of potential allies, what are the possible combinations that might produce successful, winning coalitions?
2. A person involved in a disaster faces a set of competing priorities (safety, family, shelter, neighbors, supplies), which can induce severe frustration, compounded by fear and uncertainty. Which course of action is best, or at least satisfactory?
3. A country affected by climate change must choose from among a set of competing policies, finite resources, and imperfect information. How can policy analysts arrive at defensible recommendations for policy-makers?

Questions such as these require complex social computations, not just in terms of crude costs and benefits, but also in probabilistic assessments, alternative combinatorial arrangements, fitness assessments with respect to known empirical patterns, and other computational features. The necessary science (social or natural) may also be incomplete, so allowance must be made for deep uncertainty—not just risk with known probability distributions. And yet, as scientists we wish to obtain computable answers to questions such as the three listed above.

[22]Interestingly, the structure of a terrorist organization is also that of a cellular network, as we shall see later on. What does Parnas' Principle suggest in the context of terrorist organizations, terrorism in general, or counterterrorism policy analysis? Which of those insights derived from a CSS approach are also available from traditional social science perspectives?

Computation is feasible over an immense and expanding problem-space, but it is not universal. **Computability** has to do with the effective ability to compute an algorithm, given some functions/methods/operations and data. More precisely, effective computability requires two conditions:

1. The algorithm must consist of a finite and relatively simple set of functions arranged in some proper way; and
2. Each function must execute in finite time.

Given these two requirements, a problem is not computable if either condition is not met.

Informally, **computational complexity** refers to the degree of difficulty involved in solving a computational problem of size N, in terms of space or time resources required. Formally, let $T(n)$ and $M(n)$ denote separate measures of computational complexity with respect to time and memory, respectively, where $n \in N$ denotes the size of the problem. For example, N may refer to the number of possible alliances in Problem 1, the number of alternatives in Problem 2, or similar features that measure size. In general, computational complexity has to do with how computability scales with respect to a given size. A problem that scales as a **polynomial** is said to be computationally tractable, whereas one that scales **exponentially** is not. A problem is said to be **intractable** when it cannot be solved in polynomial time.

2.12 Algorithms

So far we have used the term **algorithm** more or less as synonymous with "code" or "program." Stated more precisely, a program is a *formalization* of an algorithm, similar to the way in which an equation specifies a function. According to the *Dictionary of Algorithms and Data Structures* published by the National Institute of Standards and Technology (NIST), an algorithm is defined as follows:

Definition 2.1 (Algorithm; Black 2007) An algorithm is a computable set of steps to achieve a desired result.

In this chapter we have already seen several initial examples of algorithms, ranging from chaos to elections. We should now be able to have a better appreciation of how the concept of algorithm relates to the *Computational Paradigm* of CSS discussed earlier in Chap. 1. Such a perspective views a social system (on any scale) as an information-processing entity; i.e., as algorithmically structured. How is this possible? The information processed by social systems is structured in many ways, as discussed in Sect. 2.9. Information can be in the form of records, arrays, trees, or other data structures. Algorithms involve search, comparisons, maximization, sorting, and other fundamental and compound forms of processing information.

Algorithms are implemented in social systems using many different real-world processes. The following are some examples of significant social processes viewed in terms of "desired results" and "sets of computational steps," consistent with Definition 2.1:

Cognitive balancing (Psychology): As humans, we maintain overall cognitive coherence in our belief systems *by* adjusting beliefs through Abelsonian mechanisms (discussed in Chap. 4).

Census (Sociology): Every complex society (chiefdoms, states, empires) counts the size of its population *by* conducting surveys and other procedures for gathering data on individuals and households.

Economic transaction (Economics): Economic agents conduct a sale *by* exchanging information and agreeing on terms.

Election (Politics): A democratic polity determines a leader *by* counting votes according to some set of rules.

Legislate (Politics): Policymakers enact laws *by* aggregating preferences following constitutionally established procedures.

CSS requires us to examine social processes from an algorithmic perspective and social systems as supported by functionally significant algorithms, following the Computational Paradigm. Obviously, each of these complex processes has far more real-world complexity than can be reasonably stated in a single sentence. However, the fact that each descriptive sentence has the same algorithmic form as in Definition 2.1 is interesting and insightful. Formally, this kind of similarity is called an **isomorphism.**[23] The Computational Paradigm discussed earlier in Chap. 1 is about a general isomorphic perspective, whereby social systems are designed as adaptations (Simon's Principle) to perform complex algorithms of many kinds.

Algorithms matter greatly in CSS because through improved design of algorithms we can develop better models of social complexity and—in doing so—advance our understanding of human and social dynamics. Learning how to design and implement efficient algorithms requires both technical skill and experience through practice. Key steps involve understanding **search**, **sort**, and **recursive** algorithmic structures. For example, there are significant differences in the efficiency of various search routines (e.g., linear vs. binary) depending on input size and other considerations. **Binary search**—an example of what are called **divide-and-conquer algorithms**—is often desirable as an algorithm because it only requires time in logarithmic (i.e., less than linear) proportion to the size of a list. By contrast, **linear search** is much more time consuming (hence less computationally efficient) for relatively long lists, but is usually better for searching items in short lists. The exact tradeoff between the two strategies depends on data structures, code used, and hardware, but, in general, linear and binary search strategies are best for short and long lists, respectively.

[23]The term isomorphism comes from mathematics, where it means having the same formalism or equation in different domains. For example, a cannonball shot (physics) and a parabolic demand function (economics) are said to be isomorphic since both are described by a second degree polynomial, $y(x) = a + bx + cx^2$. Similarly, social transactions between two populations (human geography) and gravitational attraction between two masses (physics) follow an isomorphic inverse-square law, $y = kS_1 S_2 / D^2$, where S and D denote sizes (for populations and masses) and distance between them, respectively. Two systems are said to be isomorphic if the relevant equations obey the same mathematical form.

Unfortunately, a binary search usually requires a pre-sorted list, which can be a problem for sorting. **Recursive functions** for **sorting** come to the rescue! Different sorting algorithms include **select sort** and **merge sort**. Select sorting requires time that is proportional to the square of collection size (cardinality). By contrast, merge sorting is a divide-and-conquer algorithm that sorts in $n \log n$ time.

Recommended Readings

H. Abelson, G.J.Sussman, J. Sussman, *Structure and Interpretation of Computer Programs*, 2nd edn. (MIT Press/McGraw-Hill, Cambridge/London, 1996)

S.W. Ambler, *The Elements of UML 2.0 Style* (Cambridge University Press, Cambridge, 2005)

I. Asimov, *Foundation Trilogy: Foundation (1951), Foundation and Empire (1952), and Second Foundation (1953)* (Gnome Press, New York, 1953)

J. Barker, An innovative single-semester approach to teaching object modeling and java Programming, in *3rd International Conference on the Principles and Practice of Programming in Java, PPPJ 2004*, (2004)

J. Barker, *Begining Java Objects: From Concepts to Code*, 2nd edn. (Apress, Berkeley, 2005)

P.E. Black, Data structure, in *Dictionary of Algorithms and Data Structures*, ed. by P.E. Black (US National Institute of Standards and Technology, Washington, 2007)

C. Cioffi-Revilla, Simplicity and reality in computational modeling of politics. Comput. Math. Organ. Theory **15**(1), 26–46 (2008)

A.C. Clarke, *2001: A Space Odyssey* (Penguin, New York, 1968)

H.-E. Eriksson, M. Penker, B. Lyons, D. Fado, *UML 2 Toolkit* (Wiley, New York, 2004)

M. Felleisen, R.B. Findler, M. Flatt, S. Krishnamurthi, *How to Design Programs: An Introduction to Programming and Computing* (MIT Press, Cambridge, 2001)

M. Flynn, *In the Country of the Blind* (Tor Science Fiction, New York, 2003)

E. Gamma, R. Helm, R. Johnson, J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software* (Addison-Wesley, Reading, 2012)

R.L. Graham, E. Knuth Donald, O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd edn. (Addison-Wesley, Reading, 1994)

E. Grimson, J. Guttag, *MIT 6.00 Introduction to Computer Science and Programming*. Online course: http://videolectures.net/mit600f08_intro_computer_science_programming/

J.V. Guttag, *Introduction to Computation and Programming Using Python* (MIT Press, Cambridge, 2013)

Y.-T. Lau, *The Art of Objects: Object-Oriented Design and Architecture* (Addison-Wesley, Boston, 2001)

J.D. Murray, After Turing: mathematical modelling in the biomedical and social sciences, in *How the World Computes*, ed. by B.S. Cooper, A. Dawar, B. Löwe (Springer, Berlin, 2012), pp. 517–527

D.L. Parnas, On the criteria to be used in decomposing systems into modules. Commun. ACM **15**(12), 1053–1058 (1972)

E. Regis, *Who Got Einstein's Office?* (Basic Books, New York, 1988)

H.A. Simon, *The Sciences of the Artificial*, 3rd edn. (MIT Press, Cambridge, 1996)

M. Weisfeld, *The Object-Oriented Thought Process*, 2nd edn. (Developer's Library, Indianapolis, 2004)

J. Zelle, *Python Programming: An Introduction to Computer Science*, 2nd edn. (Franklin Beedle & Associates, Sherwood, 2010)

Automated Information Extraction

3

3.1 Introduction and Motivation

In the previous chapter we defined an algorithm as a computable set of steps to achieve results. The goal of this chapter is to introduce algorithms used for extracting information from data, what social scientists have traditionally called *content analysis*. The idea is to leverage computing in such a way as to minimize human, manual handling of data. Why? For multiple reasons:

- Information extraction by humans (called "coders" in this context) is very labor-intensive, requiring long periods of training and preparation.
- Even when well-trained, coders make mistakes that are difficult to correct.
- The universe of data sources has recently expanded beyond what is feasible to analyze by human coders, including many Internet sources.
- Algorithms specialized in information extraction can detect patterns that humans are not well equipped to handle, such as network structures and time-dependent features, or latent properties.

Traditionally, text data was the main target of content analysis, but increasingly these methods are also aimed at graphics, imagery, video, and audio data signals. Decades ago this was all done manually, by training coders and using manual operations that produced coded data after many months of training. The Age of Big Data has begun, with several quintillion bytes of data produced each day (1 quintillion $= 10^{18}$ on the US short scale $= 10^{30}$ on the EU long scale). Today the goal is to extract information from data (whether "small" or "big") using automated algorithms and systems. This expands scientific analysis to better comprehend the increasing volume of social data (so-called Big Data), the greater diversity of data signals, the increased accuracy, and new and exciting frontiers of cross-cultural research, among others.

3.2 History and First Pioneers

Social scientists have always been interested in the meaning of signs and other linguistic and non-linguistic (e.g., behavioral) symbols used in social interaction. The Greeks were arguably the first to ponder the meaning of signs through the study of

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_3,

EVENT DATA CODING RECORD

Date: 19621022 Actor: US Target: USSR Scale: 13

Description: President JFK announces naval quarantine on Cuba responding to Soviet missiles.

Source: New York Times, October 23, 1962.

Coder: SL

Fig. 3.1 Example of a manual coding form used to record an event based on a newspaper source. Forms such as these were used in the early days of computational content analysis to record news into machine-readable format and enable statistical analysis of large amounts of data

etymology (the study of the roots of words) and related disciplines, as shown by surviving records.[1]

Figure 3.1 shows a so-called "coding sheet" for producing a simple event data set developed from newspaper sources. Coding sheets such as these—and more elaborate ones—were common to many social science data projects based on content analysis.

Automated information extraction, under the initial name of quantitative content analysis, was invented in the 1960s, when, for the first time, digital computers made it possible to use computer algorithms to replace manual coding. However, these methods have a long history! The following are significant historical milestones and pioneers in this area of Computational Social Science:

18th century First well-documented quantitative analyses of text in Sweden (Dovring 1954; cited in Krippendorf 2013: 18).

1893 G.J. Speed publishes the "first quantitative newspaper analysis" (Krippendorf 2013: 53), to be followed by modern events data analysis many decades later (beginning in the 1960s).

1903 Eugen Löbl publishes "an elaborate classification scheme for analyzing the 'inner sources of content' according to the social functions that newspapers perform" (Krippendorf 2013: 11).

1910 Sociologist Max Weber proposes the first large-scale content analysis (Krippendorf 2013: 11).

1912 Tenney proposes the first "large-scale and continuous survey of press content" to monitor "social weather" (Krippendorf 2013: 12).

1913 Mathematician and linguist Andrey Markov (after whom the Markov "chain model" is named) publishes his statistical analysis of Pushkin's *Eugene Onegin* (Markov 1913).

[1]Much of modern science is said to have roots in the ancient Greeks. This is quite true, but others before them may have contributed earlier scientific ideas contained in media that have been lost (manuscripts, inscriptions) due to the destruction of many large ancient libraries, such as those of Alexandria, Antioch, Baghdad, Córdoba, and Damascus, just to mention some of those in the Mediterranean world. India and China also experienced the destruction of many libraries during their early history.

Fig. 3.2 Major pioneers of content analysis: Max Weber, sociologist, proposed the first large-scale content analysis in 1910 (*upper left*). Andrey Markov, mathematician, pioneered computational linguistics (*upper right*). Harold Lasswell pioneered computational content analysis (*lower left*). Charles E. Osgood discovered and quantified semantic space (*lower right*)

1934 Woodward publishes his influential "Quantitative Newspaper Analysis as a Technique of Opinion Research" (Woodward 1934).

1937 The Institute for Propaganda Analysis, founded in New York City by social scientists to counter Nazi propaganda, publishes a list of devices commonly used by extremists and propagandists.

1938 Albig publishes the first content analysis of radio media, followed by movies and television (Albig 1938).

1941 The term "content analysis" is used for the first time (Waples and Berelson 1941: 2; cited in Berelson and Lazarsfeld 1948).

1942 Psychologists Allport and Baldwin separately publish the first applications of content analysis on personality and cognitive structure, respectively (Allport 1942; Baldwin 1942).
1947 Psychologist R. K. White pioneers the application of content analysis on values (White 1947).
1948 Berelson and Harold D. Lasswell publish their pioneering and influential mimeographed text, *The Analysis of Communication Content*, published in 1952 as Berelson's *Content Analysis in Communications Research*. "This first systematic presentation codified the field for years to come" (Krippendorf 2004: 8).
1949 Claude Shannon and Warren Weaver publish their *Mathematical Theory of Communication*, formalizing the concepts of signal, message, channel, and noise (Shannon and Weaver 1949).
1949 Lasswell publishes his methodological essay on "Why Be Quantitative?" (Lasswell 1949).
1950 Sociologist Bales pioneers the application of content analysis in small-group research (Bales 1950).
1952 Berelson publishes the first integrated survey of content analysis, spreading across the social sciences (Berelson 1952).
1955 First major conference on content analysis, is sponsored by the Social Science Research Council's (SSRC) Committee on Linguistics and Psychology (de Sola Pool 1959).
1957 Charles E. Osgood [1916–1991] and collaborators publish the first semantic differential scales derived through computer-based factor analysis (Osgood et al. 1957).
1959 Osgood's contingency analysis and "cloze procedure" (Osgood 1959).
1962 Philip J. Stone [1937–2006] and collaborators publish the first paper on the *General Inquirer* in the journal *Behavioral Science* (Stone et al. 1962).
1964 Political scientist Kenneth Janda (Janda 1964; Janda and Tetzlaff 1966) creates the TRIAL (Technique for Retrieval of Information and Abstracts of Literature) system for text processing and mining of scientific literature, including use of KWIC (keywords in context) and KWOC (keywords out of context) indexing.
1972 Ward Goodenough applies content analysis in anthropology in his seminal book *Culture, Language and Society*.
1975 Charles E. Osgood and collaborators publish *Cross-Cultural Universals of Affective Meaning*, the first large comparative analysis of semantic differentials produced by computational content analysis (Osgood et al. 1975).
1997 The journal *Social Science Computer Review* publishes a special issue on "Possibilities in Computer Content Analysis" (Fan 1997).
2004 Klaus Krippendorf publishes the first edition of his classic textbook, *Content Analysis*.
2013 Kalev Leetaru, Philip Schrodt, and Patrick Brandt release the first version of GDELT (Global Data on Events, Location, and Tone), the first computer-coded big-data collection on world events, containing over 200 million ge-

olocated events from 1979 to the present. By summer 2013 GDELT was generating over 120,000 machine-coded events per day, roughly three orders of magnitude more than a team of humans could code manually, following months of intense training.

This is quite a history of scientific accomplishments that continues to expand the frontiers of social research. Today, computer-based or automated content analysis is taught in many social science departments, summer institutes, and special workshops, as well as in business schools (public relations and marketing), computer science departments (text and data mining), and communications and linguistics programs. To begin exploring these powerful methods we will also need some basic relevant background in related areas, such as linguistics, communications, and social psychology.

3.3 Linguistics and Principles of Content Analysis: Semantics and Syntax

Linguistics is the science of human language. Linguists distinguish among the following key concepts pertaining to major language components (along with many others that lie beyond this introductory survey):

Grammar: The study of rules of natural human language, which determine how a given language is spoken. There are as many grammars as there are natural human languages, a number that has been decreasing to approximately 7,000 languages that exist today, of which approximately 500 are considered nearly extinct (Lewis 2009).

Syntax: Part of grammar, which refers to how phrases and sentences are to be properly composed. Rules of syntax determine how words are arranged to convey meaning.

Semantics: The meaning of terms or words. From a concept formation perspective, semantics refers to the *definiens* of a term, while the term itself is called the *definendum*, as in a glossary. In communication theory, the term "message" denotes the meaning of a given "signal." Thus, a message (analogous to *definiens*) is said to be encoded into a signal (*definendum*) in order for it to be transmitted or conveyed, according to the Shannon and Weaver (1949) theory of communication. Formally, message:signal::definiens:definendum.

These basic ideas are significant for automated information extraction because, after all, we are dealing with how information is obtained from basic raw data in the form of text or other media (for example, graphics). **Parsing** is the process whereby a sentence of text is analyzed into syntactical components, such as object, subject, and verb. Counting the frequency of words and syntactical components is a basic procedure in automated content analysis.

Consider the following example. A simple algorithm for counting and visualizing word frequencies is Wordle^{TM}. Figure 3.3 illustrates the results of analyzing Herbert Simon's (1992) autobiography using Wordle. Each word in the autobiography is shown by size proportional to word frequency, with only "stop words" omitted

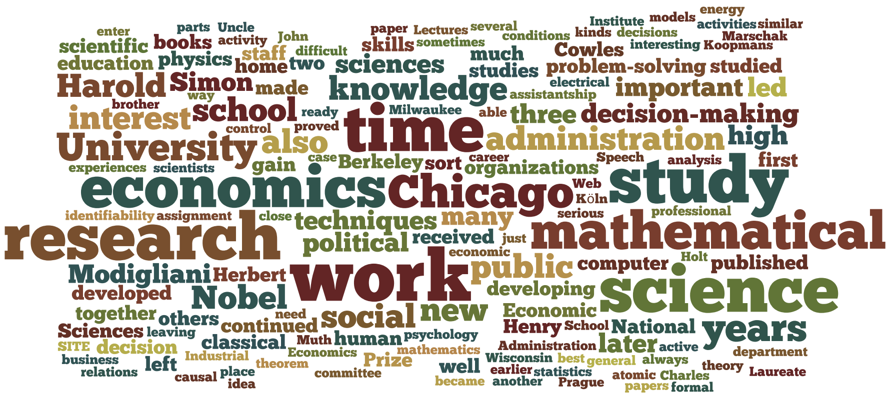


Fig. 3.3 Word frequencies automatically extracted from Herbert A. Simon's autobiography using the Wordle™ algorithm. *Source*: Simon (1992)

("a," "the," "of," and other such frequently used words that don't add information to the results). Numbers are also stopped by this particular algorithm, but could be included when they are of interest.

In the previous example all words were counted separately. This is fine for many words, but not all. For instance, from a semantic perspective, the phrase "University of Chicago" makes more sense as a compound term than as three separate words. To do this, the algorithm should process a prepared version of the raw text data that links such words together, using either a tilde character (~) or the Unicode "non-breaking space" character (U+00A0) inserted between those words that should remain linked. How would this refinement change our results? This is left as an exercise!

"Context matters," as the saying goes. Word frequencies alone, out of context, are said to provide a KWOC (or "keywords out of context") result. By contrast, a KWIC (keywords in context) analysis shows the neighboring words of each occurrence.

In computer programming, **profiling** code is a procedure for analyzing software performance by counting the frequency with which each method is called, the time required for each method, and other frequency-related features of code. Profiling is carried out through various systems called profilers, which can be passive or active. Profilers rely on program counters inside a CPU and provide a form of automated content analysis for better understanding how code works.

3.4 Semantic Dimensions of Meaning: From Osgood to Heise

Earlier we saw how semantics refers to the meaning of words. There is also a common and unfortunate misconception that semantic diversity (ambiguity) impedes or even prohibits the development of social science, due to lack of agreement on the meaning of many terms used across the social science disciplines. As shown in this section, however, social scientists have made great strides in gaining deep understanding of the meaning of words and signs, in no small way through the systematic application of computational approaches.

How do people assign meaning to words in natural language? What dimensions of meaning do we use for understanding what words mean? What do we care mostly about in terms of assigning meaning? Do we care about the source? The time of occurrence? Its location? Which attributes of a word matter most to us? These have been deeply significant, longstanding, and highly challenging questions, not just for linguists but for many other social scientists, such as anthropologists, sociologists, political scientists, and psychologists. Today this research is conducted with automated extraction algorithms. But before we address these questions we must understand what algorithms look for in terms of the structure of human information processing.

3.4.1 EPA-Space and the Structure of Human Information Processing and Meaning

A pioneer in investigating how humans think and communicate was psychologist Charles E. Osgood, who together with his colleagues made one of the most remarkable scientific discoveries of the 20th century, concerning how humans subjectively perceive the meaning of words and signs. Osgood and his collaborators at the Institute for Communication Research (ICR) at the University of Illinois at Urbana-Champaign discovered that all words used in natural language are decomposed by the human cognitive process into mostly three dimensions that he called **Evaluation**, **Potency**, and **Activity**. This 3-dimensional space or **EPA-space**, for short, consists of three continuous ranges with the following **affective values** (see Fig. 3.4):

1. Good-Bad (evaluation)
2. Strong-Weak (potency)
3. Fast-Slow (activity)

These three dimensions are the first three *orthogonal factors* extracted from a large corpus of words using standard data reduction procedures from factor analysis (Osgood et al. 1957, 1975). Roughly speaking, this means that, for each input signal (word, event, term) we perceive, we as individuals first assign a value in terms of whether the concept denoted by the input is "good" or "bad" in a normative, affective sense. This is the evaluation dimension. We then assess its potency in terms of the word or object being "strong" or "weak," as an impression. Finally, we assess the word in terms of being "fast" (dynamic) or "slow" (static), which somehow refers to its motion. This semantic space was unknown prior to Osgood's discovery—indeed, it seems remarkable that such a space exists at all, since there is nothing intrinsically necessary about its existence. The null hypothesis would be that we assign meaning in completely personal, subjective ways that are incomparable across individuals, but this is not the case, as Osgood and his collaborators discovered. We may *think* we assign meanings in highly personal ways, but—as social science in this area has demonstrated—in fact we use the same inter-personal or inter-subjective system of meaningful dimensions: Osgood's EPA-space.

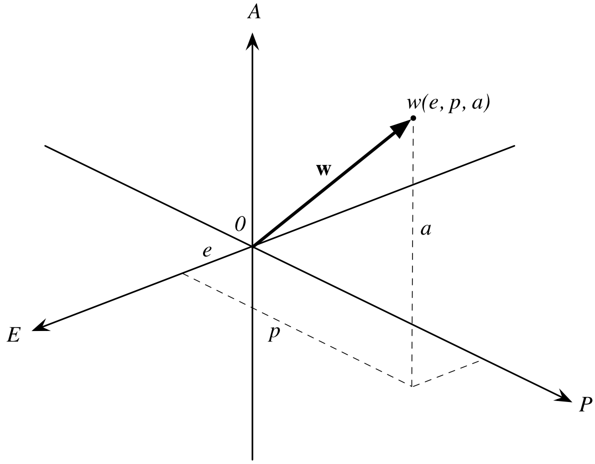


Fig. 3.4 *Osgood'd 3D Semantic Differential EPA-space.* The cognitive dimensions of evaluation E (ranging from good to bad), potency P (strong to weak), and activity A (fast to slow) span a 3-dimensional semantic space. In Osgood-space each term or word w is located by a triplet (e, p, a) or vector $\mathbf{w} = e\mathbf{i} + p\mathbf{j} + a\mathbf{k}$ with norm given by $|\mathbf{w}| = \sqrt{e^2 + p^2 + a^2}$

Why these three dimensions exist, as opposed to another system, remains quite a mystery—an unsolved scientific puzzle. Regardless, Osgood's semantic space gives us an exceptional and intriguing glimpse into how the human mind operates.[2] As it turns out, these particular three semantic dimensions of EPA-space also provide robust cognitive foundations for explaining and understanding patterns of social behavior, which is the subject of **Affect Control Theory** (Heise 1987). The core principle of affect control theory is that individuals maintain relatively stable affective impressions of others and situations, which regulates their behavior accordingly.

For example, the word "missile" would be bad, strong, and fast, whereas the word "house" would be closer to good, strong, and slow. EPA-space dictionaries now exist for many words in many languages (Osgood et al. 1975; Heise 2001). Based on this system of Cartesian coordinates, every word can be represented as a triplet of coordinates $w(e, p, a)$ in 3-dimensional EPA-space.

A significant consequence of the discovery of EPA-space is that for the first time in the history of social science it enabled measuring **semantic distance** between any pair of words, terms, or objects, assuming component coordinates (e, p, a) are known. In turn, these discoveries open the way to computational vector analysis and other directional multivariate techniques, as suggested by the system in Fig. 3.4.

3.4.2 Cross-Cultural Universality of Meaning

Do people in different cultures think differently? In many ways they do, using different metaphors and schema. But what about in terms of the semantic EPA-space used to assign meaning? Does the structure of semantic space vary cross-culturally?

[2]By contrast, John von Neumann's (1958) computer model of the human brain-mind phenomenon turned out to be wrong. Unlike von Neumann's, the EPA-space model of the human mind is empirically validated, even if it still lacks deep theoretical explanation.

Even within the same culture or language, could gender, age, or education (SES, or socioeconomic status) make a difference? It turns out that, for the most part, answers to these and similar questions are generally "no," as social scientists have been finding out in recent decades.

Much more has been investigated about 3-dimensional semantic EPA-space in the years since Osgood's seminal discovery in the 1950s. The most important exciting discovery arguably has been the cross-cultural validity of this remarkable structure about how we as humans think: the **cross-cultural universality of meaning** (Osgood et al. 1975; Heise 2001). EPA-space is a universal structure not only for words in the English language; it is universal across many other languages and cultures, including Spanish, Malay, Serbo-Croatian, Turkish, Chinese, Italian, Hebrew, Arabic, Thai, Farsi, German, French, and Japanese, among others. Gender accounts for some differences, but these are quantitatively known and measurable through the same basic methods employed by Osgood and his collaborators.

Project Magellan, based at Indiana University, is an international scientific research project aimed at automated information extraction of cross-cultural EPA ratings and related information. It employs an online Java applet system called *Surveyor*, which collects EPA ratings via the World Wide Web according to the following process (Heise 2001):[3]

> Respondents with a computer connection to the Internet go to a WWW page that fetches the Java applet and its associated stimuli files. The applet presents stimuli, and the respondent rates the stimuli with the computer's mouse, by dragging a pointer along bipolar adjective scales. The applet records the respondent's ratings in numerical form and sends the data to a central computer for storage when the respondent finishes the ratings. The Surveyor measuring instrument can be revised to work in any indigenous language. [...] At the end of each session the respondent's data are transmitted electronically via the Internet to the USA In the USA the data automatically are assembled into cleanly coded data sets. Authorized researchers, including researchers in the country of the data's origin, can download the data from the USA at any time via the Internet. [...] Ratings are recorded as decimal numbers with 430 increments from one end of the scale to the other, rather than the seven increments of early semantic differential scales, or the 80 increments of the Attitude program.

What are the main implications of these discoveries in automated information extraction and human semantic space for CSS? How do they fit within the broader field of CSS knowledge and research? There are many important CSS implications of the Osgood semantic space spanned by EPA dimensions. Automated information extraction must be informed by the nature and structure of human cognition in terms of EPA-space, regardless of source data, but especially in the case of analyzing text

[3]The predecessor of Surveyor was called *Attitude*, which was also developed by David Heise (1982) as the first computer-based extractor of EPA ratings, replacing the old paper-based forms used since Charles E. Osgood and his collaborators.

corpora. In practice, this means that CSS researchers need not "start from scratch" or invent semantic spaces based just on naive speculation, as if this were unexplored territory in social science. Rather, CSS researchers should know what has been discovered thus far—the corpus of knowledge in positive social science—and build on earlier foundations to develop the field. In the next section we will examine how CSS researchers "mine data" to extract information. EPA-space provides a natural framework for mapping such information, given what we now know about the structure of human cognition and information processing.

3.5 Data Mining: Overview

The process of automated information extraction using as input a variety of complex or unstructured data sources—a typical situation in social science—for the purpose of extracting information or patterns of various kinds is called **data mining** in computer science.[4] Extraction may pertain to monitoring, discovering, modeling, comparing, or replicating patterns in the data. Text, social media, audio, and imagery represent broad classes of data that can be mined to extract information. In a true computational sense, the pioneering work of Osgood and his successors involved data mining for the purpose of discovering the structure of human cognition and our natural semantic space used for computing overall meaning. Other instances of data mining, beyond exploration of the human semantic EPA-space, take as input many other classes of data and employ algorithms based on other data processing procedures besides factor analysis.

Who mines data? Data mining has been practiced by quantitative and computational social scientists since the dawn of computing, and by computer scientists and software engineers since the early 1980s. It is a major and growing area of research across the social sciences (and humanities), with research projects ranging from anthropology (Fischer et al. 2013) to political science (Schrodt 2000), and from archaeology to history (Williford et al. 2012). In computer science the Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) of the Association for Computing Machinery (ACM) was established for this purpose in the 1980s, offering as resources an annual international conference and proceedings, as well as a biannual academic journal entitled *SIGKDD Explorations*. There are numerous CS conferences on data mining, including the ACM Conference on Information and Knowledge Management (CIKM), the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), the IEEE International Conference on Data Mining, and the SIAM International Conference on Data Mining, among others.

[4]Unfortunately, in social science the term "data mining" has quite a negative connotation, since it is understood as lacking in theoretical understanding and symptomatic of so-called "barefoot empiricism," akin to "a fishing expedition." CSS assigns high priority to theory—the basis of understanding—while recognizing the scientific value of inductive data mining.

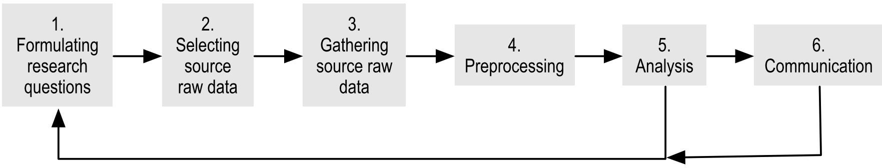


Fig. 3.5 *General Data Mining Methodological Process.* Data mining for automated information extraction involves several stages, the most important being the six highlighted here and discussed below. The core is Analysis for answering research questions, but the other five stages are just as critical for overall quality of the scientific investigation. Each of the six stages involves a variety of procedures, most of them dependent on the research questions being addressed

Data mining is a methodological process used for a variety of research purposes and in numerous domains of CSS, as we will examine in greater detail in the next section. At the core of data mining lie two fundamental analytical approaches that play major roles. Let's highlight them here in advance of a more in-depth discussion in Sect. 3.6.4:

- **Categorization**: Also known as **classification**, this type of analysis in data mining aims at producing an output set of categorized information using some degree of human intervention in the analysis; hence, categorization is a form of so-called **supervised machine learning**, computationally speaking.
- **Clustering**: By contrast, clustering is a type of data mining analysis that is far more inductive and is a form of **unsupervised machine learning**.

Both types of analysis can be considered part of **similarity analysis** within the general process of data mining, as detailed in the next section.

In computer science "data mining" also includes other algorithms for extracting associations, correlations, multivariate regression models, and other empirical data structures that are quite common in quantitative social science research. However, from the perspective of CSS those techniques would fall more commonly under traditional statistical procedures provided by software systems such as SPSS, SAS, Stata, or R—in order of increasing computational power.

3.6 Data Mining: Methodological Process

Data mining is a rapidly developing field of interdisciplinary research that has expanded from text-based documents in the initial years to social media, imagery, audio/sound, and other media in recent years (Feldman and Sanger 2007; Hsu et al. 2008; Leetaru 2011; Monroe and Schrodt 2008; Tang and Liu 2010; Hermann and Ritter 1999; Hermann et al. 2011). Regardless of the data being mined, as with most major areas of CSS, data mining for automated information extraction is a *methodological process* composed of a *sequence of stages or phases*—it is *not* a single, uniform activity or even a set of activities that can be carried out in arbitrary order. As always in science, the process of data mining (see Fig. 3.5) begins with the *formulation of research questions* and ends with *communication of results.* In between are other major, critical stages, such as those pertaining to *source raw data inputs*, *preprocessing*, and—finally—*analysis* proper, the latter being impossible without previous stages. The overall process cycles back to the first stage involving research questions, because analytical progress and communicating results often generate new research questions—as fertile scientific projects should do! **Spiraling** is another useful metaphor for understanding the general data mining process, because a project often begins with an intentionally limited corpus of data—perhaps just a sample—to test the overall procedure up to some basic analysis, after which the initial test data is gradually, incrementally, scaled up to its full final size (e.g., a whole data archive consisting of corpora of data) as determined by the research questions and data availability.

Remember: the actual media of data in a given research project can be of many different kinds, such as text, numeric, social media, geospatial, imagery, audiovisual, or other. The same general process will apply, as detailed below.

3.6.1 Research Questions

In CSS—as everywhere in science—everything begins with research questions, as we already discussed in Chap. 1. A very broad range of research questions has become feasible through data mining—and the range seems to be forever expanding, as whole new classes of questions are enabled by new theory, new data, or new methods. At one end are projects defined mostly by *data-driven* or *inductive research questions* of an exploratory and discovery nature. In this highly empirical mode of investigation the CSS researcher intentionally seeks to extract information in ways that are unbiased by previous theories, biases, or preconceptions. A classical (even dramatic!) early example of this would be Allen Newell and Herbert A. Simon's inductive rediscovery of Kepler's Third Law—also known as the Law of Harmonies—using Pat Langley's BACON.3 computer program (Langley 1981, 2004; Simon 1996; Gorman 1992). BACON found Kepler's law in three algorithmic steps, given exactly the same data used by Kepler (gathered by the 16th century Danish astronomer Tycho Brahe). In BACON's case Newell and Simon asked the research question: what is the relationship between distances of planets from the sun R and their periods of revolution T? The answer is the constant ratio T^2/R^3. It took Kepler ten years to discover the law of harmonies; BACON took seconds, although it took Simon and Newell several years to invent BACON. Another example of data-driven research was Charles E. Osgood's discovery of EPA-space using factor analysis, where the research question was: are there significant dimensions to human affective perception (semantic dimensions for the meaning of word phrases) and, if so, what are they? The answer is yes and the dimensions are three: evaluation E (good-bad), potency P (strong-weak), and activity A (fast-slow). Other dimensions don't matter or matter far less than these three. Note that in both cases answers were provided by data-driven algorithms without resort to prior theories or other domain-specific knowledge, just using a raw data input and algorithms that lacked theoretical direction.

At the opposite end of the spectrum are *theory-driven*, *deductive research questions* aimed at testing specific hypotheses and similar investigations in the more classical hypothetico-deductive mode. Many uses of data mining fit this pattern as well. An example would be Osgood's subsequent ground-breaking comparative research, where he and his collaborators sought to test the EPA-space hypothesis to confirm its cross-national validity. In this case the research questions were informed by theory and prior knowledge on the dimensionality of human semantics using factor analysis. This type of research is also known as **confirmatory factor analysis**, since it is based on some prior theory, model, or hypothesis about the dimensionality structure of the data space being investigated, as opposed to being mostly data-driven. A further example from the same domain of CSS would be David Heise's research program using Osgood's EPA-space to conduct comparative research across human languages and cultures (Project Magellan; Heise 2001).

In between the above two poles are numerous blends of data- vs. theory-driven research questions that provide great flexibility between inductive and deductive ends of the continuum. Typically, a research project may cover a range of questions, some of which are more inductive or deductive than others. Independent of orientation, the formulation of research questions should frame every well-designed data mining investigation because *research questions condition each of the subsequent stages of the process*.

3.6.2 Source Data: Selection and Procurement

The second stage in a data mining investigation focuses on the source data input itself, once research questions have been selected on the inductive-deductive continuum. Text, electronic media (including so-called social media), imagery, video, and sound are among the major classes of interest. Sensor data of many different kinds across diverse domains is also increasingly being collected and analyzed—recall the daily production of quintillions of bytes of data mentioned at the beginning of this chapter.

Data selection and procurement pose separate albeit related challenges. Research questions should guide and inform data selection. Today the Internet offers numerous sources of data—many of which can easily be found through search engines—in addition to long-standing data repositories such as those of the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan, US, and the European Consortium for Political Research (ECPR) at the University of Essex, UK. The Social Science Research Network (SSRN)—the world's largest open access repository—is an online archive containing references to numerous data sources across the social sciences. CSS research is increasingly interdisciplinary, based on the complex adaptive systems paradigm of coupled human, natural, and artificial systems, thereby requiring data sources from the physical and life sciences, engineering, and humanities. In each case, the primary principle for data selection regards the primacy of research questions in guiding or determining the choice of data. Issues regarding intellectual property rights, ethics, public vs. private funding, rights of human subjects, privacy, and similar issues are among the most prominent aspects encountered in terms of selection and procurement of source data.

3.6.3 Preprocessing Preparations

Once data has been selected and procured it almost always requires preprocessing preparation before it can be analyzed. Scanning, cleaning, filtering, initial content extraction (identifying the main body of interest), and similar preparations are among the most common preprocessing activities:

Scanning: Original texts may require OCR (optical character recognition) scanning to generate machine-readable files that can be analyzed.

Cleaning: Extracting headlines, bylines, dates, and similar information fields may also be necessary.

Filtering: Initial filtering may involve some form of preprocessing categorization, necessary for distinguishing among different actors or behaviors of interest, given the research questions. Filtering may also involve selecting elements above some selected thresholds (e.g., trade transactions above some monetary value; population centers above a given size; behaviors comprised within specific ranges).

Reformatting: A single data source, such as a whole document, often requires dividing into smaller individual component units to conduct both aggregate and desegregated analyses.

Content proxy extraction: Sometimes proxy elements in the source corpus can be used for subsequent focused analyses, as is the case for actors, locations, or events that denote or imply latent entities. An example would be certain terms (e.g., "axis of evil" in political texts or racial slurs that tag individuals).

3.6.4 Analysis

The core stage of data mining consists of one or more forms of analysis, given a properly prepared set of data. Again, analytical modes are always a function of research questions, whether the investigation is theory-driven or data-driven.

There are many kinds of analyses performed in data mining and their variety and power increase as a function of both formal methods and information technology. All of them have been in use by social scientists since the quantitative methodological revolution, but each analytical approach has undergone quantum improvements with recent computational developments. The following analytical methods are among the most widely used in CSS:

Vocabulary analysis: This is one of the most basic forms of algorithmic information extraction and aims at obtaining a catalog of words or other signs (symbols, numbers, icons, glyphs, among others) contained in the data source being analyzed. Focusing on signs irrespective of precise meaning (semantics) or grammar (syntax) is typical of vocabulary analysis, so this basic form of analysis takes a "bag of words" approach to data mining. Word counts are an example (Fig. 3.3, analyzing words in Simon's autobiography), as when analyzing text to assess a baseline, examining histograms, trends over time, or indices of readability; or testing hypotheses about their frequency distributions

Table 3.1 Measures of association depending on levels of measurement

Level of measurement	Nominal	Ordinal	Interval	Ratio
Nominal	Lambda λ	Kramer's V, ϕ (only for 2×2 tables)	Kramer's V	Kramer's V
Ordinal	Kramer's V	Gamma γ, Somer's D, Kendall's τ_b (only square tables) and τ_c (rectangular tables), Spearman's ρ	Pearson's r	Pearson's r
Interval	Kramer's V	Spearman's ρ	Pearson's r	Eta η
Ratio	Kramer's V	Spearman's ρ	Pearson's r	Pearson's r

(e.g., Zipf's Law, discussed later in Chap. 6). In turn, vocabulary analysis provides foundations for more advanced kinds of data mining analysis.[5]

Correlational analysis: A somewhat more complex form of analysis in data mining consists of looking for (data-driven) or testing (theory-driven) various kinds of associations between or among terms or signs. An association is always a mapping from one domain or set of terms to another. For example, data can be mined to establish associations between terms and any set of other features or items, such as locations, dates, contexts, or other aspects of source data. Formally, there are many kinds of associations ranging from simple concurrences or co-occurrences to more complex quantitative forms of correlational and causal relations (e.g., Granger causality). Measures of association are defined for all pairwise combinations of nominal, ordinal, interval, and ratio variables. It is important to pay close attention to this when choosing which measure to use, because the choice is not arbitrary, but most depend on the highest level of measurement supported by the data being analyzed. Table 3.1 shows proper choices and uses for measures such as Spearman's ρ, Pearson's R, and Kendall's τ, along with others commonly used.

Lexical analysis: The creation of additional lookup files, such as lexicons, thesauri, gazetteers (lexicons that associate geographic coordinates to locations), and other systematically defined auxiliary collections of entities is called lexical analysis. This form of analysis in data mining enables researchers to analyze source data files in ways that enhance the information potential of original data. Lexical analysis is used for a variety of purposes, including but not limited to **named entity recognition and extraction (NER)**, **categorization** (part of what is called similarity analysis, discussed below), **disambiguation**, and various mapping and cartographic applications. From a computational perspective, lexical analysis (including NER and other procedures) is a form of *semi-supervised learning*, where some *manual annotation of training data* is

[5] Besides its scientific value in CSS research, the popular media also uses basic forms of vocabulary analysis when counting the frequency of words used by politicians, such as in inaugural addresses or similar major speeches. The value of such anecdotal uses is rather limited, sometimes even misleading, since speechwriters and communication experts are well-versed in scientific principles of applied linguistics and human information processing, including sophisticated understanding of semantic differentials and other affect control, marketing, and propaganda devices.

still necessary, in spite of significant advances in recent decades. Another challenge is posed by differences among human languages, such as English, Spanish, Mandarin, or Arabic; each human language has its own NER challenges and mappings are still incomplete, inaccurate, or unreliable. The good news is that lexical analysis continues to improve in both effectiveness and efficiency. An important application of lexical analysis is with social, political, or economic event data, a field where machine coding has marked significant progress and is now considered equal to or more accurate than human manual coding. The **GDELT events data set** (Global Data on Events, Location and Tone; Leetaru and Schrodt 2013) was created thanks to a combination of data mining techniques that rely on lexicons or dictionaries for actors, gazetteers for locations, and other lexical analysis tools—as well as other components mentioned later—to enable computational events data analysis far beyond what was previously imaginable. While mining large data sets ($\gtrapprox$ 1 million events) is by itself a great improvement over what was feasible only a few years ago, the application of lexical analysis serves as a multiplier that greatly amplifies the range of qualitative and quantitive results by several orders of magnitude. The GDELT data set contains nearly a quarter-billion event records; updates are produced daily, 365 days a year, at a rate of more than 100,000 events per day, each record containing 58 fields of information machine-coded from scores of raw sources from many countries.

Spatial analysis: Besides being part of lexical analysis—through the role played by gazetteers—data mining techniques such as geocoding, geographic clustering, and similar geospatial techniques are used in spatial analysis. All of them can be related to earlier analyses in **quantitative human geography**. For example, spatial analysis applied to events data can be used to produce maps with various projections to examine distributions of phenomena such as social movements, migrations, disasters, and other patterns. The centroid of a spatial actor or location of an event or attribute is often used rather than its actual territorial shape. For example, the map in Fig. 3.6 illustrates the state of the world in terms of conflict and cooperation events on October 7, 2013, based on the previous 24 hours. GDELT is the most recent, global, largest, most comprehensive data set in CSS. It is also a project-in-progress. Every event data set produced by data mining must address many demanding scientific challenges such as continuous improvements in *selection of raw data sources* (newswire services), *event coding scales* (Goldstein's or other, including use of multiple scales), *categorization algorithms*, and *error propagation management*, among others.

Semantic analysis: While vocabulary and lexical analysis focus attention mainly on signals, semantic analysis focuses on meaning and actual content in terms of what various terms and entities stand for. Semantic analysis includes **machine parsing** the various **parts of speech** by means of **tagging** nouns, verbs, and other ontological components in source data. The results of semantic analysis typically consist of **noun phrases** and **verb phrases**. Semantic analysis complements lexical analysis in the construction of dictionaries such as

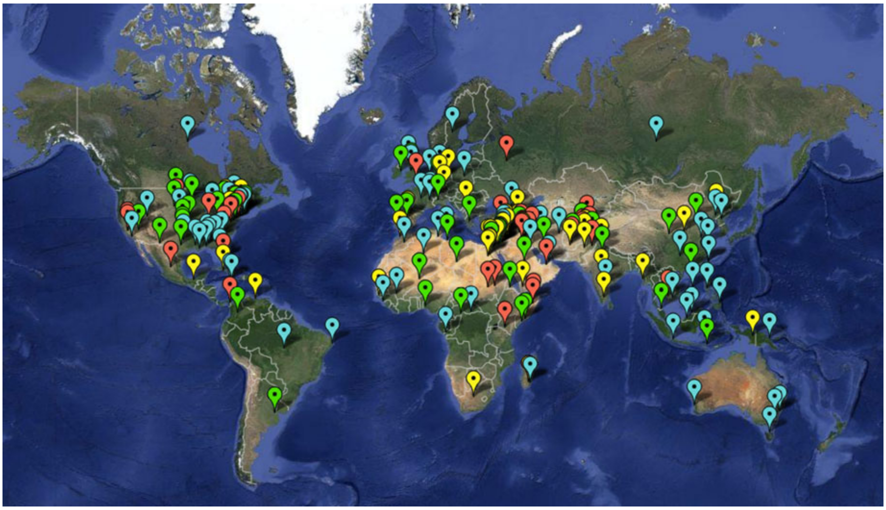


Fig. 3.6 *Spatial analysis using event data.* This Google map of the world shows the top 2,000 political events on October 7, 2013, based on the GDELT data set (Leetaru and Schrodt 2013). Color-coded events indicate degrees of conflict (*red* and *yellow*) or cooperation (*green* and *blue*). *Source*: GDELT website, downloaded October 8, 2013

CAMEO and the TABARI core extraction algorithm used in GDELT. Machine translation and other natural language processing (NLP) applications also play major roles in semantic analysis, such as entity and relationship recognition-extraction, fact and claim extraction, pronoun coreference resolution, and geographic disambiguation, among others.

Sentiment analysis: Emotional content is the main focus of sentiment analysis, a form of analysis based on Osgood's pioneering work demonstrating the primacy of the *evaluation dimension E*. Evaluative judgment (subjective assessment of good/bad) is also the basis for cognitive schema in human reasoning and belief systems (as shown in Chap. 4). Sentiment analysis is therefore a component of EPA-analysis (Azar and Lerner 1981), especially when combined with other dimensions, and is conducted at multiple levels of analysis, such as an entire document, sections of a document, or single objects/entities in the source data—all of which can be mapped onto E-space using appropriate lexicons.

Similarity analysis: Comparing and contrasting content is called similarity analysis in data mining and automated content analysis. We have already briefly mentioned two major forms of analysis that are part of similarity analysis—categorization and clustering—as (mostly) *supervised* and (mostly) *unsupervised* modes of machine learning, respectively. (This is a very rough pairing; in practice there is considerably more overlap.)

Categorization: This is a procedure that aims at classifying data based on a **training set** or data sample. A significant application of classification in CSS is for the purpose of **ontology extraction** (or **ontology generation**) from input data. This has several important applications, of which two in particular stand out: *events data analysis*, where actors and their behaviors matter greatly, and *agent-based modeling*, especially in early phases of model development such as design and implementation. In this section we examine the first, while reserving the latter for Chap. 10. A data mining algorithm that extracts this kind of information is called a **classifier**, which is also the technical name given to the actual mathematical function that implements the mapping onto the category space. Some of the simplest algorithms are the **naive Bayes classifier** and the **K-nearest neighbor classifier**. In CSS, categorization analysis was pioneered in political events data analysis by Philip Schrodt (1989) using a **Holland classifier** invented by computer scientist pioneer John Holland (1975, 1989). CAMEO (Conflict and Mediation Event Observations), the result of algorithmic entity extraction, is an example of a coding scheme for *actors* and *verbs* that describes their behaviors (Gerner et al. 2002; Schrodt et al. 2005). As demonstrated by the CAMEO-coded GDELT data set, categorization has become a major tool in events data research using online and archival data sources, now that manual human coding of newspapers and other printed sources has become mostly obsolete. Categorization is a major area of computer science and machine learning algorithms. Human supervision of categorization algorithms takes place in terms of selecting training data, establishing significant features for evaluation, selecting parameters such as thresholds, and other decisions.

Clustering: This is another type of similarity analysis for discovering low-dimensionality data structures or groupings of information, based on computational aggregation from high-dimensionality raw data. Osgood's discovery of EPA-space is an example of this use, where clusters are extracted by the factor analytic procedure. Another example of automated information extraction for clustering was the discovery of a similar 3-dimensional space spanned by national attributes such as the size S, level of economic development D, and military capability C of polities in the modern inter-state system, or **SDC-space**. This computational discovery confirmed Quincy Wright's (1942) earlier **Social Field Theory** on the existence of such a space. Note how in both cases clustering is used to uncover hidden or latent structures contained but not directly visible in the raw and "noisy" high-dimensionality data—in these cases researchers uncovered 3-dimensional Cartesian spaces that are easier to understand and visualize than the original high-dimensionality space spanned by the raw data. Clustering is considered a form of **unsupervised learning** in computer science. A common feature of clustering is the use of a large input archive of raw data from which clustering dimensions are extracted in several ways, such as optimal clustering, partitional clustering (decomposition into disjoint clusters), and hierarchical clustering (dendrograms). In addition to categorization and clustering, other important components of similarity analysis include distance and proximity measures (computed among data being compared), time warp plots (matching time-series input and target data), path distances (computed over time-warped input and target data), vector fields, difference maps, and similarity vectors and matrices. Various data mining software systems include algorithms that implement these components of similarity analysis.

Network analysis. Data mining methodology also plays an important role in the analysis of networks that arise in coupled human-natural-technological systems. Even within the confines of a purely human network, data mining can be used for extracting social communities (Tang and Liu 2010). As we saw in the Introduction chapter, network analysis is a major field of CSS, which we shall examine in the next chapter. A network consists of **nodes** and **links** (called arcs, edges, or vertices in graph theory, the branch of mathematics that studies networks). Data mining is used for extracting information pertinent to nodes and relations that constitute networks present in source data. For example, news media can be mined to automatically extract various kinds of societal network structures of interest, such as actors of various kinds (leaders, opinion-makers, supporters), roles (governmental, informal, occupational, among others), or locations, all of them linked by various kinds of social ties (Moon and Carley 2007). Network analysis enabled by data mining can also be spatial and temporal, which results in dynamic social networks that are spatially referenced.

Sequence analysis. Temporally indexed data, such as (but not limited to) time-series, lends itself to sequence analysis, a kind of data mining methodology for extracting information about the *states of a given process* and *dynamic transitions*, including *phase transitions* (Hsu et al. 2008). For example, financial data, political events data, opinion data, and others extracted through data mining algorithms can be analyzed for extracting temporal patterns. Among the most significant state-space representations of time series data are hidden Markov models (HMM), which are similar to classical Markov chains except that the state space consists of latent states, roughly similar to the idea of latent variables or invisible dimensions extracted by means of factor analysis. The states of an HMM are only approximately observable by proxies, since they cannot be directly observed. Markov models—whether classical or hidden—are similar to *UML state machine diagrams* in computing. If the main (most active) actors or entities are added to a sequence analysis, then the dynamic representation extracted from mined data may resemble a *UML sequence diagram.*

Intensity analysis: Source data can also be mined to extract intensities of observed or latent variables. For example, all kinds of *size variables* can be extracted from events data to produce *size distributions* and other quantitative features. In turn, these can be used as input for conducting subsequent analyses, as with information-theoretic measures or complexity-theoretic models—e.g., testing for power laws and other features of interest in complex systems. (We shall introduce these in a more complete way later, in Chap. 6.) From this perspective, sentiment analysis can be seen as a form of intensity analysis, except that it hardly ever goes beyond simple trends; instead, it could go much farther, to look for patterns or test hypotheses concerning generative dynamics. (Again, more on this is introduced in Chap. 6.)

Anomaly detection analysis: Some of the forms of data mining analysis seen thus far enable another form: data mining analysis for detecting anomalies or changes of some kind. In order to detect an anomaly it is first necessary to establish a base or "normal range," an idea pioneered in CSS by the late Lebanese-American political scientist Edward E. Azar (Ramsbotham 2005): the **normal relations range (NRR)** for a given series of events observed over time is defined as behavior within two standard deviations from prior average (arithmetic mean) behavior.[6] In addition, it must be assumed that the source data exhibits a significant degree of stability or persistence, in the sense that fundamental distribution moments (central tendency, dispersion, and others) do not undergo significant change during the test phase; otherwise it is difficult or impossible to detect an anomaly, unless it is many deviations away from the recent past. *Time scales also matter, because what may seem an anomaly on a short time scale may be quite normal on a longer scale*, which illustrates how anomaly detection analysis can be a very challenging procedure. Borrowing from linguistics, we can detect two forms of change: synchronic change and diachronic change. Both can be used to assess anomalies, but their dynamic context differs. Synchronic change refers to anomalies within a stationary or more or less structurally stable process or system. By contrast, diachronic change refers to much deeper anomalies being detected in the fundamental structure or generative dynamics of the process. An example of synchronic anomaly would be a change in the frequency of terms in a recurring speech pattern, as opposed to a diachronic anomaly caused by a deeper change in the actual vocabulary or grammar of the discourse. The same applies to events data analysis: some anomalies pertain to changes in the frequency of common events (synchronic anomalies), while other, much deeper, changes occur when the variety or vocabulary of events (what sociobiologists and ethologists call an ethogram) changes distribution.

Sonification analysis: We as humans have multiple senses, but most scientific analysis relies on vision. Data sonification analytics is the use of sound to learn new information or draw novel inferences on patterns in source data (Hermann et al. 2011), including Big Data. The basic idea of "sonifying" data is to listen to data features that may not be so apparent from traditional data analysis procedures. For example, the tone of multivariate time series rendered in sound (communicated by speakers) can produce harmonics that are difficult or impossible to detect in the source data. Data sonification is a form of "auditory display" (Kramer 1994) and for Big Data of interest to social scientists it is a

[6] The operationalization of the NRR in terms of two standard deviations from the process mean was suggested to political scientist and events data pioneer Edward E. Azar [1938–1991] by the mathematician Anatol Rapoport [1911–2007]. It was first applied to international relations events data series to study protracted conflicts in the Middle East. Azar was founder and director of the Conflict and Peace Data Bank (COPDAB), founded at the University of North Carolina at Chapel Hill in the 1970s and moved to the Centre for International Development and Conflict Management (CIDCM) of the University of Maryland at College Park in the 1980s.

new methodology that will likely find many applications—for example, using the recent GDELT data set to, quite literally, listen to the sound of global activity, as produced by $>10^5$ daily events worldwide. (The Smithsonian National Museum of Natural History, in Washington, DC, has an exhibit that sonifies earthquake data to communicate to the visitor seismic events around the Ring of Fire surrounding the Pacific Ocean.)

3.6.5 Communication

The final step in data mining focuses on communication of results, including implications of specific findings, broader implications for the field or research area, and perhaps also policy implications. These are very demanding communications requirements, each with its own challenges. The field of communication of information extracted from Big Data, including **visual analytics** (Thomas and Cook 2005), has become a vast area of scientific and technological research that has grown significantly in recent years—just as the Age of Big Data began to unfold. Some of the most influential concepts and principles have been contributed by political scientist Edward Tufte (www.edwardtufte.com/tufte/courses) and by the pioneering approaches developed at the US National Visualization Analytics Center (NVAC) under the leadership of visionary computer scientist James ("Jim") J. Thomas [1946–2010]. These and related efforts have recently evolved into the Visual Analytics Community, which sponsors conferences and workshops. The field of visual analytics is now considered an essential methodology for improving communication of data mining results and procedures.

Recommended Readings

N. Agarwal, H. Liu, *Modeling and Data Mining in Blogosphere* (Morgan & Claypool, New York, 2009). Available free online

E.E. Azar, S. Lerner, The use of semantic dimensions in the scaling of international events. Int. Interact. **7**(4), 361–378 (1981)

R. Feldman, J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data* (Cambridge University Press, Cambridge, 2007)

M.D. Fischer, S.M. Lyon, D. Sosna, Harmonizing diversity: tuning anthropological research to complexity. Soc. Sci. Comput. Rev. **31**(1), 3–15 (2013)

D.J. Gerner, P.A. Schrodt, Ö. Yilmaz, R. Abu-Jabr, The Creation of CAMEO (Conflict and Mediation Event Observations): An Event Data Framework for a Post Cold War World. Paper presented at the annual meeting of the American Political Science Association, San Francisco (2002)

M. Gorman, *Simulating Science* (Indiana University Press, Bloomington, 1992)

L.A. Grenoble, L.J. Whaley, *Endangered Languages: Current Issues and Future Prospects* (Cambridge University Press, Cambridge, 1998)

D.R. Heise, Project Magellan: collecting cross-cultural affective meanings via the internet. Electron. J. Sociol. **5**(3) (2001). Available online: http://www.indiana.edu/~socpsy/papers/magellan/Magellan.htm

T. Hermann, H. Ritter, Listen to your data: model-based sonification for data analysis, in *Proceedings of the ISIMADE'99*, Baden-Baden, Germany (1999)

O.R. Holsti, *Content Analysis for the Social Sciences and Humanities* (Addison-Wesley, Reading, 1969)

D.J. Hopkins, G. King, A method for automated nonparametric content analysis for social science. Am. J. Polit. Sci. **54**(1), 229–247 (2010)

W. Hsu, M.L. Lee, J. Wang, *Temporal and Spatio-Temporal Data Mining* (IGI Publishing, New York, 2008)

G. King, W. Lowe, An automated information extraction tool for international conflict data with performance as good as human coders: a rare events evaluation design. Int. Organ. **57**, 617–642 (2003)

K. Krippendorf, *Content Analysis: An Introduction to Its Methodology* (Sage, Thousand Oaks, 2004)

K. Krippendorf, M.A. Bock (eds.), *The Content Analysis Reader* (Sage, Thousand Oaks, 2008)

P. Langley, Data-driven discovery of physical laws. Cogn. Sci. **5**(1), 31–54 (1981)

P. Langley, Heuristics for scientific discovery: the legacy of Herbert Simon, in *Models of a Man: Essays in Memory of Herbert A. Simon*, ed. by M. Augier, J.G. March (MIT Press, Cambridge, 2004), pp. 461–471

D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, M. Van Alstyne, Computational social science. Science **323**(5915), 721–723 (2009)

K. Leetaru, *Data Mining Methods for the Content Analyst: An Introduction to the Computational Analysis of Content* (Routledge, London, 2011)

B.L. Monroe, P.A. Schrodt (eds.), in *Political Analysis* (2008). Special Issue: The Statistical Analysis of Political Text **16**(4), Autumn

I.-C. Moon, K.M. Carley, Modeling and simulation of terrorist networks in social and geospatial dimensions. IEEE Intell. Syst. **22**(5), 40–49 (2007). Special Issue on Social Computing

C.E. Osgood, W.H. May, M.S. Miron, *Cross-Cultural Universals of Affective Meaning* (University of Illinois Press, Urbana, 1975)

R. Popping, *Computer-Assisted Text Analysis* (Sage, Thousand Oaks, 2000)

P.A. Schrodt, Short term prediction of international events using a Holland classifier. Math. Comput. Model. **12**, 589–600 (1989)

P.A. Schrodt, Pattern recognition of international crises using hidden Markov models, in *Political Complexity*, ed. by D. Richards (University of Michigan Press, Ann Arbor, 2000)

H.A. Simon, Autobiography, in *Nobel Lectures, Economics 1969–1980*, ed. by A. Lindbeck (World Scientific, Singapore, 1992)

P.J. Stone, R.F. Bales, J.Z. Namenwirth, D.M. Ogilvie, The general inquirer: a computer system for content analysis and retrieval based on the sentence as a unit of information. Behav. Sci. **7**(4), 484–498 (1962)

L. Tang, H. Liu, *Community Detection and Mining in Social Media* (Morgan & Claypool, New York, 2010). Available free online

J.J. Thomas, K.A. Cook (eds.), *Illuminating the Path* (IEEE Comput. Soc., Los Alamitos, 2005)

C. Williford, C. Henry, A. Friedlander (eds.), *One Culture: Computationally Intensive Research in the Humanities and Social Sciences—A Report on the Experiences of First Respondents to the Digging into Data Challenge* (Council on Library and Information Resources, Washington, 2012)

T. Zhang, C.-C.J. Kuo, Audio content analysis for online audiovisual data segmentation and classification. IEEE Trans. Speech Audio Process. **9**(4), 441–457 (2001)

Social Networks

4

4.1 Introduction and Motivation

> "Social network analysis is inherently an interdisciplinary endeavor. The concepts of social network analysis developed out of a propitious meeting of social theory and application with formal mathematical, statistical, and computing methodology."—Stanley Wasserman and Katherine Faust (1994: 10).

> "Social network analysis is neither a theory nor a methodology. Rather, it is a perspective or a paradigm. It takes as its starting point the premise that social life is created primarily and most importantly by relations and the patterns they form."—Alexandra Marin and Barry Wellman (2011: 22).

This chapter introduces the fundamentals of social network analysis (SNA) as a major field of CSS, and builds on previous chapters by examining social networks from the paradigmatic perspective of emergent social structures and graph theory, supported by social theory drawn from one or more of the social sciences.[1]

Social networks consisting of actors and social relations are ubiquitous across the social science disciplines. Networks are consequential and frequent in anthropology, economics, sociology, political science, and psychology—the Big Five social sciences—as well as in interdisciplinary areas such as communication, management science, international relations, history, and geography, especially human geography. Social networks have been recorded in human history since writing was invented in the ancient Middle East over 5,000 years ago. As we shall see, social networks actually originated much earlier— at the very dawn of humanity, most likely in East Africa.

Social network analysis consists of a paradigmatic view of the social universe; it is a theoretical perspective, not just a collection of methods. Social network analysis also provides a formal language for developing the science of social networks,

[1]The field of social networks modeling and analysis is different from "the science of networks" developed by physicists. This chapters deals with social networks modeling and analysis as a field of CSS. This is because the subject matter of social networks always involves social entities, although, as in other areas of CSS, the origin of the methodologies may come from a variety of disciplines.

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_4,

including a perspective that enables and facilitates Computational Social Science. Moreover, SNA supports and extends the analysis of complex coupled human-natural-artificial systems by providing useful concepts, notation, and applied principles. The content of this chapter is intentionally selective in order to highlight the main ideas and their scientific value. Also, a word on notation: while every effort has been made to respect prevailing usage among social network analysts, inconsistencies or ambiguities present in the literature require the introduction of mathematical notation coordinated with the object-based orientation of CSS, as opposed to the more variable-based orientation of traditional social science.

4.2 History and First Pioneers

The history of contemporary social network science, which comprises analysis, modeling, and theorizing, is the result of contributions from the social, mathematical, computational, and physical sciences—with the latter as the most recent contributions and still rather tentative and hypothetical, but nonetheless intriguing.

The following chronology of social network science provides a brief history of milestones:[2]

1736 Mathematician Leonard Euler [1707–1783] solves the Königsberg bridges problem—by proving that it had no solution!—thereby initiating the field of graph theory, *the* principal mathematical structure employed by social network science.[3]

1856 Nobleman and comparative political scientist Alexis de Tocqueville coins the term "social structure" in his classic work *The Old Regime and the French Revolution*. In the United States and among political scientists worldwide, de Tocqueville is best known for his monograph, *Democracy in America*, where he discusses the significance of civic organizations for the performance of democratic political systems.

1930s The **sociogram**—the first graph-theoretic mathematical model of a social group—is invented by psychiatrist Jacob L. Moreno [1889–1974], founder of **sociometric analysis** as a modern field of social science.

1937 The journal *Sociometry* is founded with J. L. Moreno as its first editor. The aim of the journal was no less than the integration of all the social sciences through the mathematical medium of graphs for modeling social relations.

[2]Freeman (2004, 2011) provides an extensive and highly recommended history of social network analysis. In addition, most major works in SNA include historical essays or notes. However, other significant connections to applied mathematics or complexity science have often been missed.

[3]This is the gist of the Königsberg bridges problem: is it possible to follow a path that crosses each of the seven city bridges exactly once, returning to the same point of departure? The answer is no, due to the presence of odd-degree nodes (a term defined later in this chapter). Note that the referent system for the Königsberg bridge problem is an interesting example of a coupled socio-natural-technological system composed of denizens, land, river, and bridges, respectively.

1940 Anthropologist Alfred Radcliffe-Brown [1881–1955], founder of the Theory of Structural Functionalism, develops the term *social structure*—defined as a complex network of social relations—and calls for development of discrete mathematical models.

1944–1946 The foundations of Causal Attribution Theory and the Theory of Structural Balance are established by social psychologist Fritz Heider, followed in 1953 by the pioneering work of Theodore M. Newcomb [1903–1984].

1946 The matrix-based approach to social network analysis is pioneered by Elaine Forsyth Coke and Leo Katz (Forsyth and Katz 1946), followed by many others.

1948 The earliest definition of "network centrality" is proposed by Alex Bavelas (1948, 1950), including pioneering applications in laboratory experiments on communications networks.

1950s and 60s Social network concepts such as density, span, connectedness, multiplex, and others are introduced as SNA experiences significant growth across the social sciences.

1951 Anatol Rapoport (1957, 1959, 1983), one of the greatest mathematical social scientists of the 20th century, publishes the first paper on random graphs (Solomonoff and Rappaport 1951), a decade ahead of Erdős's and Rényi's (1960) more influential paper.[4]

1953 The formalization of Cognitive Balance Theory using graph–theoretic models is pioneered by Frank Harary [1921–2005], one of the most prominent graph-theoretic mathematicians of the 20th century.

1954 The term "social network" is first used by anthropologist John A. Barnes [1918–2010].

1956 Heider's Theory of Structural Balance is formalized and significantly extended and generalized by Dorwin Cartwright and Frank Harary (Cartwright and Harary 1956).

1957 Anatol Rapoport publishes the first of what is now called the "preferential attachment mechanism" in biased networks: well-connected nodes (with high degree) attracting yet more connections—as in a snowballing effect—a stochastic process pioneered by statistician George U. Yule in 1925.[5] In the same year the US Navy invents PERT (Program Evaluation and Review Technique), a network method for complex project management of the Polaris nuclear submarine program—an artifact, in the sense of Simon, of unprecedented complexity.

---

[4]In 1960 mathematicians Paul Erdős and Alfréd Rényi published their own paper on random graphs, reinventing the wheel nine years after Rapoport's seminal publication, and proposing new results.

[5]In 1999 the same mechanism of preferential attachment was re-proposed for the emergence of scaling in random networks (Barabasi and Albert 1999), decades after Anatol Rapoport's work on biased networks.

1958 The so-called "small world phenomenon" is conjectured for the first time using a mathematical model to predict how the world population is interrelated (de Sola Pool and Kochen 1978). The original paper was published 20 years later in the inaugural issue of the journal *Social Networks*.

1960s American sociologist Harrison White, from Harvard's Department of Social Relations, establishes social network analysis as a field in its own right.

1961 The first social network analysis of international relations, based on empirically-referenced graph–theoretic models applied to the Middle East, is published by Frank Harary in the *Journal of Conflict Resolution*.

1963 Mathematical foundations for the formal theory of roles and positions in social networks are established in the anthropological study of kinship systems by White (1963) and Boyd (1969).

1965 Thomas Saaty, one of the greatest applied mathematicians of the 20th century, publishes his influential monograph on *Finite Graphs and Networks: an Introduction with Applications*, followed in 1968 by his essay "On Mathematical Structures in Some Problems in Politics." First demonstration of the power law in networks of scientific collaborators (de Solla Price 1965).

1967 Social psychologist Stanley Milgram demonstrates the so-called small world phenomenon conjectured ten years earlier by de Sola Pool and Kochen, showing that a random sample of the US population was separated by approximately six links.

Mid-1970s Social network analysts and graph theoretic modelers begin the study of networks over time, what is now called dynamic networks (Wasserman and Faust 1994: 16; Breiger et al. 2003).

1971 The concept of social role is formalized by social network analysts François Lorrain and Harrison White. The computer program SOCPAC I for structural analysis of sociometric data, written in Fortran IV, is published in the journal *Behavioral Science* by S. Leinhart.

1977 The International Network for Social Network Analysis (INSNA), the world's leading professional social science SNA organization, is founded by Barry Wellman.

Early 1980s The SNA computer software UCINET 1.0 is released by Linton Freeman.

1981 The First International Sunbelt Social Network Conference of the INSNA (Sunbelt I) is held in Tampa, Florida, with anthropologist H. Russell Bernard (2012) as keynote speaker.

1983 Sociologist Mark Granovetter discovers "the strength of weak ties."

1991 M. Granovetter initiates the Cambridge University Press monograph series on Structural Analysis in the Social Sciences.

1994 Stanley Wasserman and Katherine Faust publish the first (and to this day most) comprehensive SNA textbook, consisting of 825 pages.

1996 B. Wellman and collaborators initiate the study of computer-supported social networks (CSSNs) as a new domain generated by the Internet.

1998 A *small-world model*, based on the exponential random graph model, is proposed as a highly abstract model of a simple social network with g nodes and uniform constant node degree d (the number of links attached to a node), to enable analytical approaches from statistical physics (Watts and Strogatz 1998).[6]

1999 The power law or scale-free structure of both the Internet and the World Wide Web network are demonstrated (Faloutsos et al. 1999; Albert et al. 1999).

2000 A new measure of clustering, the clustering coefficient C, is introduced by Barrat and Weigt (2000: 552).

2001 Swedish sociologist Fredrik Liljeros and collaborators demonstrate that sexually promiscuous individuals span a scale-free network, such that sexually transmitted deceases spread quickly through high-degree nodes.

2002 A binary decision model of so-called "global cascades" in d-regular random networks is proposed (Watts 2000).

2003 The first comprehensive survey of dynamic networks is published by the US National Academy of Sciences (Breiger et al. 2003).

2004 Computer simulations of a logit-type p* exponential random graph (ERG) network model demonstrate how combinations of parameter values can lead to a variety of network structures, including small worlds (Robins et al. 2005, 2007).

2011 The *SAGE Handbook of Social Network Analysis* is published as "the first published attempt to present, in a single volume, an overview of the social network analysis paradigm" (Carrington and Scott 2011: 1).

HOW DID SOCIAL NETWORKS ORIGINATE? Between ca. 100,000 years ago and ca. 10,000 years ago—i.e., for most of our common history as a species—humans lived *exclusively* in **kin-based networks** or family, household, and extended family networks. Migratory flows of these primary social networks wandered "out of Africa" ca. 100,000 years ago maintaining the same social structure for tens of thousands of years. Beginning just 10,000 years ago the very first non-kin networks emerged from social dynamics in hunter-gatherer societies in the form of simple **chiefdoms**—the first *networks-of-networks*. Some networks of chiefdoms eventually evolved shortly after into **states**, forming the first social *networks-of-networks-of-networks*, where `State = networkOf(Chiefdom = networkOf(Family))`. States formed the first **interstate networks** by

[6]The Watts-Strogatz model is d-regular with $Var(d) = 0$, a class of very rare social networks (Wasserman and Faust 1994: 100–101). Terminology and notation are confused by physicists using the symbol k to denote node degree δ. Other physics terms for node degree δ include number of neighbors, node connectivity, nearest neighbors, wired vertices, and so on, which is reminiscent of the Tower of Babel lamented by social scientists (Sartori 1970; Collier and Gerring 2009). Node degree δ is the standard terminology of SNA used here.

Table 4.1 Origin and evolution of the earliest social networks between 100,000 and 5,000 years ago (100–5 kya) according to system-of-systems network order $O(N)$

Year (B.C.)	Network N	Composition	Network order $O(N)$
3,000	Alliances	Groups of states[a]	4
4,000	States	Groups of non-kin-based groups	3
10,000	Chiefdoms	First non-kin-based groups	2
100,000	Families	Kin-related individuals	1

[a]*Organizations* of states, known in contemporary social (political) science terminology as *international organizations*, did not form until the 19th century A.D., following a **phase transition** (a term explained in Chap. 6) initiated by the 1815 Congress of Vienna

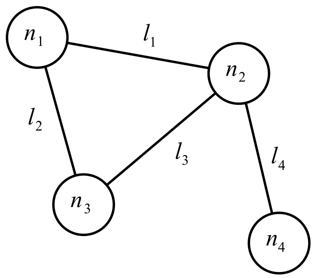


Fig. 4.1 A social network consisting of nodes and links. In this network $g = 4$ nodes and $L = 4$ links

ca. 6,000 years ago in the Middle East during the so-called Middle Uruk period (ca. 3750–3500 BC; Rothman 2001; Algaze 2008). These phase transitions have marked what may be called "The World History of Human Social Networks," summarized in Table 4.1.

4.3 Definition of a Network

A social network consists of several constituent parts which include *entities* (actors, values, sentiments, ideas, locations, attributes), *relations* (links, ties, associations, affiliations, interactions, evaluations), and *aggregations* (dyads, triads, groups, and subgroups). In this section we examine these ideas before introducing some necessary quantitative, mathematical, and computational aspects. Graph theory, algebraic methods, matrix algebra, and probability theory provide the main mathematical foundations of social network analysis. Together they represent a scientifically fertile and powerful suite of ideas, which explains why social network models play such a prominent role in computational social science. In particular, "graph theory provides both an appropriate representation of the social network and a set of concepts that can be used to study formal properties of social networks" (Wasserman and Faust 1994: 15). Formally, graphs are to networks as decision-theoretic models are to decision-making, differential equations are to dynamical systems, and game-theoretic models are to strategic interactions.

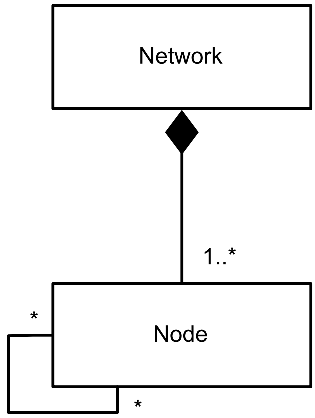


Fig. 4.2 UML class diagram of a social network as an object composed of node objects associated to the network by composition

A **network** $\mathscr{N}$ consists of a finite set $\mathbb{N}$ of entities (called nodes or vertices), denoted by $\{n_1, n_2, n_3, \ldots, n_g\}$, and a set of relations $\mathbb{L}$ (called lines, links, or edges), $\{\ell_1, \ell_2, \ell_3, \ldots, \ell_L\}$ defined on the set of nodes $\mathbb{N}$.[7] Note that g is the cardinality of $\mathbb{N}$ or total number of nodes in $\mathscr{N}$. The cardinality of $\mathbb{L}$ is $L = \binom{g}{2} = g(g-1)$ for directional pairs. A directional relation between node i and node j is denoted by $n_i \to n_j$ or x_{ij}. Figure 4.2 shows a simple example.

This is a fundamental concept upon which many other kinds of network concepts, models, and methods are built. As we shall see, the possibilities are practically infinite—and, most important, scientifically insightful—for advancing our understanding of social networks.

4.3.1 A Social Network as a Class Object

As we just saw, the classical formal definition of a social network is as a finite graph, a tradition dating back to the founding pioneers in the late 1950s and early 1960s, before the origins of the object-orientation to modeling. Recall the distinction between composition (denoted by a *solid diamond-head* ♦) and aggregation (*blank diamond-head* ◊), introduced in Chap. 2. Based on the same definition of a social network $\mathscr{N}$ as a graph, from a computational perspective we can also view a network as a class, a very general type of social object that is *composed* (i.e., *not* merely an aggregation) of nodes of various kinds that can have any number of relations among them.

This idea of a social network as a class having object instances is illustrated in Fig. 4.2 using a UML class diagram. We use the class diagram in short form (no attributes or methods are specified yet) to focus attention just on the main entities of interest: the network $\mathscr{N}$ with its nodes $\mathbb{N}$ and relations $\mathbb{L}$ (later we examine more closely the attributes of each). In other words, the *self-association* of nodes has *arbitrary multiplicity* in a network. This is an insightful perspective for understanding

[7]Note the formal mathematical translation of social entities into graph-theoretic nodes and social relations into edges.

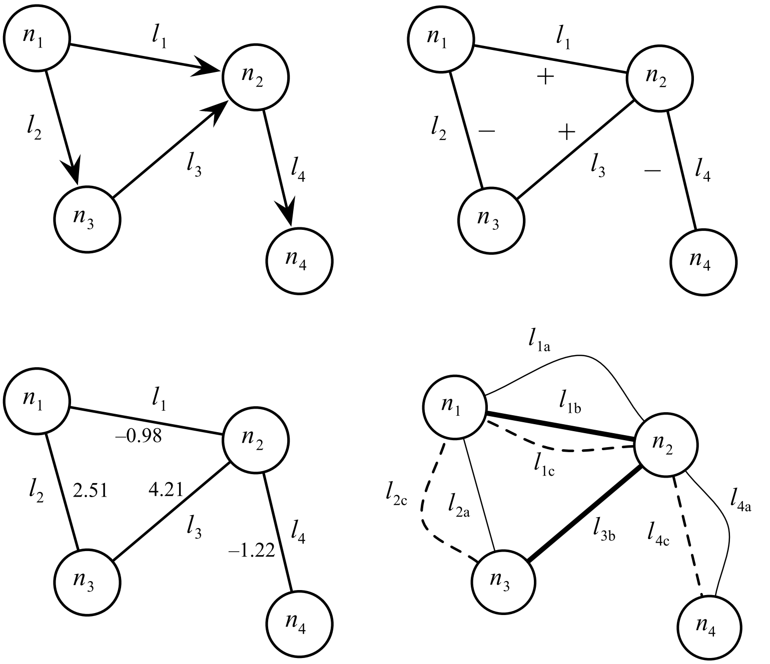


Fig. 4.3 Types of social networks according to their social relations $\mathbb{L}\{\ell_{1,2,\ldots,L}\}$. *Upper left*: a directed graph or digraph $\mathscr{D}$. *Upper right*: a signed graph $\mathscr{S}$ with valences. *Lower left*: a weighted network $\mathscr{W}$. *Lower right*: a multiplex $\mathscr{M}$ with various kinds of social relations possible between nodes

the essence of a social network, one that is not apparent from a graph-theoretic perspective. This object model of a social network complements the graph model in the same way as alternative models of the same phenomenon complement each other.[8]

Note also that the type of association between a node and its network is one of *composition*, not mere aggregation. Why? Because a node has no social meaning outside a network; a node is socially meaningful only within the context of some network, even if it is isolated from other nodes in the network, in which case it is called an **isolate node**.

4.3.2 Relational Types of Social Networks

Several interesting variations on the core concept of a social network $\mathscr{N}$ are highly significant in terms of the nature of social relations and the state of a network (see Fig. 4.3).

[8]A classic example of complementary models of the same phenomenon are the wave model and the particle model of light.

A **directed network** or **digraph** $\mathscr{D}$ (in Fig. 4.3, upper left) is a social network with directional social relations. While in a simple network the links between nodes lack specific direction, in a digraph or directed graph each line or association has a definite orientation or direction. This large class of social networks in social science includes the vast variety of transaction networks, such as those consisting of flows between nodes. Transaction flows typically refer to persons (e.g., flows of migrants, tourists, refugees, diplomats, or international students), money or goods (trade transactions), or other resources (imports/exports, information). All directional data is generally susceptible to social network analysis using digraphs.

A **signed network** or **valued network** $\mathscr{S}$ is a social network where the links have valence signs: $+, -, 0$ (see Fig. 4.3, upper right). For example, in politics, allies, adversaries, and neutrals have these kinds of relations. In psychology, belief systems are composed of ideas that are congruent, opposed, or unassociated—which are states marked by signs. Affect Control Theory is based on valence networks and the logic of cognitive consistency pioneered by F. Heider, L. Festinger, and R. Abelson.

A **weighted network** $\mathscr{W}$ is one where the links have weight or intensity of some kind (in Fig. 4.3, lower left). For example, a network of cities is related by pairwise distances between them, as shown by tables in travelers' maps. Similarly, airports are linked by flying times between them. Other weighted networks include volume of trade between countries, strength of friendship ties, and many other common social networks.

A **multiplex** $\mathscr{M}$ is a social network with one or more multiple/parallel associations between node pairs (in Fig. 4.3, lower right). In other words, the set of social relations $\mathbb{L}$ contains multiple social ties or links between nodes. For example, let $\mathscr{N}$ denote a small company with a set of employees $\mathbb{N}$. In this case, employees may be related/associated in a variety of ways, not just in a single way through their working association in the same small company. For instance, they may be related by kin relations, residential neighborhood, shared enthusiasm for the goals or products of the company, or through friendship ties, among many other interesting social possibilities. Empirically, many real-world networks of interest—from families and other "simple" networks to large and complex networks such as international organizations—are multiplexes. In practice, however, most SNA is confined to single-relation networks.

Paths are of interest in social network analysis. An **Eulerian path** is one that crosses each link exactly once. A **Hamiltonian path** is one that visits each node only once. **Hamiltonian distance** is defined by the number of nodes along a Hamiltonian path.

4.3.3 Level of Analysis

The level of analysis is a significant aspect in the architectural structure of a social network. Several levels of analysis are distinguishable and insightful. From micro- to macro-level ("bottom up"):

- **Nodal level**: The most detailed level of social network analysis focuses on attributes of node-entities, such as nodal degree, centrality, prominence, status, and other significant roles, such as being a bridge or an isolated entity. We have already seen that a node is an object, so attributes are encapsulated in nodes. Nodal attributes come in all kinds of data types (integer, string, Boolean, and so on, or corresponding values on the Stevens scale: nominal, ordinal, interval, and ratio). Nodal level analysis of a social network often involves statistical frequency distributions and their associated mathematical models: probability distributions. We will examine these in Sect. 4.6.
- **Dyadic level**: A relational pair can be analyzed as a binary unit from a number of perspectives, including but not limited to the attributes of the relationship. All the networks in Fig. 4.3 contain dyads. Given the different types of networks already seen in Sect. 4.3.2, the fundamental significance of the dyadic level should be clear: the qualitative type of dyads comprised in a social network can determine the very character of the network.
- **Triadic level**: Social triads are often significant, given the role they can play in balancing processes and transitive relationships, among others. Network triads are significant at all scales of social networks, from cognitive balance in psychological belief systems ("the friend of my enemy is my enemy") to international relations and political dynamics in alliance systems. Triads can also be the building blocks of more complex social networks.
- **N-adic level**: By induction, social network analysis can examine any aggregation of unit nodes and relations, up to the entire size of the network. If $N = g$ denotes the total number of nodes in a network, then the g-adic level of analysis is the same as analyzing the whole social network $\mathscr{N}$. These N-adic levels of analysis are significant in the field of communication research, among others, where audiences of various kinds can be defined in terms of sub-networks ranging from dyads to the complete network, with combinations in between.
- **Network level**: A set of concepts, measures, and properties is also defined for the most aggregate level of a network, which examines macro-level, aggregate attributes such as size, diameter, connectedness, centralization, density, and others. Analysis at the network level can involve aggregate or emergent properties and phenomena. For this reason, the network level is most commonly associated with complex systems analysis.

Most of what we know today about social networks is at the node level and the network level. However, a set of measures is defined for each of these other intermediate levels, as we shall examine further below, so in principle, any social network can be described in great quantitative detail, given sufficient data, regardless of the specific structure of the network. In fact, such detailed quantitative descriptions are important for understanding network structure.

Cross-level analysis, which, as the name indicates, investigates properties and dynamics involving multiple network levels, is also of significant scientific interest in computational social science. An example of this are critical changes in properties at the level of nodes that are consequential for inducing phase transitions at the global network level. These and other network dynamics will be examined subsequently, after we have learned more about the properties and structures of social networks.

4.3.4 Dynamic Networks

So far we have considered social networks examined from a static perspective. Such a perspective is legitimate for situations when network composition or structure are relatively invariant, stable, stationary, or static within a given time period (epoch). Obviously, that is not always the case in the system of interest. A **dynamic network** $\mathscr{N}(t)$ is a social network whose state changes as a function of time t. Dynamic networks can exhibit many interesting forms of behavior: nucleation, growth, evolution, transformation, disintegration, decay, or termination, among other patterns. The history of a dynamic network can range from relatively simple to highly complex, depending on the social network in question and its circumstances. For example, the history of a small group with fixed start and termination times, such as an airline flight with passengers or a ceremony with organizers and participants, spans a relatively simple dynamic network. By contrast, the evolution of international organizations, from the Concert of Europe to the United Nations system today, or the evolution of global terrorist networks such as al-Qaeda and affiliate organizations, represent hugely complex dynamic networks. Historically, the most ancient non-kin-based dynamic networks were trade networks that originated in Asia, perhaps as long as 5,000 years ago, and—somewhat later—in the Americas. We shall return to dynamic networks in the next chapter.

Note that all these important ideas about the concept of a social network $\mathscr{N}$ are defined *independently* of the specific structure or special features of a network. That is to say, these and many other properties of social networks hold true regardless of the specific nature of the social network being investigated.

4.4 Elementary Social Network Structures

Social networks in the real world vary significantly according to structure or "architecture." However, certain types of structural patterns—we may call these "elementary structures"—are significant for their properties and recurrence, either in pure form or in combination with others. In this section we define and illustrate these different types of social networks and in the next section we introduce quantitative methods for measuring their properties at various levels of analysis.

Graphs of different types of networks are illustrated in Fig. 4.4, together with their associated matrices and attribute measures—which are explained in the next sections.[9] The social networks in all figures are shown without reference to their

[9]The first four network structures represented in Fig. 4.4—known as the chain, the wheel, the Y, and the circle—can be called **Bavelas networks**, after the MIT social psychologist who first investigated their properties in the context of communication networks.

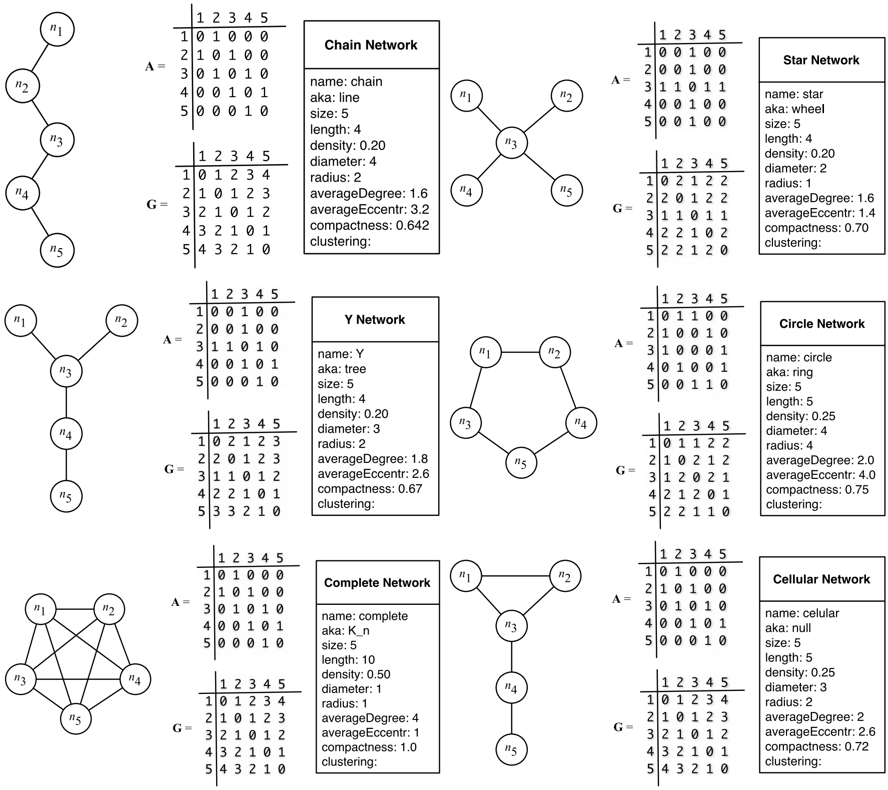


Fig. 4.4 Structural types of social networks according to their architecture. *Upper left*: chain or line network. *Upper right*: star network. *Middle left*: Y network. *Middle right*: circle network. *Lower left*: complete network. *Lower right*: cellular network. Each structural type is represented by its associated graph, adjacency matrix **A** and geodesic matrix **G**

relational type (i.e., directed, valued, weighted, or multiple); each of them can have any relational type, depending on the nature of its dyads. Moreover, all six networks in the figure have the same size (number of nodes = 5, a property defined later in this chapter), but most other structural features vary across the six cases.

Later in this chapter we will examine the attributes of each in greater detail. The purpose right now is to understand the variety of structural types, by moving from the simplest to some of the most complex. The following is a brief description of some of the most important structures in social networks. Familiarity with the terminology of network structures is important for communicating and discussing their characteristics and properties. These are among the most common, in approximate order of increasing complexity:

Simple network: A network without loops or parallel/multiple links. All social networks in Fig. 4.4 are simple.

Chain network: A string of nodes, also known as a **line network**. Supply chains and multi-stage processes of many kinds are common social examples.

Star network: Central node is radially linked to all the other nodes around it. Also known as a **wheel network**. This network has a more centralized structure. Hierarchical organizations have this common structure.

Y-network: A chain with split or frayed terminal path. This structure is also known as a **tree network**. Social examples include many organizational charts, all games in extensive form, and branching processes, among others. A tree structure is also common in computational algorithms.

Forest network: Set of disconnected trees. Although disconnected, a social network can be composed of a set of trees or other networks.

Circle network: A closed chain where nodes are linked in a circle fashion. This is also known as a circle and it resembles the chain network but without terminal nodes. This is the least hierarchical of the structures seen so far.

Cyclic network: A graph containing one or more cycles. The smallest cycle is a triad. The complete network and the cellular network in Fig. 4.4 are cyclic.

Acyclic network: Contains no cycles. The chain network, the star, and the Y network are all instances of acyclic networks.

Connected network: Every pair of nodes is joined by at least one chain. All six social networks in Fig. 4.4 are connected.

Component network: A disconnected subgraph. A tree is a component of a forest.

Complete network: Every node is connected to all others. A complete network is shown in Fig. 4.4. A complete network has maximum communication and may or may not indicate lack of hierarchy, depending on the nature of nodes.

Bipartite network: A network with a node set that can be partitioned into two disjoint sets, $\mathbb{N}_1$ and $\mathbb{N}_2$, such that every link has one end in $\mathbb{N}_1$ and the other in $\mathbb{N}_2$. Political party affiliations, a list of refugee camps and countries where they are located, phone directories, a price list, and a list of countries and capitals are common examples of bipartite networks.

Cellular network: A network in which one or more nodes has a complete graph attached to it. The last example in Fig. 4.4 is a cellular network. Terrorist networks are often organized this way.

Nonplanar network: A network that cannot be drawn on only two dimensions. Most social networks are nonplanar, as is typical of "hair-ball" graphs in the popular media. All five Bavelas networks are planar, as are all forest networks and composites of these.

Random network: A network model with the property that the probability of links forming between nodes is governed by some probabilistic process. Social examples include networks of relations in which people become acquainted by chance; social networks containing dyads intentionally drawn from a lottery; and a variety of growth processes.

Small-world network: Social structure in which most nodes are not adjacent to one another, but can be reached from other nodes by just a small number of links. This social structure lies more or less between a complete network and a much simpler network structure having only neighbors.[10]

Scale-free network: Social structure in which degree distribution follows a power law, such that most nodes in the network have few neighbors, some have many more neighbors, and just a few nodes have a huge number of links.

Broad-scale network: Same as a scale-free network but with sharp cutoff, such that there are not as many highly connected nodes as would be expected by a power law.

Single-scale network: Social structures with degree distribution characterized by a fast decaying tail; i.e., not power law.

The Internet and the World Wide Web are two distinct networks. The former refers to the physical network of computers, while the latter is a network of hyperlinks via URLs. The more social of the two is the World Wide Web, since people are more closely associated with URLs (e.g., social media websites, personal pages, and so forth), whereas the Internet is mainly a network of servers linked by communications systems and related hardware.

4.5 The Network Matrix

The relational structure of a given social network $\mathscr{N}$ is represented by a matrix $\mathbf{M}_{\mathscr{N}}$. Several graph matrices can provide formal canonical definitions of a social network. When a social network is defined in terms of linked or adjacent neighbors, the network matrix $\mathbf{A}$ is called a **sociomatrix** (Moreno 1934) or **adjacency matrix**, where a_{ij} denotes an element of the binary $g \times g$ sociomatrix $\mathbf{A}_r$. The sociomatrix is defined strictly in terms of the node set. Other social network matrices of interest can also be defined by selecting different sets of interest (e.g., $\mathbb{L}$) and combinations thereof.[11]

Social network analysis uses both conventional matrix notation from linear algebra and simple tabular notation to represent a sociomatrix in full form:

$$\mathbf{A}_{g\times g} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1g} \\ a_{21} & a_{22} & \dots & a_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ a_{g1} & a_{g2} & \dots & a_{gg} \end{pmatrix} = \begin{array}{c|cccc} & n_1 & n_2 & \dots & n_g \\ \hline n_1 & a_{11} & a_{12} & \dots & a_{1g} \\ n_2 & a_{21} & a_{22} & \dots & a_{2g} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_g & a_{g1} & a_{g2} & \dots & a_{gg} \end{array}. \tag{4.1}$$

The **distance matrix** $\mathbf{D}_{\mathscr{N}}$ is defined in terms of minimal path distances between all connected nodes, where each element $d_{ij} \in \mathbf{D}_{g\times g}$ denotes the minimal number of links between node n_i and node n_j.

[10] See Amaral et al. (2000) for an excellent survey of the main classes of small-world networks.

[11] From a graph-theoretic perspective, see Busacker and Saaty (1965: Chap. 5), Wilson (1985).

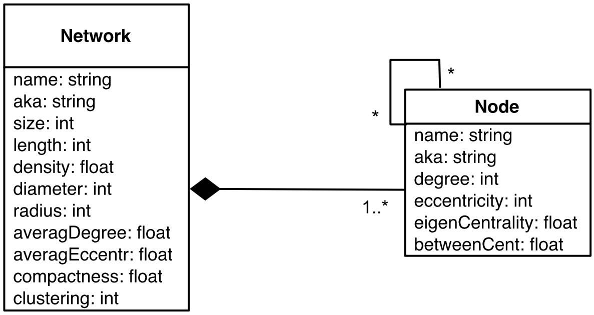


Fig. 4.5 Long-form UML class diagram of a social network modeled as an object composed of node objects associated to the network by composition. This model highlights the nodal composition of networks while placing network links in the background

4.6 Quantitative Measures of a Social Network

There are two main classes of social network measures: *micro-level nodal measures*, which are attributes of nodes, and *macro-level network measures*, which are aggregate attributes that characterize features of network structure as a whole. Sub-group or sub-network measures (e.g., for cliques) are just constrained versions of the latter (e.g., the size or density of a clique). A computational way of thinking about these measures at various levels of analysis is as attributes of their respective object, be it the nodal or the network level of analysis. This idea is summarized in Fig. 4.5 and each of the attribute-measures is examined in this section.

4.6.1 Nodal Measures: Micro Level

The following nodal measures are all defined with respect to node $n_i \in \mathscr{N}$. Each nodal measure is an attribute of the node object, so each node has all of these measures—plus any number of others that may be of interest. New measures are being invented all the time, some more significant than others. Historically, the first nodal measure is the so-called "degree" of a node. While some of these measures have intrinsic value, they are also used to define macro-level measures for the network as a whole.

Degree $\delta(n_i) = \delta_i = \sum_j a_{ij}$. Number of links incident on a node. Sum of a node's $a_i j$ elements in the sociomatrix. Number of incident alter nodes. Degree is a measure of centrality, sometimes called degree centrality (as opposed to other kinds of centrality defined below).

Distance between n_i and $n_j = d(n_i, n_j) = d_{ij}$. The minimal (so-called **geodesic**) number of links in any chain connecting n_i and n_j. Thus, $d(n_i, n_i) = 0$ for all $n_i \in N$.

Eccentricity $\epsilon(n_i) = \epsilon_i$. Maximum geodesic (i.e., shortest-path) distance between node n_i and any other node n_j. Nodal eccentricity is a measure of how far the node is from the most remote terminal node (boundary) of the entire network. A graph has as many eccentricities as there are nodes, since eccentricity is a nodal attribute.

Eigenvector centrality $c_e(n_i) = \lambda \sum_j a_{ij} e_j$, where λ is the eigenvalue and e_j is the eigenvector centrality score. Same as nodal degree but weighted by the centrality of each incident/adjacent node. Measure of a node's influence. Has inspired the model for Google's PageRank measure, which is a version of eigenvector centrality. Given two nodes with the same degree, the one linked to other nodes with high degree will have greater influence (eigenvalue centrality).

Betweenness centrality. Number of times that a node is a bridge in the shortest path between two other nodes. Number of geodesic paths from all vertices to all other paths that pass through that node.

4.6.2 Network Measures: Macro Level

The following are macro-level measures defined with respect to a given network $\mathcal{N}(\mathbb{N}, \mathbb{L})$, consistent with previous notation. These are illustrated in Fig. 4.4 for the six elementary network structures.

Size $S = card(\mathbb{N}) = |\mathbb{N}|$. Total number of nodes in $\mathbb{N}$. Note that the size of all the elementary networks in Fig. 4.4 is the same ($S = 5$). Social networks vary greatly by size, from small to large (e.g., Big Data networks).

Length $L = card(\mathbb{L}) = |\mathbb{L}|$. Total number of links in $\mathbb{L}$.

Density $Q = L/S(S-1) = L/(S^2 - S) \approx L/S^2$ for large S. Number of actual links relative to total number of possible links in N. Thus, network density is linearly proportional to network length and inversely proportional to the square of network size. Interestingly, for networks of equal length (same number of links), $Q \propto 1/S^2$, which is a power law and a *universal property* because it emerges independent of network structure.

Diameter $D = \max_{n_i \in \mathbb{N}} \epsilon(n_i)$. Maximum nodal eccentricity. Maximum geodesic distance in the network.

Radius $R = \in_{n_i \in \mathbb{N}} \epsilon(n_i)$. Minimum nodal eccentricity. Minimum geodesic distance in the network.

Average degree $\overline{\delta} = 2L/S = Q(S-1)$. Measures the general connectedness of nodes in the network. This is perhaps the most common network statistic besides size, which is informative so long as its distribution is fairly well-behaved (e.g., not multi-modal or highly skewed).

Degree skewness $Skew(\delta) = (\overline{\delta} - \hat{\delta})/\sigma_\delta$ (following Pearson's equation). Significant for detecting non-equilibrium distributions, because the distribution of degree can have many forms.

Average eccentricity $\bar{\epsilon}$. Measures the general "width" of a network. As with all averages, it should be interpreted conditionally upon information about its distribution.

Compactness C. Defined by the equation

$$C = \frac{\sum_{i \neq j} (1/d_{ij})}{S(S-1)}, \tag{4.2}$$

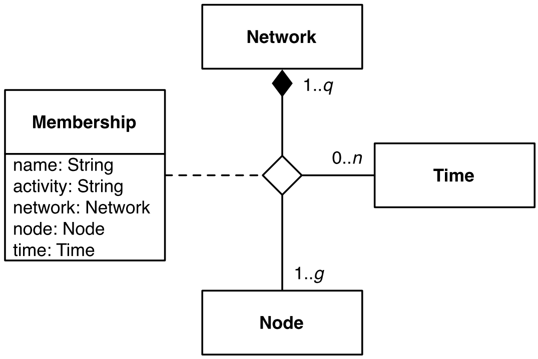


Fig. 4.6 UML class diagram of a dynamic social network represented as a ternary association class with multiplicities. Each link in the association corresponds to a membership in one or more (up to q) concurrent networks over a period of n time units

where d_{ij} are the dyadic distances in the network. Note inverse distances must be computed using the geodesic distance matrix $\mathbf{G}$ to derive $\mathbf{G}^* = \{1/d_{ij}\}$. The elementary social structures in Fig. 4.4 vary in compactness from 0.642 (chain network) to 1.0 (complete network, as expected).

4.7 Dynamic (Actually, Kinetic) Networks as Ternary Associations

All the networks we have discussed so far in this chapter have been formally static, in the sense that we have been assuming that their basic structural features do not change over time. A **dynamic network** is one that experiences change in the number of nodes or links.[12]

Earlier we saw how a social network could be seen as a binary association—i.e., between a network and node objects (recall Sect. 4.3.1 and Fig. 4.2). In the real social world, binary associations—as between $\mathscr{N}$ and $\mathbb{N}$—are quite common. However, sometimes social systems and processes are best modeled as ternary or higher associations. A dynamic network is a membership type of **ternary association** among the network, its nodes, and time.

An n-ary association consists of a relationship among n classes. A set of concurrent dynamic networks is an example of this for $n = 3$, as shown in Fig. 4.6. Note that the association in this case does not belong exclusively to any of the three classes. Rather, the association depends on all three classes simultaneously. In Fig. 4.6 the multiplicities are constrained as follows: (1) A node (actor) may belong to as many as q networks at any given time; (2) each network can have between one and g nodes in a given year; and (3) a node (actor) may belong between zero and n time units in any given network.

[12]Etymologically speaking, the term "dynamic" should be reserved for analysis of change as a function of forces of some kind, as indicated by the Greek root *dynamos*—which means force. The term kinematic or kinetic also means change, but without attribution to or explicit treatment of causal forces. Loosely speaking, unfortunately, it has become common in social science to call dynamic anything that changes with time. The proper term in "kinetic" or "kinematic."

4.8 Applications

Networks are ubiquitous and highly significant throughout social science. In this section we look at several classic and contemporary applications in a variety of domains. A useful way to approach such a large number of applications across domains of the social universe is to examine them from “micro” or individual-based models, which exist in the minds of actors, to “macro” or global-based models that form among collective social groups, such as nations and international organizations. This is also a *consilient* or hierarchical approach, in the sense of E.O. Wilson (1998), since most micro models are in some sense embedded in macro models, although they are not always explicitly treated as such.

4.8.1 Human Cognition and Belief Systems

We as humans form mental **images** of the world we perceive. Such images are significant to recognize and understand, for we use them all the time for judgment and decisionmaking, rather that basing our decisions on direct, unmediated data from the real world. In other words, we perceive the world through our personal receivers (senses, paradigms, schemata, theories, and similar cognitive structures), and then form a mental image of such a world. Images support human decisionmaking and subsequent actions. Another term for the concept of image is **individual belief system**, which is useful for highlighting the complexity of these constructs.

A belief system may be more or less realistic, depending on its empirical validity. What matters most is that images exist—whether real or imaginary—and we use them all the time. In a sense, therefore, the degree of realism of images or belief systems is secondary (an attribute among many others) relative to the fact that they exist. Belief systems are also a cross-cultural universal of humans, a feature not unique to any particular group or culture. Of course, different cultures develop different, sometimes even conflicting images of the same phenomenon—but the fact that all human decisionmaking is based on individual-based belief systems is a valid assumption about the social world.

Another salient feature of a belief system is that it is often shared. **Collective belief systems** are those shared among a group of people, such as beliefs about social identity, cultural norms and traditions, or national history. Clearly, collective belief systems are highly consequential and also universal across human cultures. For example, the concept of political culture can be defined as the set of beliefs that members of a given polity hold with respect to issues such as governance, fairness, equity, social role, and justice.

An image consists of a set of conceptual entities (nodes), which may be tangible or intangible, connected by various kinds of mental associations (links). Some examples of simple beliefs are shown in Fig. 4.7. In some cases nodes refer to actors (USA, North Korea) that contain **attitudinal values** (also called **affective valuations**), whereas in others they may represent ideas or concepts (friend, ally, freedom, tyranny).

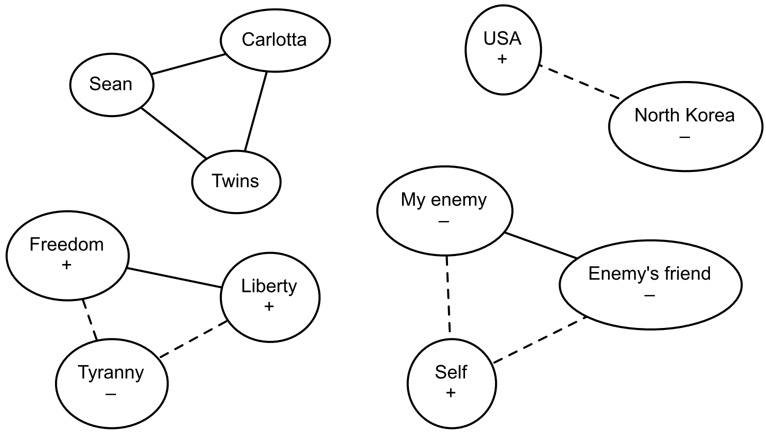


Fig. 4.7 Some simple beliefs modeled as valued networks

Two computationally remarkable and highly challenging properties of human belief systems are their sheer size and evolutionary dynamics. Moreover, our understanding of their full complexity remains rather incomplete, due to both of these features, among others. Human belief systems consist of networks that can span many orders of magnitude in size (no one has measured this with much precision)—a feature that is true with regard to both individual and collective belief systems. We all hold simple beliefs, such as those in Fig. 4.7. However, those are only small components (subgraphs) that are part of vast networks of ideas. Linguists estimate that the average person knows somewhere in the order of 10^4 words. Although this is less than 100,000 words, the number of possible associations and higher-order connections is of the order of 10^8, or tens of millions of dyadic links, not counting triads and higher-order (N-ary) associations. Collective belief systems are arguably orders of magnitude larger still. If each node and link holds a certain amount of information (say, in some proportion to the person's education or knowledge), it is easy to see how the total amount of information held by a human belief system is staggering.

There is another feature of human belief systems that is remarkable: in addition to being huge, belief systems are also dynamic, not static, as discovered almost a century ago by social psychologists such as Fritz Heider and, subsequently, Robert Abelson. Belief systems change over time because valuations can change, perhaps as a result of new information, or because new nodes and links are added to prior beliefs. For example, the simple belief self+friend changes when a person learns about a friend's other friends or enemies, resulting in self+friend+friendOfFriend or self+friend−EnemyOfFriend, as the case might be. What is remarkable about this change is that the overall belief system maintains consistency, an important property or principle that is also known as **cognitive balance**.

In fact, cognitive balance obeys the logic of the **algebra of signs**:

$$+ \cdot + = + \tag{4.3}$$

$$- \cdot - = + \tag{4.4}$$

$$+ \cdot - = - \tag{4.5}$$

$$- \cdot + = - \tag{4.6}$$

This can be easily verified by the simple examples in Fig. 4.7, where positive links are denoted by solid lines and negative by dashed. In each case the algebra of signs yields a positive result, even in cases of multiple links, not just in dyadic cases. The same is generally true for much larger belief systems—both individual and collective belief systems—as has been demonstrated by numerous studies. The overwhelming cognitive structure of human belief networks is balanced.

How does this occur? How do humans maintain overall cognitive consistency as their belief systems evolve? This is apparently due to the existence of four cognitive balancing mechanisms, as discovered by Robert Abelson, who called them "modes of resolution of belief dilemmas" in one of the most famous papers of 20th-century social science:

1. **Denial**. The simplest way to balance an imbalanced belief is to deny or simply ignore any problematic parts. For example, one may choose to ignore the fact that a friend's friend is one's adversary and simply carry on normal good relations with the neighbor. This is quite common. The denial mechanism is not a true form of balancing because the inconsistency is not actually resolved, only ignored. Denial is sometimes referred to as a psychological defense mechanism.
2. **Bolstering**. A somewhat more sophisticated mechanism consists of emphasizing the balanced parts of a belief system and upholding those as being more important. For example, one might choose to highlight one's friendship with a neighbor as being more important than the fact that the neighbor is a relative of one's adversary. Again, this is quite common and not a true process for resolving inconsistency.
3. **Transcendence**. A third way is to appeal to a higher principle that—as the term suggests—transcends an imbalanced inconsistency. During the Cold War it became necessary to avoid nuclear war among the superpowers in spite of deeply conflictive relations. When truth is the victim of peace, "in the interest of peace" is a common form of balancing by transcendence, as is the principle of maintaining sociality "for the common good." Transcendence is a common, powerful, and important balancing mechanism for maintaining social cohesion, and it is frequently invoked, especially in times of crisis.
4. **Differentiation**. The most interesting mechanism works by splitting a concept into two (or more) newly derived concepts with a resulting structure that is somewhat more complex but also balanced and hence more stable. For example, one $(+)$ may dislike some group $(-)$ but for some reason it seems necessary to maintain good relations $(+)$, which produces cognitive imbalance: $+ \cdot + \cdot - = -$. This can be balanced by distinguishing between the "bad leaders of the group" $(-)$,

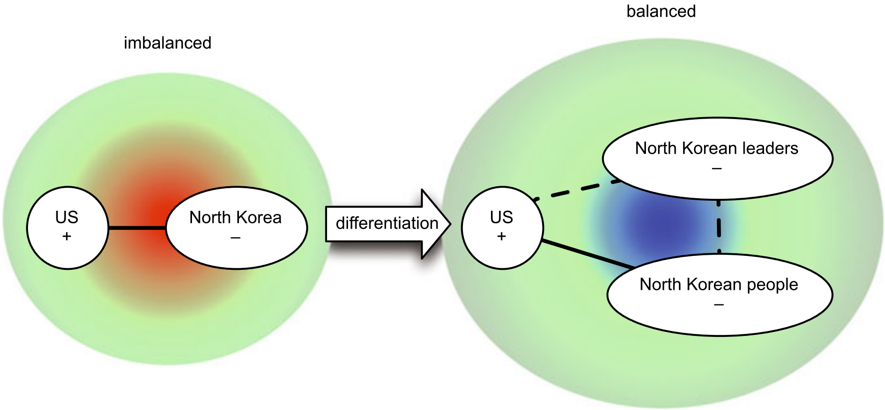


Fig. 4.8 Cognitive balancing by Abelson's differentiation mechanism. *Left*: Having positive relations with a country that is disliked results in an imbalanced cognition. This belief is balanced by differentiating between evil rulers and good people, and reassigning valuations to each of the new relations

whom we dislike, and the "good members of the group" (+) that are "oppressed" (−) by the nasty leaders. This differentiated structure is now balanced, as shown in Fig. 4.8.

Several features of the four cognitive balancing mechanisms are particularly noteworthy. First, they differ with respect to producing true balance, with differentiation producing complete balance and the other three maintaining some degree of inconsistency or pseudo-balance. Second, as a result of this first property, differentiation is a powerful mechanism because it produces highly stable, persistent beliefs that are more complex than the original imbalanced system but are more enduring. This explains its widespread occurrence. Third, all four mechanisms are cross-cultural universals found in all societies. Fourth, all four cognitive balancing mechanisms are also significant instruments of social control, as effective leaders understand. They can be used individually as well as in combination. Finally, from a computational perspective, relatively little use has been made of these mechanisms, although they are highly relevant and profoundly human. For example, they can and should be more extensively used in agent-based models and social simulations, as well as investigated in terms of complexity-theoretic properties since all four produce emergent phenomena.

4.8.2 Decision-Making Models

Going beyond the cognitive level, to the level where actors make decisions, we can also view human decision-making as a network. A *decision* can be defined as a choice within a *set of alternatives*, each of which has a *set of outcomes* associated with each alternative. In turn, each outcome has two significant attributes: the *utility* or value of the outcome and its *probability* of occurrence. Utilities and probabilities are then used to compute the *expected value* of each alternative, in order to choose the alternative having the highest expected value. In rational choice theory this is known as the Bayesian decision model, which is the basis for a large literature across the social sciences.

Figure 4.9 illustrates the network structure of the classic Rational Choice Model. Note that the overall network structure is that of a set of line subnetworks of equal length joined at the root (the decision $\mathbb{D}$, so to speak), as in a tree or n-star with embedded circle leaves.

Note that even a model of bounded rationality, with a limited set of alternatives and outcomes, as well as imperfectly known utilities and probabilities, will still span a network. Or, put somewhat differently, bounded rationality decision-making can still be usefully viewed as a network structure by modeling its components and associations in terms of nodes and links. In contrast with complex belief systems at the lower level of analysis, decision networks are relatively simple, especially those under assumptions of bounded rationality. The network structure of human decision-making is recognizable and remarkable.

4.8.3 Organizations and Meta-Models

A classic application of social networks, and one of the areas that originated the analysis of networks in social science, is to human organizations of many different kinds—from small groups or teams to large corporations and international organizations, global or regional. This is a very natural application of social network analysis, because human organizations lend themselves to multiple representations in terms of individuals and roles or functions within an organization. The well-known visual example of this is the *organizational diagram*, also known as an organigram(me) or organizational chart.

Another network model of organizations defines the set of nodes as consisting of various subsets that include people (agents), goals, knowledge, tasks, locations, resources, organizations, and the like. This type of heterogeneous network model—originally proposed by David Krackardt and Kathleen Carley—is called an **organizational meta-matrix** model or meta-network (Carley 2001). Figure 4.10 shows an example of a meta-model network of leaders, locations, and other relevant features. From a computational perspective these network models constructed by means of specialized algorithms, such as ORA (see Sect. 4.9 below), can process large corpora of text and other media.

4.8.4 Supply Chains

A **supply chain** is a linear array of sequential operations required to produce an end result. Complex societies (and even those that are not so complex) rely on supply chains of many different kinds to provide a vast array of goods and services. Such goods and services may be private or public. Some of these chains originated thousands of years ago, at the dawn of civilization. In fact, it is no exaggeration to note

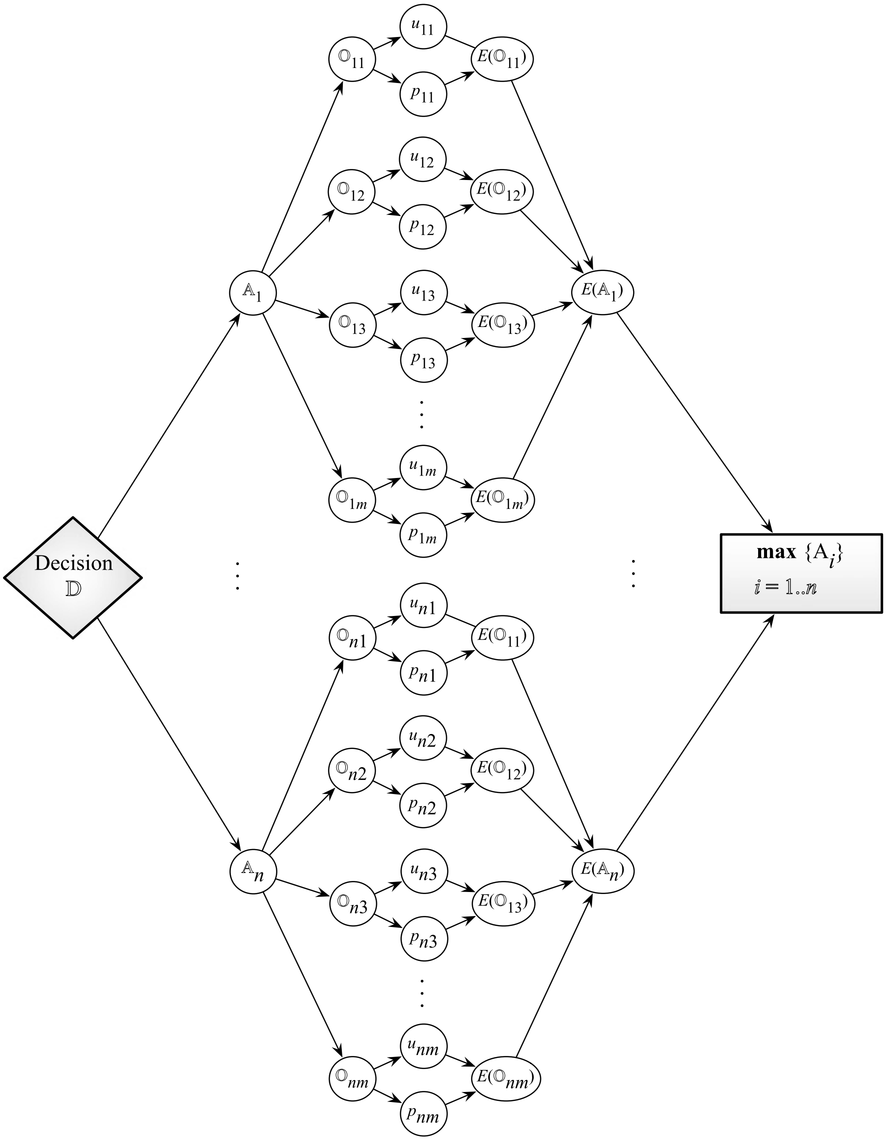


Fig. 4.9 Network structure of the Rational Choice Model. *Left*: A decision $\mathbb{D}$ consists of choosing an alternative $A^* \in \{A_i\}$ that has the maximum expected utility over the entire set of n alternatives

that the rise of civilization was rendered possible thanks to the design, implementation, and maintenance of complex supply chains. For example, the production of bronze—which occurred for the first time in the ancient Near East (Mesopotamia, present-day Iraq) during the 4th millennium B.C.—is an excellent example of a supply chain network that required the coordinated extraction of minerals, such as cop-

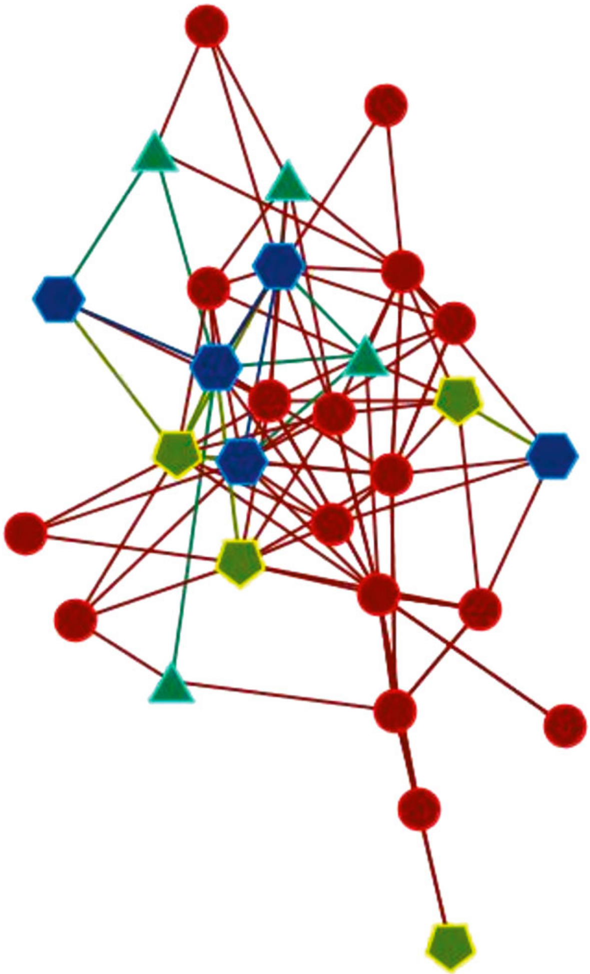


Fig. 4.10 Meta-network model of a social event involving actors, locations, resources, and other entities denoted by nodes and links of various shapes and colors. Produced by the ORA software at the Center for Computational Analysis of Social and Organizational Systems (CASOS), Carnegie Mellon University. A complex humanitarian crisis can be represented by a meta-network linking victims affected by the disaster, relief workers, supplies and equipment, locations, and responder activities. Similar examples include financial crises and conflicts of various kinds, all of them consisting of data n-tuples that can be extracted from raw sources

per, tin, zinc, and lead, involving hundreds and in some cases thousands of workers organized in a systematic way so as to produce the desired bronze artifacts. Today, modern manufacturing processes, as well as all kinds of services, involve supply chains. A particularly important class of supply chains involves those that support critical infrastructure and emergency services that are essential for the operational performance of contemporary societies.

The **first-order network structure** or basic organization of the supply chain is obviously a line or chain network. However, in almost all real-world examples, at least one and often all nodes require some degree of parallelization. Hence, the composite structure of complex supply chains involves a combination of serial and parallel networks. The field of systems science that studies such networks is called **systems reliability** and the mathematical foundations for developing models of complex supply chains and similar networks is very well developed.

Supply chains can be modeled through a variety of mathematical approaches. One particularly useful approach is to view the outcome of the supply chain—the end result—as a probabilistic outcome. Since the outcome depends on the successful completion of all prior, necessary stages in the production process, we may view the outcome of a supply chain as a compound event in the sense of elementary probability theory. Let P denote the probability of the outcome and $P_1, P_2, P_3, \ldots, P_N$ the probabilities associated with each of the necessary stages. Then,

$$P = P_1 \times P_2 \times P_3 \times \cdots \times P_N. \tag{4.7}$$

Equation (4.7) is based on the probability of a compound event and models the first-order network structure of a supply chain.

Now let Q denote the probability of a parallelized activity associated with one or more of the serial nodes, and let $Q_1, Q_2, Q_3, \ldots, Q_M$ denote individual parallel activities. Then,

$$Q = 1 - (1 - q_1)(1 - q_2)(1 - q_3) \cdots (1 - q_m). \tag{4.8}$$

Equation (4.8) models the second-order network structure, or substructure, of the supply chain. Combining both equations by substituting P_i component probabilities in Eq. (4.7) by their respective Q-equation, it is possible to derive a second-order equation for the probability of performance or production in a serial-parallel supply chain—as we examine later, in Chaps. 6 and 7.

Modeling real-world supply chains in social systems and processes often requires many levels of embedded serial and parallel components. Not surprisingly, this also is an area where computational approaches are essential and provide powerful and often counterintuitive results. In particular, human intuition is a very poor guide when it comes to understanding emergent patterns in serial and parallel systems such as supply chains and similar organizations. For example, human judgment almost always overestimates the overall reliability of the supply chain or serial system. The common saying that "a chain is as strong as its weakest link" is erroneous and can be very misleading. The correct saying should be " the chain is always less strong than the weakest of its links." This is because probabilities are values between 0 and 1, so, when they are multiplied, the resulting probability is always smaller—most times much smaller!—than the smallest probability in the chain. The opposite is true for parallel systems: the reliability of a parallelized system is always greater than the highest of the component probabilities. Given the supply chain system and that it combines various patterns of serial and parallel structures, the only way to really understand how the system will behave is to mathematically model the composite structure and conduct a computer simulation.

4.8.5 The Social Structure of Small Worlds

Earlier in Sect. 4.1 we saw how Stanley Milgram was the pioneering discoverer of the so-called **small-world structure** of social networks. In recent years others have rediscovered the same phenomenon in different social domains (as well as outside the social sciences, such as in biology, physics, and computer science).

A small-world network is a rather sparse network structure situated somewhere between a fully connected, complete network where every node is connected to every other node, and a random network that has minimal density. In a small-world network most nodes are not directly connected, but can be reached from other nodes by a small number of links.

An intriguing characteristic property of a small-world $\mathbb{S}$ at the network level of analysis is that the geodesic distance d_{ij} between two randomly chosen nodes n_i and n_j is proportional to the logarithm of the size S of the network:

$$d_{ij} = k \log S, \tag{4.9}$$

where k is a constant. This regularity may be called the **Watts-Strogatz Law**, after the discoverers Duncan J. Watts and Steven H. Strogatz. Given Eq. (4.9), it follows that the greatest increases in geodesic distance occur as a small network increases its initial size (as in a club that grows from just two or three friends), since $\partial d/\partial S < 0$. Similarly, the logarithmic effect vanishes in proportion to S, so large networks have typical distances largely insensitive to their size.

Why do small-world structures matter from a social perspective? Basically it is because things can propagate very quickly in small worlds, relative to more sparsely linked networks. For example, infectious diseases spread far more rapidly in a small-world community than in a society with higher "degrees of separation." The small-world phenomenon also explains the frequent occurrence of discovering friends in common, especially among people who do not know each other.

4.8.6 International Relations

Networks are also ubiquitous in the field of international relations—as already implied by the term itself. Some of the most common and well-known examples include trade networks (one of the most ancient forms of social networks); diplomatic relations that link foreign ministries to embassies, consulates, and other foreign posts; and politico-military alliances and international organizations. Networks in international relations are well documented since the 4th millennium BC, although it was not until recently that full data coverage became available for recent centuries.

Trade networks are usually modeled by sets of nodes that represent countries or economies and links that represent exports and imports. However, trade networks can also include much more detail. For example, nodes can be described in terms of various sectors of an economy, and links can represent detailed flows of raw materials, semi-manufactured and manufactured goods, and all kinds of services. Whereas trade networks used to be modeled by transaction matrices, today they are modeled using SNA as well as complexity-theoretic methods and related approaches.

Diplomatic networks in the international system can be of two kinds. A **national diplomatic network** is spanned by the Ministry of Foreign Affairs as the hub, and embassies and other diplomatic missions as end nodes, with regional offices or bureaus in between the two. Therefore, a national diplomatic network has the classical structure of a tree or star. By contrast, the **international diplomatic network** consists of all the countries and sovereign entities as nodes, and two-way, reciprocal diplomatic ties linking the nodes. Obviously, such a network is not complete, since not every country in the world has relations with all other members of the international system. In addition, countries have diplomatic relations with international organizations, such as the United Nations and a host of other international governmental organizations in the UN family, the European Union, NATO, and others. There is also a vast network of working relations among nongovernmental organizations in numerous fields that cover social, economic, cultural, and political affairs. The number of international organizations, including governmental and nongovernmental varieties, has skyrocketed since the first ones were established in the 19th century.

A particularly important type of international network consists of *alliances* in the global system. Well-known historical examples include the Triple Alliance and the Dual Entente during World War I, and the contemporary NATO alliance, among numerous others that have existed in the international system since the formation of early states and empires. A complete record of all alliances that have existed in history is not yet available, but in principle it should be possible to compile such a dataset, based on historical sources. Some of the earliest alliances documented in the historical record pertain to the so-called Amarna period in the 2nd millennium BC, involving Egypt, the Hittite Empire, and Assyria. Today, thanks to the increasing availability of empirical data on alliances, it has been possible to trace the international structure of alliance networks since 1815.

4.9 Software for SNA

When social network analysis was invented in the 1930s by Jacob Moreno and his contemporaries, computers did not yet exist. Even until a few decades ago, most researchers had limited access to computing resources necessary for manipulating large matrices—a much-needed facility in social network analysis, as we have seen in this chapter. It wasn't until a few years ago that computational social network analysis became practical for matrices of meaningful size. For example, computational social network analysis of small groups of size up to, say, a dozen or so members (like a team), has been feasible since the 1960s. However, social networks with hundreds or thousands of nodes, as they occur in many domains across the social sciences (for example, in international relations, where just the number of countries in the international system has size in the order of 10^2), were not very tractable. The good news is that the situation today has vastly improved because the computational brawn available to computational social scientists is much greater than even just a few years ago.

A critical computational consideration in the theory and practice of social network analysis concerns computation time, data structures, algorithms, and tractability—topics already covered in Chap. 3. While most small social networks are computable in polynomial time, many larger networks are not. Wallis (2000: Chap. 13) provides background and an overview of these issues.

Today, one of the most widely utilized software packages for social network analysis is **UCINET** (Borgatti et al. 2002), which was developed at the University of California-Irvine. It comes complete with useful tutorials and a large and growing users' group with many international members, and is a system recommended for social network analysis for up to approximately 5,000 nodes. Moreover, UCINET is well illustrated in several textbooks on social network analysis, including *Analyzing Social Networks* (Borgatti, Everett, and Johnson 2013) as well as other monographs and textbooks.

Pajek software, winner of the 2013 W. Richards, Jr. Software Award of the International Network for Social Network Analysis (INSNA), is another commonly used SNA software program. Pajek is also free and has an online wiki (URL: pajek.imfm.si).[13] Along with UCINET, Pajek is frequently featured in leading social network analysis journals, including *Connections* and *Social Networks*, both published by INSNA.

AutoMap, which is Java-based and developed at Carnegie Mellon University, is described as a "text-mining tool that supports the extraction of relational data from texts. [It] distills three types of information: content analysis, semantic networks, [and] ontologically coded networks. In order to do this, a variety of natural language processing/information extraction routines is provided (e.g., stemming, parts of speech tagging, named-entity recognition, usage of user-defined ontologies, reduction and normalization, anaphora resolution, email data analysis, feature identification, entropy computation, reading and writing from and to default or user-specified databases)" (Carley 2013).

ORA, another system from CMU designed for dynamic network analysis, is described as "a dynamic meta-network assessment and analysis tool containing hundreds of social networks, dynamic network metrics, trail metrics, procedures for grouping nodes, identifying local patterns, comparing and contrasting networks, groups, and individuals from a dynamic meta-network perspective. ORA has been used to examine how networks change through space and time, contains procedures for moving back and forth between trail data (e.g. who was where when) and network data (who is connected to whom, who is connected to where ...), and has a variety of geo-spatial network metrics, and change detection techniques. ORA can handle multi-mode, multi-plex, multi-level networks. It can identify key players, groups and vulnerabilities, model network changes over time, and perform COA analysis. It has been tested with large networks. Distance-based, algorithmic, and statistical procedures for comparing and contrasting networks are part of this toolkit" (Carley et al. 2013).

NodeXL by Microsoft is a *free* computational tool based on solid SNA foundations. Useful for learning with social media data, such as Twitter and Flickr, but can be used for analyzing and visualizing any network dataset. The companion book by Hansen et al. (2011) is a must and very well-prepared.

[13] "Pajek" means spider in Slovenian, referring to the web-like metaphor of a social network.

In addition to specialized social network analysis software, there are several other sources of computational tools for social network analysis. For example, Mathematica, R, NetworkX library for Python, Stata, SAS, and SPSS all have social network analysis facilities.

Recommended Readings

R.P. Abelson, Modes of resolution of belief dilemmas. J. Confl. Resolut. **3**(4), 343–352 (1959)

L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, Classes of small-world networks. Proc. Natl. Acad. Sci. USA **97**(21), 11149–11152 (2000)

D.L. Banks, K.M. Carley, Models for network evolution. J. Math. Sociol. **21**(1–2), 173–196 (1996)

P. Bearman, J. Moody, R. Faris, Networks and history. Complexity **8**(1), 61–71 (2003)

S.P. Borgatti, M.G. Everett, J.C. Johnson, *Analyzing social networks* (Los Angeles, Sage, 2013)

S.P. Borgatti, A. Mehra, D.J. Brass, G. Labianca, Network analysis in the social sciences. Science **323**, 892–895 (2009)

R. Breiger, K.M. Carley, P. Pattison (eds.), *Dynamic Social Network Modeling and Simulation: Workshop Summary and Papers* (The National Academies Press, Washington, 2003)

K.M. Carley, Smart agents and organizations of the future, in *The Handbook of New Media*, ed. by L. Lievrouw, S. Livingstone (Sage, Thousand Oaks, 2001)

K.M. Carley, D. Columbus, P. Landwehr, *AutoMap User's Guide 2013*. Technical Report, CMU-ISR-13-105, School of Computer Science, Institute for Software Research, Carnegie Mellon University (2013a)

K.M. Carley, J. Pfeffer, J. Reminga, J. Storrick, D. Columbus *ORA User's Guide 2013*. Technical Report, CMU-ISR-13-108, School of Computer Science, Institute for Software Research, Carnegie Mellon University (2013b)

M. Coscia, F. Giannotti, D. Pedreschi, Towards democratic group detection in complex networks, in *SBP 2012*, ed. by Yang et al. (Springer, Heidelberg, 2012), pp. 105–113

C. Freeman Linton, *The Development of Social Network Analysis: A Study in the Sociology of Science* (Empirical Press, Vancouver, 2004)

D.L. Hansen, B. Shneiderman, M.A. Smith, *Analyzing Social Media Networks with NodeXL: Insights from a Connected World* (Elsevier, Amsterdam, 2011)

Z. Maoz, *Networks of Nations: The Evolution, Structure, and Impact of International Networks, 1816–2001* (Cambridge University Press, Cambridge, 2010)

J. Moody, D. McFarland, S. Bender-deMoll, Dynamic network visualization. Am. J. Sociol. **110**(4), 1206–1241 (2005)

J. Scott, P.J. Carrington (eds.), *The SAGE Handbook of Social Network Analysis* (Los Angeles, Sage, 2013)

M.A. Serrano, M. Boguna, A. Vespignani, Patterns of dominant flows in the world trade web. J. Econ. Interact. Coord. **2**(2), 111–124 (2007)

J. Skvoretz, Complexity theory and models for social networks. Complexity **8**(1), 47–55 (2003)

S. Wasserman, K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994)

D.J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton University Press, Princeton, 1999)

D.J. Watts, The new science of networks. Annu. Rev. Sociol. **30**, 243–270 (2004)

Social Complexity I: Origins and Measurement 5

5.1 Introduction and Motivation

What is social complexity? How did it originate in human societies thousands of years ago? How is social complexity measured? How is the emergence of complexity detected in a previously simple society? What do we know about the long-term evolution of social complexity? What does current knowledge about social complexity tell us about the likely features or plausible trajectory of future trends? This chapter covers both the “Cosmology” or “Big Historical Picture” of social complexity, as well as underlying foundations in CSS. It introduces facts, methods and theories about social emergence and subsequent dynamics, starting with the simplest social systems that originated in early antiquity and their long-term evolution. The chapter leverages materials from previous chapters, showing how ideas learned in previous chapters are essential for a deeper understanding of how social systems operate and can be modeled computationally.

There are concepts, measurement methods, and theoretical models of social complexity in early, contemporary, and future societies. Accordingly, this generates something like a 3×3 matrix of topics. These are presented from a scientific perspective (i.e., the main sections of this chapter) rather than by historical epochs. The chapter ends with an overview of measurement, which leads to more formal approaches to description (laws) and explanation (theory) in the next chapters.

5.2 History and First Pioneers

The first extant systematic study of social complexity was arguably the one by Greek philosopher Aristotle, who conducted the first comparative research on what we would now call “critical phase transitions” between different regimes of government (which he called *stable* and *degenerative* forms) in three types of political

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_5,

systems:

$$Monarchy \rightsquigarrow Tyranny \tag{5.1}$$
$$Aristocracy \rightsquigarrow Oligarchy \tag{5.2}$$
$$Democracy \rightsquigarrow Ochlocracy, \tag{5.3}$$

where the symbol "$\rightsquigarrow$" denotes decay.

The modern roots of the scientific study of social complexity date to the time of the French Enlightenment, as do so many other areas of systematic social science research. In this case the history and pioneers of social complexity origins and measurement are intertwined through developments across political science, anthropology, and computational science. Moreover, many milestones are relatively recent, since the core concept of social complexity became a focus of scientific investigation in large part during the past half-century. The following pertain to origins and measurement of social complexity. (Laws and theories are discussed in the next two chapters.)

18th century Archaeologists begin uncovering material evidence of early social complexity through excavations in Asia and elsewhere.

1944 Anthropologist Bronislaw Malinowski publishes his classic, *A Scientific Theory of Culture and Other Essays*, where he conceptualizes human institutions as instrumental in achieving basic human needs.

1952–1958 Archaeologist Kathleen Kenyon excavates the ancient neolithic and walled settlement of Jericho, Palestine, dating it to ca. 7000 B.C.; it is still among the earliest known sites of primary social complexity.

1962 Social scientist Elman R. Service publishes his influential monograph on *Primitive Social Organization* with the ordinal-level scale of rank values of tribe-band-chiefdom-state that is still in common use today.

1968 Anthropologist Lewis L. Binford publishes his influential paper on "Post-pleistocene Adaptations."

1972 Anthropological archaeologist Kent V. Flannery of the University of Michigan publishes his influential paper on the cultural evolution of civilizations.

1973 Political scientist Giovanni Sartori of the University of Florence publishes his paper on "What Is 'Politics'" in the inaugural issue of the journal *Political Theory*.

1989 Anthropological archaeologist Timothy Earle of Northwestern University publishes his paper on the evolution of chiefdoms in *Current Anthropology*, followed by other influential work on the theory of chiefdoms during the 1990s (1991, 1997).

1994 Archaeologist Henry Wright of the University of Michigan publishes his influential paper on pre-state political formations.

1995 Douglas T. Price and Anne Birgitte Gebauer publish *Last Hunters—First Farmers*, a highly influential collection of papers on the emergence of agriculture and social complexity, including the important paper by Patty Jo Watson.

1995 The same year Smithsonian scholar Bruce D. Smith publishes his classic monograph on *The Emergence of Agriculture*.

1996 Political scientists Yale H. Ferguson and Richard Mansbach propose the concepts of vertical and horizontal polities in *Polities: Authority, Identities, and Change*, a conceptual innovation for understanding complex societies and political systems.

1997 Archaeologist Joe W. Saunders and collaborators publish their paper on initial social complexity at the site of Watson Break, Louisiana, the oldest mound complex in North America, dated to the 4th millennium B.C., in the journal *Science*.

1998 Archaeologists Gary Feinman and Joyce Marcus publish their influential edited volume on *Archaic States*, including the first comparative, cross-cultural analysis of Marcus' "Dynamic Cycles Model" of chiefdoms, and other important papers on early social complexity.

2001 Oxford historian Felipe Fernández-Armesto publishes his comprehensive monograph on *Civilizations*, a descriptive world history in remarkable harmony with Simon's computational theory of social complexity through adaptation to challenging environments in ecosystems.

2001 The earliest origins of primary social complexity in South America are dated to the late 3rd millennium B.C. at Aspero and Caral, in the Supe River Valley, a short distance north of Lima in present-day Peru.

2005 Computational social scientists and other scholars hold the first international conference on sociogenesis in St. Petersburg, Russia, inviting mathematicians, computer scientists, historians, and social scientists from the various disciplines.

This braided history of social complexity science demonstrates how diverse disciplinary strands have finally begun to interact in more systematic fashion only in recent years. The main result of this process is that today there exists a critical mass of facts and measurement methodologies for conducting research on social complexity, including specific scientific knowledge about origins thousands of years ago in a few and quite special regions of the world. Modeling and theoretical milestones are highlighted in the next two chapters.

5.3 Origins and Evolution of Social Complexity

The primary purpose of this section is to provide an empirical, factual base to learn about the precise geographic locations and specific historical epochs—i.e., the space-time coordinates—of social complexity origins within the broader context of global history. This brief long-range survey has intrinsic value in addition to providing foundations for better appreciating the significance of concepts, measurements, models, and theories presented later in this chapter. A long-range perspective is also needed for understanding the substantive, interdisciplinary, and methodological demands on CSS theories and research on social complexity.

When, where, and how did social complexity originate in the global history of human societies? For now, by **social complexity** we mean simply the extent to which a society is governed through non-kin-based relations of authority. In simple, precomplex societies (e.g., in hunter-gatherer groups before the invention of agriculture) individuals are governed by kin-based authority, such as the older member of a household. At the other extreme of social complexity, a modern democracy is governed by elected officials who exercise authority through the executive power of large state bureaucracies comprised of government agencies and specialized government workers. This initial definition of social complexity, based on **relations of authority**, is sufficient for now. Later we will use a more precise definition.

As we shall see later in this chapter, the **chiefdom** represents the simplest form of complex society, one that is governed by rulers who derive their authority from a source that is different from family ties (although the latter never quite disappear entirely from the scene).[1] Hence, the previous, general, and more abstract questions concerning social complexity origins now translate into the more specific, and hence more scientifically tractable, quest for the origins of the earliest chiefdoms.

The **Service scale** is named after American anthropologist Elman R. Service, who was the first to propose the following ordinal-level scale of social complexity:

$$\textit{band} \prec \textit{tribe} \prec \textit{chiefdom} \prec \textit{state} \prec \textit{empire}, \tag{5.4}$$

where the symbol $\prec$ denotes an ordinal relation on ranked values of social complexity.[2] The Service scale of social complexity in expression (5.4) is extended to empires, which are polities that display significantly greater social complexity than states. We shall examine this scale and others more closely later in this chapter.

Specifically, we are most interested in those chiefdoms that eventually developed into states. By **state**, for now, we mean a polity more developed than a chiefdom, in the sense that (1) authority relations are sanctioned by institutions and (2) government operates through a system of public administration that carries out specialized functions. (Later we will also examine the concept of **empire** as a polity that is significantly more complex than a mere state).

5.3.1 Sociogenesis: The "Big Four" Primary Polity Networks

The earliest developmental stage of social complexity—what is often called " primary" social complexity—consists of the formation of the earliest polities or "chiefdoms," a major social milestone that occurred after the great Ice Age in their most simple form approximately 10,000 years ago (the early Holocene Period) in both northern and southern hemispheres. These early polities were not yet "states," but rather societies that departed from egalitarian norms in public activities (e.g., in communal worship, warfare, and major monumental works, among others) through non-kin relations of authority. As a consequence, a chiefdom polity is also *centralized* in the person of a *paramount* leader, chief, or strongman (an individual who is

---

[1] A more formal definition of "chiefdom" and "state" is provided later in Sect. 5.4.

[2] This is the same notation used to denote preferences, since they too are usually expressed on an ordinal-level scale. In LaTeX, these are written as backslash-prec for $\prec$ and backslash-succ for $\succ$. Symbols such as greater than or less than should be avoided for ordinal relations, because they imply interval- and ratio-levels of measurement.

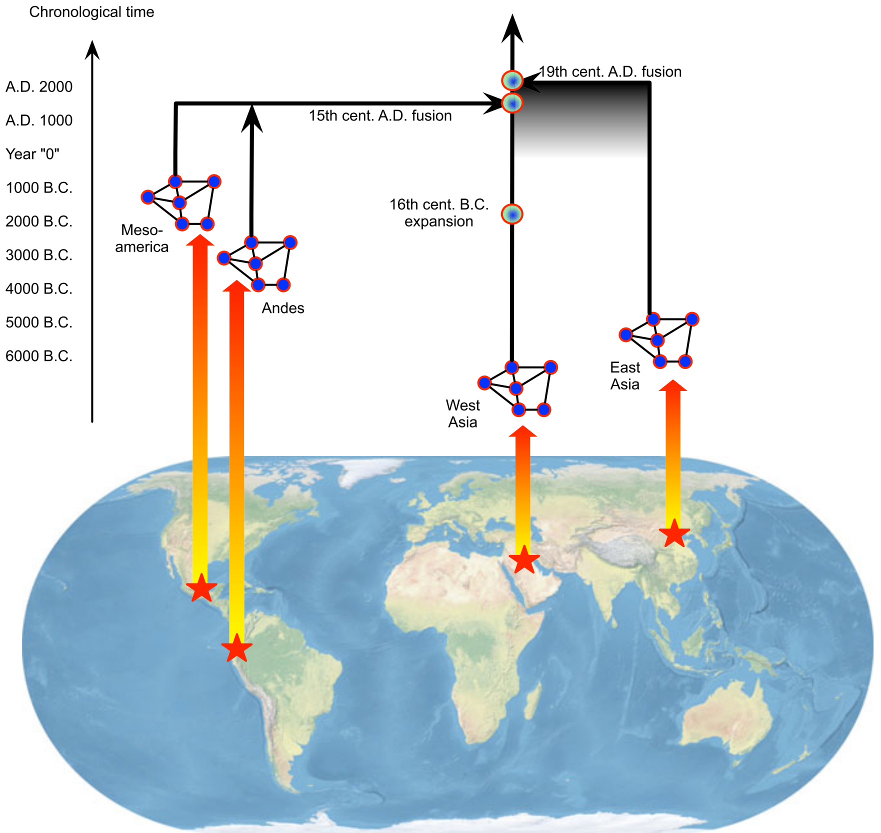


Fig. 5.1 Global geo-chronology of origins of social complexity in the four "cradles of civilization." *Source*: Adapted from Cioffi-Revilla (2006)

primus inter pares, or first among equals, relative to other local leaders); governance is *hierarchically* organized (the leader commanding local sub-leaders or confederates); and it has a *ranked* social order (the family of a leader, whether paramount and confederate, being more important and richer than a commoner family). A chiefdom is an intermediary society between an egalitarian simple society and a state. Therefore, the formation of a chiefdom in a region previously populated by a set of simple egalitarian societies marks a distinctive **phase transition** on the Service scale, and understanding the origins of social complexity—that is to say, when, where, and how the simplest chiefdoms emerged for the first time in human history—is fundamental for understanding not just the origin but also the evolution of complex societies.

Complex societies originated in four separate regions of the world thousands of years ago, during the early Neolithic Period, as summarized in Fig. 5.1. In each regional case a set of local polities generated the first regional interaction network for that part of the world. The description of each region of original social complexity—

based on the evidence currently available for each case (which will certainly increase due to current and future archaeological research!)—is described in terms of first-generation chiefdoms, which were the earliest polities to appear, followed by first-generation states, in chronological order by region. Numerous other states and empires later followed in these regions during subsequent epochs.

How do we know all this? Or, more specifically, how were these determinations of space and time in the initial social complexity of each region, and globally, arrived at in the first place? We will answer questions like these in the next section when we examine the measurement of social complexity from a methodological perspective.

5.3.1.1 West Asia

The earliest chiefdoms in human history formed in the ancient near East (Mesopotamia and the Levant), in the region presently occupied by the countries of Iraq, Israel, Palestinian Territories, Jordan, Iran, Lebanon, Syria, and Turkey—the region known as the Levantine Fertile Crescent. This occurred about 8,000 years ago (8 kya),[3] or by the middle of the sixth millennium B.C.. Early polities centered at Jericho, Çatal Hüyük, and other Neolithic sites in this region are among the oldest extant manifestations of social complexity or individual chiefdom-level polities. Although the Pre-pottery Neolithic-B (PPNB) polity of Jericho (7500 B.C.) once stood in relatively temporal isolation from the earliest West Asian chiefdoms of the 'Ubaid period (5500–4000 B.C.), archaeological investigations have uncovered other pre-'Ubaid polities chronologically situated between PPNB-Jericho and 'Ubaid. Umm Dabaghiya (Iraq) and Ain Ghazal (Jordan) are two examples. Therefore, it is quite possible that the antiquity of the West Asian system of regional polities may some day be pushed back to ca. 7000 B.C., or almost two thousand years earlier than the current dating.

The earliest West Asian system of polities formed between ca. 5800 and 4000 B.C., or during the 'Ubaid period, and consisted exclusively of chiefdoms involved in trade, warfare, and other regional interaction relations. Eridu, Ur, Uruk, Kish, Umma, and Haggi Muhammad were among the most important chiefdoms in Lower Mesopotamia, with Susa [Sush in Persian], Boneh Fazili, Choga Mish, and Farukhabad to the East, and Brak, Gawra, Hacilar, Gian Hasan, and Mersin to the north and northwest.

The first true inter-state system formed in Lower Mesopotamia by ca. 3700 B.C. (Middle Uruk period). Although the exact complete composition of this pristine inter-state system is still unknown, some of the most important states were Uruk and its neighbors in Lower Mesopotamia (Rothman 2001; Algaze 2008); Mish, Susa, and Fanduweh in the eastern regions (present-day Iran); and a number of Anatolian states to the northwest (present-day Turkey).

5.3.1.2 East Asia

The second original polity system emerged in East Asia after ca. 7000 kya, approximately 1,000 years after the formation of the West Asian polity system in the Fertile Crescent. This system emerged pristine, not by any known direct process of diffu-

[3]The acronym "kya" has the standard meaning of "thousands of years ago."

sion from West Asia (*ex nihilo*). This hypothesis might change, as investigations uncover previously unknown links between West and East Asia, but for now we continue to assume socially disjoint separation between the two Asian polity networks. Whereas the traditional Chinese paradigm (Han ideology)—based largely on Confucian culture—has been to view the origins of social complexity in East Asia as centered solely in the Yellow River basin, this belief has now been proven to be a misconception. Today, archaeological investigations are documenting the origins of the East Asian polity system in a multitude of regions across China, not just in the traditional Han homeland. Future investigation will no doubt further clarify the social complexity landscape and show a multi-cultural spectrum of societies at the dawn of East Asian history, perhaps a more diverse social landscape than the spectrum of societies that generated the earlier West Asian system a thousand years earlier.

The first East Asian polity system probably formed over a large area during the Early Banpo to Yangshao and Dawenkow periods (ca. 5000–3000 B.C.), among chiefdoms such as Banpo, Chengzi, Jiangzhai, Dawenkou, Daxi, Hutougou and other Hongshan chiefdoms (4500–3000 B.C.). During the subsequent Longshan period (3000–2000 B.C.) the East Asian polity system already consisted of numerous chiefdoms scattered across a vast area in virtually all regions of present-day China—not just the north.

The Erlitou period (ca. 2000–1500 B.C.) and early Shang period (1900–1700 B.C.) witnessed the emergence of the first interstate system in East Asia, with a core area comprising the polities of Xia (capital at Erlitou) and Shang (capital at Xiaqiyuan), as well as other states that emerged soon after nearby. Traditionally, this is when the Xia dynasty is supposed to have ruled, but today the evidence for these polities is established by anthropological and dirt archaeology, not by epigraphy alone, as we shall examine in the next section. In addition to the state of Shang and the state of Xia, other states also formed, probably at Panlongcheng (Hubei) and Suixian (Wuhan), although the complete system composition is still unknown.

5.3.1.3 South America

The third oldest polity system emerged in South America after 5 kya, or Late Preceramic period, ca. 2500–1800 B.C., and was centered in present-day Peru. A well-known characteristic of this network system is that it functioned for over three-thousand years without a written language, which remains a puzzle from a political science perspective. Another remarkable feature of the South American social complexity is the highly constrained natural environment in which it emerged and evolved for thousands of years, specifically its north-south linear form, in contrast to the more diversified natural environments of the other three original polity regions.

The first phase of South American social complexity took place with the emergence of interacting chiefly polities located up and down the Peruvian coastal regions irrigated by numerous mountain valleys and river basins draining from the Andes: Aspero (Supe river drainage, 2700 B.C.), El Paraíso (near Lima 2000 B.C.), La Galgada (Santa river basin, 2400–1900 B.C.), Río Seco, Salinas de Chao, and other polities.

According to most Andean specialists the first state in the South American region—Moché or Mochica—emerged in the first centuries B.C. from this landscape of warring chiefdoms. However, the material and cultural influence of the much earlier Chavín de Huántar polity (900–300 B.C.) could support an alternative hypothesis that Chavín—earlier than Moché—may have been the first state of the Andean region, given additional evidence besides its own monumentality, as we shall examine later.

The first true interstate system in South America probably emerged after the fall of the Moché state (ca. A.D. 600, after the Middle Horizon period), when two powerful contemporary states emerged—Wari in the north (centered in the Peruvian highlands) and Tiwanaku in the south (centered in northern Bolivia)—and competed for primacy. This was also the first bipolar system of the Western Hemisphere. Both Wari and Tiwanaku were extensive territorial states governed from large capitals and powerful provincial administrative centers.

5.3.1.4 Mesoamerica

Last but not least, Mesoamerican social complexity occurred most recently, having emerged only approximately three thousand years ago, perhaps 3.5 kya. Similar to the oldest polity system in the Old World—the West Asian world system—Mesoamerican social complexity also had a highly diversified set of cultural origins: Olmec, Zapotec, Maya, and other major early Amerindian cultures that shared some common attributes but also differed in significant respects. Another commonality with both Old World primary systems—West Asia and East Asia—lies in the variety of ecotopes (natural environments) in which the Mesoamerican polity system originated and subsequently evolved.

The earliest Mesoamerican polity network that formed was arguably among Olmec chiefdoms, such as those centered at La Venta, San Lorenzo, and others nearby, but regional clusters of chiefdoms developed early in the Zapotec and Maya areas as well. In fact, prior to the emergence of a true interstate system, Mesoamerica was politically organized into chiefdom clusters or subgraphs of chiefdoms with weak links among clusters. Calakmul and El Mirador provide examples in the Maya area; San José Mogote and other Zapotec chiefdoms are examples in the Oaxaca Valley.

The earliest Mesoamerican state probably formed in the Valley of Oaxaca—the Zapotec state, ca. A.D. 400—and had its capital at Monte Albán. On a much larger regional scale, the first interstate system of Mesoamerica was formed by no later than the Late Formative period, and consisted of the Zapotec state, the state of Teotihuacan to the northwest, and the cluster of powerful Maya states to the southeast. After ca. A.D. 500, the composition of this system included Tula in the Mexican central highlands, El Tajín in the Gulf of Mexico, and the post-Classic Maya states in the Yucatán Peninsula. The polity of Teotihuacán may have been an empire during the period A.D. 200–600, with colonial outposts as far south as Kaminaljuyú in present-day Guatemala City (reminiscent of Uruk's Tell Brak in Mesopotamia) and possibly others.

5.3.2 Social Complexity Elsewhere: Secondary Polity Networks

In other regions of the ancient world besides the four original ones we have just discussed—in Africa, Europe, North America, and Oceania—systems or networks of polities also developed. However, such systems were not pristine and persistent in terms of having produced original social complexity extending to large-scale imperial complexity. For example, the Indus Valley region gave rise to the polities of Harappa, Mohenjo-daro, and others in the same region, but most likely these polities were inspired by or at least influenced by the much earlier and powerful polities of West Asia, in Mesopotamia, and in the Levant. Similarly, the network of Egyptian polities in the Nile Valley was also influenced by earlier and more complex developments in Mesopotamia and the Levant. Both cases—the Indus Valley polities and the Nile Valley polities—were linked by trade networks (and possibly migration as well) to the pre-existing West Asian polity network.

In Africa (excluding Egypt) the emergence of social complexity came much later, perhaps as late as the 11th century A.D. during the late Iron Age. In Europe, chiefdoms formed earlier, but they formed states much later than in the near East, as in Greece and Italy and elsewhere, or they were conquered by nearby Asian polities.

Social complexity also originated in North America, but only after A.D. 600. The most complex polities before the European invasion and conquest were centered at Chaco Canyon (New Mexico) and Cahokia (Illinois). The scientific consensus today is that both were chiefdoms, not states. A **complex chiefdom** is a term that would best describe them, because they may have been at the threshold of the phase transition to statehood. The history of the two largest and most complex North American polities overlapped chronologically, but there's no evidence of contact between them. Both had declined by the time of the arrival of the Europeans in their former territories. We shall return to these later, after some further ideas that are necessary to appreciate their great significance from a CSS perspective.

5.3.3 Contemporary Social Complexity: Globalization

How do we arrive at the state of contemporary social complexity in the global system from the four original regional networks that we have just examined? In terms of social complexity, most of the history between those early origins and the present consists of second-generation polities, both chiefdoms and states, as well as empires, which we shall examine later.

Globalization, defined as a significant and relatively rapid increase in the size (network diameter) and connectivity of a world system of polities, is an ancient social complexity phenomenon that began thousands of years ago, not a recent or unprecedented occurrence that is unique to modern history. In a certain sense, globalization began in conjunction with the very origins of social complexity, because each of the four primary polity systems began to globalize almost as soon as it originated.

Two quantitatively and qualitatively distinct classes of globalization events are observable in world history. **Endogenous globalization** occurs as a process of growth or expansion that takes place within a given polity region (e.g., the expansion of the Uruk polity in Mesopotamia, Rome in the Mediterranean basin, or Chaco in the American Southwest), while **exogenous globalization** occurs between geographically distant polity network systems that had been previously disjoint as isolated subgraphs (e.g., the 16th century A.D. merging of Eurasian, South American, and Mesoamerican world systems during the European expansion to the Western Hemisphere).

As shown by the evolutionary model in Fig. 5.1, four disjoint and distinct politico-military polity network systems were evolving in parallel—i.e., each of these systems was oblivious of the other since the time that each had originated—around the end of the third century B.C. By this time, several episodes of endogenous globalization had occurred in world history, as we have just seen. By contrast, there have been only two events of exogenous globalization in world history.

The first true episode of exogenous globalization began with the emergence of the Silk Road, which for the first time linked the already vast Euro-Afro-West Asian world system with the equally vast East Asian system by 200 B.C. This new large-scale network of interacting polities was unprecedented, creating an Afro- Eurasian world system in the Eastern Hemisphere and unleashing a set of social and environmental transformations with aftershocks that are still reverberating in today's world system. The formation of the Silk Road and its subsequent development was by no means a linear or uniform process, because it experienced several phases of growth and decline, but its significance cannot be overstated in terms of having caused the first truly massive collapse of world systems—in this case the merging of the Euro-Afro-West Asian world system and the East Asian world system into a single Eastern Hemisphere world system. Thus, only three of the original four truly autonomous world systems remained after the rise of the Silk Road.

The second and last exogenous globalization event occurred when the Euro-Afro-Asian (or eastern hemispheric) world system became linked by politico-military conquest and commercial expansion with the two separate world systems of the Western Hemisphere, around 500 years ago. This time the fusion or catalytic event was the European conquest of the Americas, an event in important ways systemically analogous to the emergence of the Silk Road more than a thousand years earlier. This time the fusion was even greater than it had been with the emergence of the Euro-Afro-Asian world system (which collapsed two systems into one), since this time a single and truly global world system emerged from the previous three that had existed in isolation.

After A.D. 1600 the global world system has greatly increased its connectedness and further reduced its connectivity diameter—down to the "small world" compact structure observable today; no further exogenous globalization is possible. The contemporary world in which we live today consists of a vast, relatively compact or dense network of socially complex units, which range in scale from tiny countries to huge superpowers linked by governmental and non-governmental international and transnational organizations. The recent emergence of networks of international organizations is especially significant from a social complexity perspective, because it indicates that global society has begun to produce structures of governance that exercise some degree of authority and policymaking activity beyond the state level—especially since their dismantling is increasingly unthinkable. Viewed from this long-range perspective, the contemporary global system could either (1) endure in its present level of social complexity (with a hybrid ecology of states and international and transnational organizations, as it has during the past 200 years); (2) continue to grow towards the emergence of world government at some future point (which would mark another major phase transition); or (3) recede toward a prior situation of autonomous nation states linked by relatively weak international organizations that are purely technical and lack any authority—such as, for example, the international system prior to World War I, or before 1914.

5.3.4 Future Social Complexity

The inventor and social philosopher Charles Franklin Kettering [1876–1958] once said that he was interested in the future because he was going to spend the rest of his life there. (He also said that "the whole fun of living is trying to make something better,"[4] which is consistent with the drive to improve quality of life, which generates increasing social complexity.) Future social complexity is uncertain in its details, of course, but its general features are not difficult to sketch out. The best scientific way to predict future social complexity is to understand its causes, based on proven principles informed by data. Based on this approach, the current state of social complexity indicates that human societies will continue to develop artificial systems, both engineered and institutional, to address threatening challenges, exploit opportunities, or enhance our quality of life.

A highly significant feature of contemporary human civilization—from a social complexity perspective—has been the development of the space program, which has been in progress for many decades. The space program is an excellent example of how humans have generated a remarkable array of complex systems and processes within the same logic of strategic adaptation to meet the challenges of space exploration, travel, and eventually colonization away from the earth. The space program that exists today can be considered an embryonic form of **spacefaring civilization**, both in the form of (1) vehicles and their engineered physical facilities that constitute a complex network of infrastructure systems, as well as (2) in the human organizations and institutions that have been decided, planned, and implemented to support space missions. In August, 2012, NASA confirmed that the spacecraft Voyager 1 became the first man-made artifact to reach interstellar space.

A future spacefaring civilization is entirely compatible with the history of human social complexity, as we will see in greater detail following the examination of some additional concepts and theories that are necessary in order to assess its plausibility.

[4] As quoted in *Dynamic Work Simplification* (1971: 12), by W. Clements Zinck.

However, the incipient spacefaring civilization that we already have today displays a large number of features related to social complexity.

1. Computation and information-processing not only play a major role in the current space program but also provide critical infrastructure for maintaining and enhancing performance.
2. Highly complex artifacts, such as space vehicles (capsules, shuttles, and stations), have enabled the performance of human activities of unprecedented complexity in environments with extreme by hostile physical conditions for humans. Such conditions include the vacuum of space, exposure to intense solar radiation, and small and large asteroids while in orbit, in addition to re-entry and landing failures, among the most common lethal hazards.
3. Societal dependence on an increasingly complex and vast array of space-based systems (both orbiting and geostationary systems of systems), ranging from GPS to highly sophisticated remote-sensing satellites, among others, is arguably irreversible. All critical infrastructure systems in the majority of countries in the world now rely on essential links to space assets.
4. A space-based economic sector is already in its formative stage, with examples such as commercial weather satellites, private navigational systems that support surface and airline travel, soon to be followed by other economic activities already making the headlines.
5. Numerous and unprecedented challenges in design, implementation, management, and integration of complex human organizations and technical systems (i.e., coupled socio-techno systems) have been overcome, and there is no indication— at least not judging from all relevant evidence from university training programs, the manufacture of vehicles and systems, professional conferences and associations—that such a trend will end anytime soon.

The dependence of contemporary civilization on spaced-based systems today may be quite unobtrusive—and it is admittedly so for most members of society, concerned as they are with issues in everyday life—*but from a scientific point of view that does not make it less real.* Solar flares and electromagnetic storms are also real, and space weather has major effects on our planet. These and other indicators do not seem easily reversible patterns, barring some extreme, catastrophic event. Even the threat of major hazards posed by such catastrophic events, such as near-Earth objects and asteroids, provide, further impetus toward a spacefaring civilization by generating new programs, economic growth, and international collaboration, under at least some imaginable set of reasonable conditions. Understanding future social complexity, with or without a spacefaring civilization, requires further development in our conceptual, methodological, and theoretical foundations.

5.4 Conceptual Foundations

In this section we take a closer look at key concepts in the study of social complexity in ways that are more specific than discussed so far. Several of these have already been introduced, but require more powerful definition, while others are new and introduced here for the first time.

5.4.1 What Is Social Complexity?

Earlier we introduced the concept of **social complexity** in the context of Simon's theory, which applies universally to societies both ancient and contemporary, and more recently discussed it in our survey of how the first sociopolitical systems formed in early human history (sociogenesis), based on the Service scale—specifically as *the extent to which a society is governed through non-kin-based relations of authority*.

These ideas already suggest basic features of social complexity that merit highlighting:

Goal-seeking behavior: Humans are goal-seeking actors, not purely passive agents.

Basic goals sought: Basic goals sought by humans, and society as a whole, include **survival** and **improvement**. The former includes meeting existential challenges while the latter refers to the human desire to improve one's quality of life, if not for oneself then for one's kin, friends, or descendants. Both goals are universal cross-cultural drives.

Adaptation: Goal-seeking behavior generally requires adaptation, because individual and collective environments in which humans are situated can be challenging or shifting. Quite commonly the goals being sought are pursued in difficult environments or adverse circumstances.

Artifacts: Implementing adaptive behavior requires the activities of planning and constructing artifacts which, as we have already discussed, can be tangible or intangible, generally corresponding to engineered and organizational systems, respectively.

Polity: The complexity of a society is expressed by its polity and economy, which represent the way it is governed and sustained.

Ordinal scale of social complexity C: Let $a(C) \prec b(C)$ denote an ordinal relation defined with respect to social complexity C, such that the complexity of b is greater than the complexity of a. A society's level of complexity is expressed by the ordinal level of its polity (band/tribe $\prec$ chiefdom $\prec$ state $\prec$ empire $\prec$ world system) and economy (barter $\prec$ monetary), which represent the way it is governed and sustained, respectively. Other ordinal features of social complexity include the authority of leaders (decentralized $\prec$ centralized), territorial control (putative $\prec$ effective), tax extraction ability (null $\prec$ effective), among others.

5.4.2 Defining Features of Social Complexity

We use these basic ingredients of social complexity to understand other facets of this concept. Among these are the fundamental notions of bounded rationality, emergence, near-decomposability, modularity, and hierarchy.

5.4.2.1 Bounded Rationality

Goal-seeking behavior by humans situated in real-world conditions or normal circumstances —i.e., the context where social complexity occurs—is never completely based on rational choices, often not even remotely. Humans make decisions and behave according to what is known as **bounded rationality**. This is best understood by briefly examining the **model of perfect rationality** in terms of its main assumptions when compared to assumptions of human bounded rationality. The basic ingredients of the rational choice model consists of goals, alternatives, outcomes, utilities, and probabilities.

Assumption 1—Goals: *Decision-making goals are clear/precise*. By contrast, humans often have an imprecise understanding of the goals they seek, particularly when deciding under stress.

Assumption 2—Alternatives: *The complete set of available alternatives is known*. Similarly, humans usually have an incomplete understanding of available alternatives. This problem is compounded by numerous circumstances, including the presence of stress, incomplete information, and similar factors.

Assumption 3—Outcomes: *Each alternative entails a set of known outcomes*. The estimation of outcomes that can follow from alternatives is difficult, to say the least, since it involves prediction. This is further compounded by human biases, such as wishful thinking, group-think, and many other well-documented biases.

Assumption 4—Probabilities: *Each outcome occurs with known probability*. Probabilities derive from a mathematical theory, whereas we humans normally employ intuition, which is well-known as a poor guide for estimating true probabilities.

Assumption 5—Expected utilities: *Expected utilities can be computed for each outcome and integrated for each alternative*. Human reasoning is incapable of conducting expected utility computations except in the simplest circumstances or through extraordinary efforts.

Assumption 6—Utility maximization: *The alternative with the highest expected utility is chosen*. By contrast, humans often decide to act by what they feel obligated to do, which may not be in their best interest, or by what their friends appear to be doing, or they choose a course of action through some other principle that may not bring the highest expected utility.

Since the rational choice model is critically dependent on these six stringent assumptions—both individually and as a set, since they are formulated as jointly necessary conditions—perhaps it is not so difficult to understand why the model fails to meet even a mildly realistic test, especially because each assumption is difficult if not impossible to obtain.

Behavioral social science is founded on the bounded rationality model.[5] It is interesting to note that violations of the perfect rationality model occur because humans have **imperfect information** or they experience **faulty information-**

[5] Herbert A. Simon, Daniel Kahnemann, and other social, behavioral, and economic scientists have been recognized for their pioneering work in this area by receiving the Nobel Prize.

processing even when the quality of the information itself may be excellent. Human processing of information— analysis and reasoning—is not fault-free, because it, too, is affected by biases and other cognitive effects. This is another instance in which information-processing is highlighted in CSS, this time specifically in the context of social complexity.

The estimation of outcomes and probabilities, by individual humans and groups, constitutes a large area of research in behavioral science. *Experimental work* in this area has now documented literally scores if not possibly hundreds of **human biases** caused by our incapacity, under common circumstances, to correctly estimate true outcomes and probabilities. Besides wishful thinking and group-think, other biases include referencing and other distortions.

The bounded rationality that is natural in humans also has significant institutional consequences: *humans often create institutions* (i.e., organizational artifacts) *precisely for the purpose of managing or attempting to overcome their faulty rationality*. For example, the purpose of deliberative bodies and agencies in contemporary polities (such as legislative or executive branches of government) is to discuss, discern, and agree on goals, explore alternative options, and conduct assessments of outcomes and probabilities in order to improve cost-benefit analyses that support policymaking—from legislation to implementation. Hence, *increased social complexity through creation of institutions and procedures, often in the form of large bureaucracies, is explained by social complexity theory as simply an adaptation strategy for coping with our innate lack of perfect rationality.* In other words, social institutions are causally explained by bounded rationality. Institutional growth and development is also a major occurrence of "emergent" phenomena.

5.4.2.2 Emergence

The term emergence denotes the processes whereby aggregate, macroscopic phenomena result from individual, microscopic behaviors. The study of social complexity comprises many forms of emergence. Social complexity itself is an emergent phenomenon, because it results from goal-seeking decisions under bounded rationality conditions and adaptive behaviors on the part of many individuals or groups. *All artifacts, whether engineered or institutional, are emergent phenomena.* Networks, polities, economies, and culture itself, among many other macroscopic phenomena in the social universe, represent instances of emergence.

An emergent phenomenon is particularly interesting and well-defined when the *aggregation association* among micro-level components is *strong*, in the sense of *composition*, rather than mere aggregation, in an object-oriented sense. (Recall the earlier discussion of the aggregation association in Sect. 2.8.2.1.) This is because in the case of association by composition the component objects or entities are strictly defined in terms of the aggregate, macro-level entity. Instances of this include polities, networks, organizations, social movements, public moods, all forms of collective behavior including the significant class of collective action phenomena, and numerous other significant entities in the study of social complexity. By contrast, simple aggregation is not considered a form of emergence in the strict scientific sense of the term (e.g., a meeting of persons without a collective action outcome is an instance of simple aggregation but not an emergent phenomenon; collective action would turn the meeting into an instance of an emergent phenomenon).

5.4.2.3 Near-Decomposability

The structural organization of social systems and processes is highly significant, because not all structural forms are characteristic of social complexity. For example, a fully connected network may be considered complicated—such as when in a given group everyone is speaking with everyone else—but it is not complex. At the other extreme, a network composed exclusively of singletons is also not complex. Social complexity lies at a specially structured location in between these two extremes, specifically when the organizational structure in question is said to be "near-decomposable." Near-decomposability refers to a system having subsystem components interacting among themselves as in clusters or subgraphs, and interactions among subsystems being relatively weaker or fewer but not negligible. A classic example of a near-decomposable structure is a hierarchical organization that is divided into divisions and department units.

High-level descriptions of social systems and processes often conceal near-decomposability in their social complexity. For example the near-decomposability of a polity system is not revealed by its first-order composition in terms of a societal component (Society) and a governance subsystem component (Government) interacting for managing Public Issues through Policies. Society and Government are subsystems that compose a polity system, such that Polity is a **system-of-systems**. However, each major component of a polity is, in turn, composed of strongly connected components. Society is composed of individuals, households, and groups that interact among themselves in terms of numerous social relations. Similarly, Government is composed of numerous agencies and entities (e.g., legislative, executive, judicial) that are linked by numerous tightly coupled interactions. Hence, while the first order composition of a Polity does not appear to be decomposable, its second- and higher-order structures, especially those of the operational level, are decomposable.

The property of near-decomposability applies equally to the complexity of social systems *and* processes, not just the former. Accordingly, a process is nearly decomposable when each of its subsequent stages is, in turn, composed of multiple activities. An example of this is the *legislative process* within a given polity, whereby the enactment of law consists of several major stages (such as caucusing, drafting, bargaining, initial voting, reconciliation, final voting), each of which entails numerous other intermediate interactions. *Policy implementation* is another classic example of near- decomposability in social processes, as a policy cascades down from the central administration to local agencies to the point where policy consequences reach individuals and groups that are part of society.

A nearly-decomposable structure is also said to be **modular** or modularized. Therefore, **modularity** or modularization is a defining feature of social complexity. A related feature of modular organizational structure is the presence of **hierarchy** as a characteristic of social complexity. This explains why so many forms of social organization are also hierarchical: chiefdoms, states, and empires, as well as the structure of social relations and bureaucratic institutions that support them vary according to scale, but they are all hierarchical and modular in their organization.

5.5 Measurement of Social Complexity

Social complexity is a **latent variable**, which means that it is a property (i.e., a variable or attribute) that is *measurable* but not directly observable. Although we may not be able to measure social complexity directly, we are certainly able to measure it, assuming we are clever enough to use appropriate **proxy indicators** or empirical, operational measures for recording it. For example, the size of artificial systems that support a given society, such as the size of the bureaucracy (measured, say, by the number of public employees), among other dimensions, is a proxy measure of social complexity. This is also true for the size and sophistication of infrastructure systems, which are highly indicative of social complexity. Latent variables are common throughout the social sciences, not just in CSS and the study of social complexity: social status, literacy, wealth and poverty, inequality, unemployment, socioeconomic development, the size of wars, or something even as seemingly observable and countable as voter turnout, all refer to latent variables that rely upon proxy indicators for purposes of measurement. All theoretical concepts are latent, by definition, since they rely on operational variables or empirical indicators for assessing their values. The Service scale (expression (5.4)) is defined in terms of latent values, because data-based proxies are needed to determine the ordinal-level polity value of a given society on the basis of all available empirical evidence.

Social complexity is measured by means of proxy indicators defined at various Stevens-levels,[6] which can be qualitative (nominal or categorical) and quantitative (ordinal, interval, ratio). In this section we present both types, and later in this chapter others will be added.

5.5.1 Qualitative Indicators: Lines of Evidence

Six important and relatively independent **lines of evidence** are used for detecting and measuring social complexity, especially for detecting original formation in the earliest societies (sociogenesis), although these are also applicable to contemporary society.

Structural: The built environment constitutes structural evidence of social complexity, especially structures intended for collective or public use as opposed to private. Temples, plazas, fortifications (walls, gates, towers, barracks, and other types of military engineering), storehouses, cisterns, irrigation canals and networks, monumental tombs, and palaces are examples used to establish emergence of complexity in the earliest societies. Today, airports, public

[6]The Stevens level of measurement of a given variable refers to whether it is a nominal-, ordinal-, interval-, or ratio-scale variable.

buildings, metropolitan transportation systems, and the coupled network of critical infrastructure systems, are common examples of structural evidence of 21st-century social complexity. Structural evidence is among the strongest signals of social complexity, because it is often large, sometimes massive, and long-lasting.[7]

Pictorial: Imagery depicting leaders, ceremonies, or places of government, and similar visual representations indicative of social complexity, constitute another line of evidence. Court scenes, formal processions, depictions of conquerors and vanquished, portraits of leaders, including those on coins, and heraldry, among others, are diagnostic of initial social complexity. Leaders of ancient polities often used extravagant imagery and exotic pictorial representations of themselves or their allies or territories for propaganda purposes. This is another universal, cross-cultural pattern, not unlike that observed in many modern leaders today. In more modern times, similar evidence persists, in addition to imagery associated with social complexity in a large variety of information media.

Artifactual: Artifacts made by humans are diagnostic of social complexity when their production or technological process requires organization beyond the private, household, or strictly kin-based level. Handmade household pottery for daily utilitarian purposes is not indicative of social complexity; however, an elaborate jade artifact or, even more so, a bronze vessel, are both diagnostic of social complexity. This is because both jade and bronze artifacts require considerable social organization and proven technology in their respective production processes, including specialized knowledge of production, sourcing the appropriate raw materials (minimally copper, tin, and lead in the case of bronze, often from different sources found only at remote locations), specialized workers and facilities (high temperature ovens), warehousing, and a system of accounting. Today, some typical examples of artifacts indicative of contemporary social complexity include computers, cell phones, airplanes, satellites, and other artifacts that, in turn, require hugely complex organizations and supply chains in order to produce them. The global world economy is based on organizational and technological systems with unprecedented complexity.

Epigraphic: Written evidence in the form of many types of documents or inscriptions can provide direct evidence of social complexity. In ancient societies some of the earliest forms of epigraphic evidence was provided by clay tablets written in the cuneiform system of writing for purposes of accounting, teaching, correspondence, and maintaining court records. The Mesopotamian government produced a large quantity of historical chronicles and other epigraphic records. Epigraphic evidence is also abundant in the form of inscriptions on artifacts and buildings, providing **compound evidence** of social complexity. In modern times, history books and a panoply of media, both in print

[7] A classic example of this is the Great Wall of China, but there are also numerous other examples of similar long-lasting structures, such as irrigation canals in ancient Mesopotamia, road networks in Mesoamerica, among others that are only visible through modern satellite imagery and remote sensing.

and electronic form, provide clear examples of epigraphic evidence of social complexity.

Forensic: The condition of human skeletal remains provides another line of evidence for measuring social complexity. In ancient times such practices as cranial deformations, encrustations (such as onyx decoration of the front teeth among the Maya aristocrats of early Mesoamerica), and features of bone tissue indicative of particular diets available only to elites, provide evidence of initial social complexity. In modern times, human remains are relatively less susceptible to forensic analysis that is specifically diagnostic of social complexity.

Locational: Finally, the geographic location of human settlements can be another line of evidence for measuring social complexity. Defensible locations, as on high ground or places with difficult access, are often indicative of widespread warfare, which in turn can imply complex social organization. Numerous chiefdoms and early states were established on such locations, often requiring organizations and infrastructure to render them sustainable. Even in modern times, cities located in inhospitable environments, such as deserts or high mountain regions, require extraordinary complexity in terms of urban support systems.

The level of confidence in the measurement of social complexity is proportional to the number of lines of evidence that provide positive support—the more the better, because the probability of a false positive decreases exponentially with the number of lines that provide evidence of social complexity. A single line of evidence is generally viewed as insufficient, although it may be useful because it suggests that additional lines of evidence may be found. This is because social complexity exhibits numerous manifestations which should be measurable by all available data from multiple lines of evidence, rather than confined to a single source of information.

It should be stressed that lines of evidence for measuring social complexity are relevant not only for establishing initial, formative stages—such as identifying the phase transition from egalitarian to ranked societies in chiefdoms (and later states and empires)—but are also necessary for measuring the complexity of modern societies, such as different levels of social, economic, and political development. There is much more than a simple, nominal difference between advanced and developing societies; the difference can also be quantified in terms of numerous indicators such as critical infrastructure systems, especially when viewed as coupled socio-technological systems.

5.5.2 Quantitative Indicators

We have already been using Service's ordinal-level scale of social complexity, which measures and ranks polities using the ordered values of **chiefdom** (base level) and **state**, to which one can add subsequent ordinal values of **empire** and **global system**. Other quantitative indicators of social complexity include, for instance, the size and structural features of infrastructure present in a given society, since infrastructure is a proxy diagnostic measure of social complexity. The percentage of the population that is not involved in basic subsistence activities (such as individuals involved in education, government, national defense, and a host of others that rely upon that portion of the population not engaged in the production of food and similar basic needs) is increasingly large in advanced, contemporary societies. It too can be considered a proxy measure of social complexity.

Quantitative measures of social complexity can be divided into two broad categories, based on the nature of operational independent variables used to define each measure: formal measures and substantive measures. These should be viewed as heuristic, complementary categories, not necessarily mutually exclusive. They should also be used for comparative purposes.

5.5.2.1 Formal Measures of Social Complexity

Formal measures of social complexity are based on mathematical approaches, such as network-based or graph-based metrics, or information-theoretic measures, among others, all of which use formally defined independent variables. These measures assume that a network matrix is available for computing appropriate indices.

Near-decomposability, a defining feature of social complexity (Sect. 5.4.1), is a latent variable that can be measured by a clustering coefficient proxy. In general, a clustering coefficient measures the number of nodes that are linked by triangles forming subgraphs of various size. Several clustering coefficients have been defined in the context of various near-decomposable structures. The standard **undirected network clustering coefficient** is the average of the clustering coefficient of nodes in an undirected network (such as in an organizational diagram), where the **node clustering coefficient** C_i of node i is defined as

$$C_i = \frac{2\lambda_i}{\delta_i(\delta_i - 1)}, \tag{5.5}$$

where λ_i is the number of connected pairs between all neighbors of node i and δ_i is the degree of i (number of neighbors, defined in Sect. 4.6.1). Therefore, the network clustering coefficient $C_{\mathcal{N}}$ of network $\mathcal{N}$ is given by

$$C_{\mathcal{N}} = \bar{C}_i \tag{5.6}$$
$$= \frac{1}{g}\sum_{i=1}^{g} \frac{2\lambda_i}{\delta_i(\delta_i - 1)}, \tag{5.7}$$

where $g = card(\mathbb{N}) = |\mathbb{N}|$ is the total number of nodes in network $\mathcal{N}$, or the size S of $\mathcal{N}$.

The **Barrat-Weigt clustering coefficient** is defined as

$$C_{BW} = \frac{3(g-1)}{2(2g-1)}(1-p)^3, \tag{5.8}$$

where g is the number of linked neighbors (degree) and p is the probability of rewiring (Barrat and Weigt 2000: 552).

Another quantitative proxy measure of social complexity is **Shannon's entropy** H, which can be measured over the degree of nodes. In this case,

$$H(\delta) = -\sum_{i=1}^{g} P(\delta_i) \log_2\big[P(\delta_i)\big], \tag{5.9}$$

where $P(\delta_i)$ is the probability that node n_i has degree δ. A structure consisting mostly of singletons will have high entropy, and hence not be near-decomposable. At the other extreme, a fully connected graph will have maximum entropy, because the degree distribution will have a single peak given by $\delta = g - 1$. A near-decomposable complex system indicative of clustering and hierarchy will have an intermediate value of entropy somewhere in between.

The *comparative statics* of each of these formal measures of social complexity are interesting, because they are mostly nonlinear functions.

5.5.2.2 Substantive Measures of Social Complexity

By contrast, **substantive measures of social complexity** are based on specific social, economic, political, or other cultural variables. Traditional social science methods can be used to construct proxy measures of social complexity. For example, **multi-dimensional scaling (MDS)** is one such method widely used for comparing scores on multiple indicators that measure dimensions of latent social phenomena. Both classical and nonparametric versions are available in the R programming language. Classical MDS uses Euclidean distances across objects aimed at plotting low dimensional graphs.

The **Peregrine-Ember-Ember ordinal Guttman scale of social complexity** is used for measuring the earliest phase transitions into chiefdoms and states.[8] It contains the following items ranked from minimum to maximum values:

1. Ceramic production
2. Presence of domesticates
3. Sedentarism
4. Inegalitarian (status or wealth) differences
5. Population density > 1 person/mi^2
6. Reliance on food production
7. Villages > 100 persons
8. Metal production
9. Social classes present
10. Towns > 400 persons

[8] The Peregrine-Ember-Ember (2004) scale of social complexity is one of the current Guttman scales developed by anthropologists. It is based on the most comprehensive sample of early human cultures, based on the worldwide *Outline of Archaeological Traditions* from the Human Relations Area Files (HRAF), based at Yale University, and builds on earlier scales of social complexity developed by R.L. Carneiro, L. Freeman, G.P. Murdock, and C. Provost, among others.

11. State (3+ levels of hierarchy)
12. Population density > 25 person/mi^2
13. Wheeled transport
14. Writing of any kind
15. Money of any kind

Chiefdoms form between levels 3 and 7, whereas states form between levels 8 and 11. A defining feature of a Guttman scale is that each ordinal value includes all previous value–traits. For example, villages consisting of 100 or more persons (level 7) also rely on food production (level 6), have population density of more than one person per square mile (level 5), experience marked inequality (level 4), and so forth down to level 1 (ceramic production). Similarly, states consist of towns with more than 400 persons, have social classes and metal production, in addition to traits associated with lower scale values.

For modern polities, the United Nation's **Human Development Index** D_H is a specific example of a proxy measure of social complexity at the country or polity level, designed to assess aggregate socioeconomic conditions (Table 5.1).

The Human Development Index is a composite indicator consisting of three other indices: life expectancy L^*, education level E^*, and national income per capita I^*. These three components are strongly associated with significant levels of social complexity, individually but especially in combination. Simple or primitive societies generally score very low across all three indices. Life expectancy is high in all countries where social complexity is also highest, such as in the advanced industrialized economies. High levels of education are attainable only in societies that can sustain the most expensive universities. High income indicators are similarly observed only in complex societies, where cost of living is also highest. Simple societies measure the lowest scores in lifetime expectancy, level of education, and income-related indices. Formally, D_H is defined as the geometric mean of the three component indicators

$$D_H = \left(L^* \cdot E^* \cdot I^*\right)^{1/3} \tag{5.10}$$

$$= \sqrt[3]{\frac{L - \alpha_1}{\alpha_2} \cdot \frac{\sqrt{S \cdot \langle S \rangle}}{\beta} \cdot \frac{\ln(I/P) - \gamma_1}{\gamma_2 - \gamma_1}}, \tag{5.11}$$

Table 5.1 Social complexity according to the polity-level Human Development Index D_H (2012) in the top fifteen countries. *Source*: United Nations Development Programme, *2013 Human Development Report*

Rank	Country	D_H	Rank	Country	D_H	Rank	Country	D_H
1	Norway	0.955	6	New Zealand	0.919	11	Canada	0.911
2	Australia	0.938	7	Ireland	0.916	12	South Korea	0.909
3	United States	0.937	8	Sweden	0.916	13	Hong Kong	0.906
4	Netherlands	0.921	9	Switzerland	0.913	14	Iceland	0.906
5	Germany	0.920	10	Japan	0.912	15	Denmark	0.901

where independent variables and constants are operationally defined as follows:[9]

$L =$ life expectancy at birth
$S =$ mean years of schooling multiplied by a factor of 1/13.2, or "mean years of schooling index"
$\langle S \rangle =$ expected years of schooling by a factor of 1/20.6, or "expected years of schooling index"
$I =$ gross national income
$P =$ populations
$\alpha_1 =$ 20 years
$\alpha_2 =$ 62.3 years
$\beta =$ 0.951 years^{-1}
$\gamma_1 =$ 100 dollars/inhabitants
$\gamma_2 =$ 107,721 dollars/inhabitants

Several aspects of the human development index are noteworthy as a quantitative measure of social complexity. The geometric mean in Eq. (5.11) defines a *cubic function* for D_H with respect to its three component indices. It also defines D_H as a function of five independent variables and parameters, in terms of *multiple nonlinear dependencies*. Therefore, the comparative statics are interesting also in this case of measuring social complexity. Empirically, all countries in Table 5.1 are also well-known for operating advanced infrastructure systems, which are necessary for adaptation and achieving high quality of life in complex environments.

Numerous measures of complexity have been proposed for generic systems. For example, the minimal description necessary to describe the features of a system (such as an algorithm) can be viewed as a measure of the system's complexity. In the context of a social system's complexity, we can define a **lexical measure of social complexity** based on the length of the minimal description of its functional structure. Rigorous definitions of chiefdoms, states, and contemporary polities, written with minimally necessary and systematic vocabulary, based on comparative social science terminology, provide viable examples. Another operational approach of the same lexical measurement procedure could be based on formal graphic notation, such as UML class, sequence, and state diagrams for describing specific social systems, such as a chiefdom, a state, or a contemporary polity.

Let S denote a social system with complexity $C(\mathsf{S})$. A lexical measure of C can be defined as the minimal number of characters κ, including spaces, that is minimally necessary to describe S. For example, later in Chap. 7 we will examine the formal, theoretically based definitions of a chiefdom and a state. Definition 7.9 (chiefdom) yields $C(\text{chiefdom}) = 289$ characters, whereas Definition 7.10 (state) yields $C(\text{state}) = 339$ characters, consistent with the fact that a state is more complex than a chiefdom.

Different definitions of the same social system S can be expressed in somewhat different number of characters $(\kappa_1, \kappa_2, \kappa_3, \ldots, \kappa_N)$. However, since they are all describing the same system S, only in different words, and all descriptions are assumed

[9] Notation here is different from the original UN annual report, which uses abbreviations and acronyms rather than proper mathematical symbols.

to be minimally necessary, the number of characters can be assumed to be normally distributed. Therefore, the simple arithmetic mean taken over the set of κ_i values provides a composite lexical indicator of social complexity:

$$C(\mathsf{S}) = \sum_{i=1}^{N} \kappa_i. \tag{5.12}$$

Alternatively, if S is defined in terms of graphic models—such as when using a set of associated UML class, sequence, and state diagrams of S—then the set of features contained in the graphics can be used as information to define $C(\mathsf{S})$. For example, suppose the UML class diagram of social system S consists of a number of objects and a number of associations among objects, denoted by discrete variables O and A, respectively, where $O = 1, 2, 3, \ldots, o$ and $A = 1, 2, 3, \ldots, a$. Similarly, the UML sequence diagram of S consists of O objects and S sequential interactions among objects in separate "lanes," where $S = 1, 2, 3, \ldots, s$. Finally, suppose the UML state diagram of S has X states and Φ transitions among states, where $X = 1, 2, 3, \ldots, x$ and $\Phi = 1, 2, 3, \ldots, \phi$. Then, social complexity based on the three graphic models can be defined by functions of these metrics. For instance, the graphic complexity measure

$$C(\mathsf{S}) = (O + A) + (O + S) + (X + \Phi) \tag{5.13}$$

$$= O(A + S) + X + \Phi \tag{5.14}$$

provides a simple but viable aggregate indicator, as do other similar functions defined in terms of graphic features that specify the complexity of social system S. For example, the norm of a vector $\mathbf{C}(\mathsf{S})$ consisting of graphic values in the UML diagrams,

$$|\mathbf{C}(\mathsf{S})| = \sqrt{o^2 + a^2 + s^2 + x^2 + \phi^2}, \tag{5.15}$$

is another viable graphic-based measure of social complexity.

Social complexity is also measurable on a temporal scale, where **long-range correlations** are diagnostic of complexity in social processes. The **Hurst parameter** is a temporal indicator for measuring the complexity of a time series of social data in terms of its **long-range dependence** (LRD). Let $X_1, X_2, X_3, \ldots$ denote a time series of values at times $t_1, t_2, t_3, \ldots$ with mean μ and variance σ^2. The Hurst parameter is defined by the **autocorrelation function** $\rho(k)$ of a time series as

$$\rho(k) = \frac{E(X_t - \mu) \cdot E(X_{t+k} - \mu)}{\sigma^2} \tag{5.16}$$

$$\sim C_\rho |k|^{-2(1-H)}, \tag{5.17}$$

where $|k|$ denotes time lags or leads of length $0, 1, 2, 3, \ldots$ in either direction, the symbol $\sim$ denotes asymptotic equality as $k \to \infty$, and $C_\rho > 0$ is a scale parameter. Note that $\rho(k)$ decays algebraically as a **power law**, so the autocorrelations are

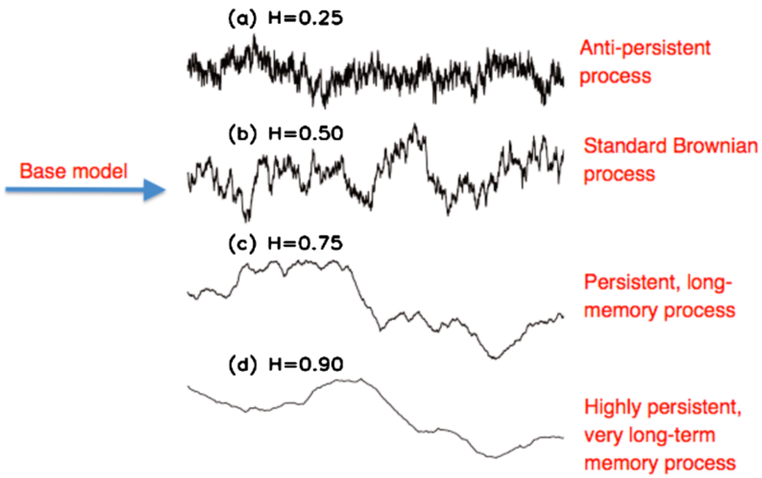


Fig. 5.2 Long-range dependence (LRD) or memory structure in time series measured by the Hurst parameter H. *Source*: Adapted from Gao et al. (2013: 16)

scale-free and, therefore, the process is said to be **self-similar**, that is, **fractal**. We shall examine these properties more closely later, when we focus on power laws of social complexity. **Spatial autocorrelation** is similarly characteristic of social complexity.

The value of the Hurst parameter estimated from empirical data is indicative of process complexity as determined by the following ranges:[10]

Case 1: When $0.5 < H < 1$ the process has long-term memory, or LRD, so the process is also called **persistent**.

Case 2: When $H = 0.5$ the process is standard **Brownian motion** with normal or Gaussian distribution, mean $\mu = 0$, variance $E[(B_H(t))^2] = t^{2H}$, and power spectral density $1/f^{2H+1}$. This is *not* a case indicative of complexity, but rather one of **equilibrium dynamics**.

Case 3: When $0 < H < 0.5$ the process is **anti-persistent**, meaning that it is significantly more jagged than the Gaussian process.

Cases 1 and 3 are driven by **non-equilibrium dynamics** typical of complex systems and processes, as shown in Fig. 5.2. Standard Brownian motion is a base process or **phase transition boundary** (**critical bifurcation value**, $H = 0.5$) for the temporal complexity of a social process. Above the critical value the process has persistent memory ($H > 0.5$), indicative of the status quo or dynamic stability, the process looks increasingly smooth as the autocorrelation length increases, and the distribution of X is heavy-tailed (extreme events have a significant likelihood)). By contrast, below the critical value the process has anti-persistent memory ($H < 0.5$) indicative of high volatility or dynamic instability, and the process looks

---

[10] Many estimators of the Hurst parameter are available, as reviewed by Gao et al. (2007).

more jagged. The "jaggedness" of a time series is inversely related to the Hurst exponent.

If policy is based on assumptions other than those warranted by a time series analysis of the Hurst exponent for temporal complexity, then the provision of public goods will be misguided. The causes of LRD are often difficult to determine. Sometimes it is related to the cumulative effect of prior processes responsible for generating a time series.

Spatio-temporal autocorrelation is diagnostic of social complexity. By contrast, it is noteworthy that traditional data analysis in social science research generally dislikes spatio-temporal autocorrelation, because it violates standard assumptions of correlational analysis of data. The use of various transformations (logarithmic, inverse, square, among others) to obtain "normal" Gaussian-distributed data destroys information necessary for measuring social complexity and should therefore be avoided in social complexity analysis. The same is true for skewed distributions, as we shall see in the next chapter.

Recommended Readings

G. Algaze, *Ancient Mesopotamia at the Dawn of Civilization: The Evolution of an Urban Landscape* (University of Chicago Press, Chicago, 2008)

C. Cioffi-Revilla, The big collapse: a brief cosmology of globalization, in *Globalization and Global History*, ed. by B. Gills, W.R. Thompson (Routledge, London, 2006), pp. 79–95

C. Cioffi-Revilla, D. Lai, War and politics in Ancient China, 2700–722 B.C.: measurement and comparative analysis. J. Confl. Resolut. **39**(3), 467–494 (1995)

C. Cioffi-Revilla, T. Landman, Evolution of Maya polities in the Ancient Mesoamerican system. Int. Stud. Q. **43**(4), 559–598 (1999)

G.M. Feinman, J. Marcus, *Archaic States* (School of American Research Press, Santa Fe, 1998)

K. Flannery, J. Marcus, *The Creation of Inequality: How Our Prehistoric Ancestors Set the Stage for Monarchy, Slavery, and Empire* (Harvard University Press, Cambridge, 2012)

J. Gao, Y. Cao, W.-W. Tung, J. Hu, *Multiscale Analysis of Complex Time Series: Integration of Chaos and Random Fractal Theory, and Beyond* (Wiley-Interscience, Hoboken, 2007)

J. Marcus, Ancient Maya political organization, in *Lowland Maya Civilization in the Eighth Century* A.D., ed. by J.A. Sabloff, J.S. Henderson (Dumbarton Oaks Research Library and Collection, Washington, 1993), pp. 111–183

J. Marcus, K.V. Flannery, *Zapotec Civilization: How Urban Society Evolved in Mexico's Oaxaca Valley* (Thames and Hudson, London, 1996)

J. Marcus, P.R. Williams, *Andean Civilization: A Tribute to Michael E. Moseley* (Cotsen Institute of Archaeology Press, Los Angeles, 2009)

P.N. Peregrine, C.R. Ember, M. Ember, Universal patterns in cultural evolution: empirical analysis using Guttman scaling. Am. Anthropol. **106**(1), 145–149 (2004)

C. Renfrew, P. Bahn, *Archaeology: Theories, Methods, and Practise*, 6th edn. (Thames & Hudson, London, 2012)

R.J. Sharer, A.K. Balkansky, J.H. Burton, G.M. Feinman, K.V. Flannery, D.C. Grove, J. Yaeger, On the logic of archaeological inference: early formative pottery and the evolution of Mesoamerican societies. Latin Am. Antiq. **17**(1), 90–103 (2006)

Social Complexity II: Laws

6

6.1 Introduction and Motivation

In science, laws describe and theories explain. Laws provide understanding of "how" social complexity occurs; theories answer questions of "why" it occurs. Laws are like mappings between variables; theories are causal stories that account for observed social complexity. Which patterns of social complexity have empirical validity as universal laws that hold cross-culturally and across domains of social science research? How is social complexity explained in terms of existing theories?

This chapter develops the analysis of social complexity by presenting theoretical and empirical laws that describe emergence and subsequent dynamics. The main emphasis in this chapter is on formal description for understanding social complexity. The next chapter progresses toward explanatory theories of social complexity. Understanding of basic patterns in laws of social complexity is necessary for developing viable computational models.

6.2 History and First Pioneers

The history of laws of social complexity dates to the early twentieth century, when pioneers such as Vilfredo Pareto, Max O. Lorenz, Corrado Gini, and Felix Auerbach demonstrated the first power laws in human and social domains of science, half a century before power laws entered physics. These early discoveries were soon followed by social power laws discovered by Alfred Lotka, George K. Zipf, Lewis F. Richardson, Herbert A. Simon, and Manus I. Midlarksy. Most recent work on these and other non-equilibrium distributional models focuses on discovering additional domains (e.g., the Internet) as well as replicating earlier discoveries with newly available and better data.

By contrast, research on structural laws of social complexity is more recent, beginning in the Cold War years with the pioneering work of Albert Wohlstetter, William Riker, Martin Landau, Jeffrey L. Pressman, Aaron Wildavsky, Elinor Ostrom, and John W. Kingdon. Research on both types of laws of social complexity is

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_6,

still active and promises new discoveries as CSS researchers expand the domains of universal patterns.

1896 Economist Vilfredo Pareto [1848–1923] pioneers power laws through his comparative research on income and wealth in his classic textbook, *Cours d'economie politique*.

1905 Max Otto Lorenz [1876–1959] publishes his seminal paper on the curve named after him in the *Journal of the American Statistical Association*, while still a doctoral student at the University of Wisconsin.

1912 Sociologist Corrado Gini [1884–1965] proposes his classic coefficient of inequality in *Mutabilitá e Variabilitá*.

1913 Physicist Felix Auerbach [1856–1933] discovers the rank-size law of human settlement sizes, published in *Das Gesetz der Bevölkerungskonzentration* (The Law of Population Concentration), rediscovered years later by Zipf.

1926 Statistician Alfred Lotka [1880–1949] publishes his discovery of the inverse-square law in the "The Frequency Distribution of Scientific Productivity," *Journal of the Washington Academy of Sciences*.

1935 Linguist George Kingsley Zipf [1902–1950] publishes his first papers on the rank-size distribution of settlements.

1941 Meteorologist Lewis Fry Richardson [1881–1953] discovers the scaling power-law of conflicts, inaugurating the modern scientific study of war through a series of papers in 1941, 1945, and 1948. His first monograph dates to 1919, on "The Mathematical Psychology of War."

1955 Herbert A. Simon publishes his classic paper "On a Class of Skew Distributions" in the journal *Biometrika*, followed in 1958 by his first paper on the power-law distribution of business firms in the *American Economic Review*.

1958 Gutenburg-Richter Law for earthquakes is discovered, arguably the first true power law in the physical sciences.

1959 Albert Wohlstetter publishes his classic paper on Deterrence Theory, "The Delicate Balance of Terror," based on the Conjunctive Principle examined in this chapter and the next, in the influential policy journal *Foreign Affairs*.

1960 Richardson's *Statistics of Deadly Quarrels* is published posthumously.

1962 William H. Riker formalizes the Theory of Political Coalitions and demonstrates the Conjunctive Law for minimal-winning coalitions.

1969 Martin Landau explicitly identifies conjunctive redundancy in his seminal paper published in the *Public Administrative Review*, followed in 1972 by his classic *Political Theory and Political Science: Studies in the Methodology of Political Inquiry*.

1973 Jeffrey L. Pressman and Aaron Wildavsky publish the classic *Implementation: How Great Expectations in Washington Are Dashed in Oakland*, based on the Conjunctive Law.

1978 Gabriel Almond and Bingham Powell publish their influential input-output model of a complex polity, where policies in the outcome space follow a sequential conjunctive law.

1984 John W. Kingdon publishes his classic *Agendas*, *Alternatives*, *and Public Policies*, demonstrating the sequential conjunctive law for policy-making processes in complex polities.

1985 Elinor Ostrom [1933–2012] and colleagues from Indiana University (Vincent Ostrom, Roger Parks, Harvey Starr), the University of Illinois (Claudio Cioffi-Revilla, Richard L. Merritt, Robert Muncaster, and Dina A. Zinnes), and the University of Iowa (Robert Boynton) establish the Triple-I Seminar on Complex Systems.

Since 1990 Power laws are replicated in numerous domains of social science research, such as elections, budgetary processes, finance, terrorism, and the Internet.

1999 Cioffi-Revilla discovers that civil wars scale across the global system, demonstrating long-range spatio-temporal correlations.

2003 Economist Christian Kleiber and statistician Samuel Kotz [1930–2010] publish *Statistical Size Distributions in Economics and Actuarial Sciences*, the first comprehensive treatise on the Pareto Law and related distributions of social complexity.

2003 The same year Cioffi-Revilla and Midlarsky demonstrate that a uniform distribution can be critically misjudged as a power law (Type II error) when diagnostic bending in the lower and upper tails is ignored. In the same paper they demonstrate power law scaling for the deadliest wars.

6.3 Laws of Social Complexity: Descriptions

In this section we examine descriptive laws of social complexity. These are grouped into two main categories, structural and distributional, each of which consists of a variety of models. The *comparative statics* of these laws are interesting, because most equations are nonlinear in nature. This often results in non-intuitive or counter-intuitive consequences on the emergent behavior of social complexity. Both share two additional, scientifically deep properties: they are *related* to one another, as well as being *universal* across domains of social complexity.

6.3.1 Structural Laws: Serial, Parallel, and Hybrid Complexity

The **structure of social complexity** refers to the way systems and processes are organized across social domains, including coupled socio-techno-natural systems and components within them, as we have already seen in the case of near-decomposability. Figures 6.1 and 6.2 illustrate isomorphic examples of structural configurations found in social systems and processes, which can often (not always!) be expressed in terms of networks or trees, respectively. A salient feature of structural laws of social complexity is that they have dual isomorphic representation as logic and probabilistic formalism, which facilitates computational modeling. Here we examine more closely the character of causal structures and how they generate emergent social complexity.

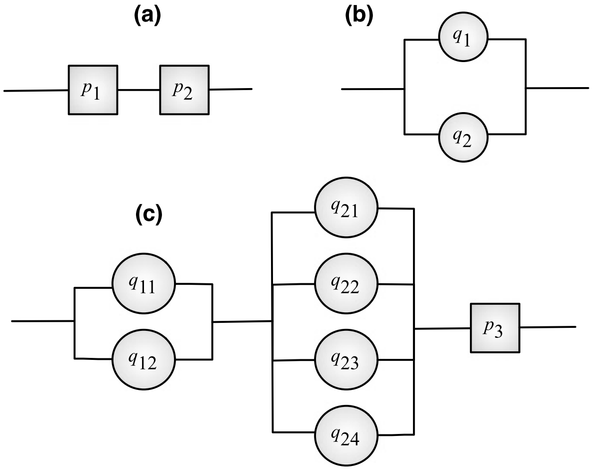


Fig. 6.1 *Structural patterns of social complexity by causal necessity and sufficiency.* (**a**) Serial complexity by causal conjunction; (**b**) parallel complexity by causal disjunction; and (**c**) a case of hybrid serial-parallel complexity with some parallelized disjunctive components within an overall serialized 3-conjunctive structure

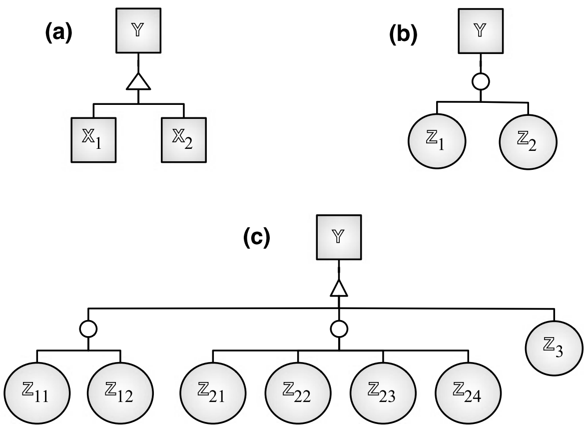


Fig. 6.2 *Structural patterns of social complexity by logic conjunction and disjunction.* (**a**) Serial complexity by causal conjunction; (**b**) parallel complexity by causal disjunction; and (**c**) a case of hybrid serial-parallel complexity with some parallelized disjunctive components within an overall serialized 3-conjunctive structure

6.3.1.1 Serial Complexity by Conjunction

The fundamental structure of complexity in social systems and processes is generated by **compound events**, which emerge from the conjunction of causal events. For example, in the standard model of a polity, the occurrence of successful governance is an emergent compound event generated by a sequential process that begins with (1) an issue collectively affecting a significant sector of society; followed by (2) pressure groups placing demands on government to act; followed by (3) decision-makers doing something to relieve societal stress by enacting policies; and, finally, (4) the public issue being mitigated.

The example just seen is that of a **serial** system (Figs. 6.1(a) and 6.2(a) with 4 components rather than just 2), which is based on *necessary* causal events occurring as a *conjunction* (by Boolean logic AND operator) and emergent overall probability Y_s given by its associated indicator structure function $\Psi_\cap$ according to the following set of related equations:

$$\begin{aligned}\mathbb{Y}_s &= \Psi_\cap(\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \ldots, \mathbb{X}_n) && (6.1)\\ &\Leftarrow \mathbb{X}_1 \wedge \mathbb{X}_2 \wedge \mathbb{X}_3 \wedge \cdots \wedge \mathbb{X}_n && (6.2)\\ Y_s &= p_1 \cdot p_2 \cdot p_3 \cdots p_n = \prod_{i=1}^{n} p_i && (6.3)\\ &= P^\Theta, && (6.4)\end{aligned}$$

where $\mathbb{Y}_s$ denotes the compound event for overall conjunction with necessary causal conditions, $\mathbb{X}_i$ are the n causal events, the symbol $\wedge$ denotes conjunction (Boolean AND), p_i are the probabilities of the causal events, P is their probability when they are all the same, and $\Theta = 1, 2, 3, \ldots, n$ denotes the number of causal events.

An important variation of serial conjunction is when necessary conditions occur in sequence, called **sequential conjunction**, equivalent to Boolean logic SEQAND. Note that probabilities are conditional for sequential causal events. In this case Eqs. (6.1)–(6.4) are simply edited to take into account conditional probabilities, which still require multiplication.

Regardless of whether causal probabilities are conditional or unconditional, overall probability P_s is always *decreased* when social complexity is serialized. **Hypoprobability**, defined by the inequality $Y_s < \min p_i$, is a fundamental property of serial social complexity. It means that serially structured social systems have an overall probability of performing that is smaller than that of the most poorly performing component. Accordingly, the popular aphorism of a chain being as strong as its weakest link ($P = \min p_i$) is objectively wrong, because it overestimates overall serial probability.[1]

6.3.1.2 Parallel Complexity by Disjunction

By contrast, at other times a social system or process may operate according to concurrent activities, as when policy is based on a set of multiple public programs. For example, anti-inflationary policies used by governments are often based on a mix of (1) price controls, (2) subsidies of various kinds (for food, housing, medicines), and (3) other programs that are implemented simultaneously. This example is represented in Figs. 6.1(b) and 6.2(b) with three as opposed to just two causal component events.

This is an example of a **parallel** system, which is based on *sufficient* causal events occurring as a *disjunction* (by Boolean logic OR operator) and emergent overall probability Y_p given by its associated indicator structure function $\Psi_\cup$ and the following set of related equations:

$$\begin{aligned}\mathbb{Y}_p &= \Psi_\cup(\mathbb{Z}_1, \mathbb{Z}_2, \mathbb{Z}_3, \ldots, \mathbb{Z}_m) && (6.5)\\ &\Leftarrow \mathbb{Z}_1 \vee \mathbb{Z}_2 \vee \mathbb{Z}_3 \vee \cdots \vee \mathbb{Z}_m && (6.6)\end{aligned}$$

[1] The correct aphorism should be that a chain is *weaker* than its *weakest* link, which is an even worse condition than being *as weak as* the weakest link.

$$Y_p = 1 - (1 - q_1) \cdot (1 - q_2) \cdot (1 - q_3) \cdots (1 - q_m) = 1 - \prod_{j=1}^{m} (1 - q_j) \quad (6.7)$$

$$= 1 - (1 - Q)^{\Gamma}, \quad (6.8)$$

where notation follows the same conventions as for Eqs. (6.1)–(6.4). By De Morgan's Law, it can be easily demonstrated that parallelization equations (6.5)–(6.8) follow from serialization equations (6.1)–(6.4).

An important variation of parallel disjunction occurs when sufficient conditions are mutually exclusive, called exclusive disjunction, equivalent to the Boolean logic XOR operator and the common language phrase "either." In this case the probabilities of causal events must add up to 1, so the parallel complexity equations we just presented now become

$$\mathbb{P}_p = \Psi(\mathbb{Y}_1, \mathbb{Y}_2, \mathbb{Y}_3, \ldots, \mathbb{Y}_m) \quad (6.9)$$

$$\Leftarrow \mathbb{Y}_1 \veebar \mathbb{Y}_2 \veebar \mathbb{Y}_3 \veebar \cdots \veebar \mathbb{Y}_m \quad (6.10)$$

$$P_p = q_1 + q_2 + q_3 + \cdots + q_m = \sum_{j=1}^{m} q_j \quad (6.11)$$

$$= mq. \quad (6.12)$$

There is a symmetrical result for hypoprobability. Regardless of whether causal disjunctive probabilities are inclusive (OR) or exclusive (XOR), overall probability P_p is always *increased* when social complexity is based on a parallel structure—which is also common at the second- and higher-order of causation. **Hyperprobability**, defined by the inequality $Y_p > \max q_j$, is the fundamental property of parallel social complexity. It means that parallel structured social systems have an overall probability of performance that is greater than that of the best performing component.[2]

6.3.1.3 Hybrid Structural Complexity

Most social systems and processes in the real world operate through some combination of serial and parallel structure, especially those that are complex artifacts or complex policies. Examples of this kind of structural complexity are shown in Figs. 6.1(c) and 6.2(c), which show first-order 3-conjunction that embeds 2- and 3-disjunctions of the second-order.

The following two kinds of symmetrical patterns (serial-parallel and parallel-serial) serve as building blocks for modeling far more complex social forms, to *any* desirable degree of structural complexity.

[2]Popular culture is silent about an analog of the serial chain metaphor for the case of a parallel structure. If it existed, it should say: a parallelized system is stronger than its strongest component.

A **serial-parallel system** has first-order Θ-degree serialization, second-order Γ-degree parallelization, and overall probability equation given by

$$Y_{sp} = \left[1 - (1 - Q)^{\Gamma}\right]^{\Theta}. \tag{6.13}$$

This is the kind of structural complexity shown earlier in Figs. 6.1(c) and 6.2(c). In this instance, we may have a 3-stage social *process* where the first and second stages are carried out by two and four parallel activities, respectively. Alternatively, the same structure may represent a social *system* that requires three operating components to undertake action (e.g., legislative, executive, judicial branches of government), the first of which relies on two parallel components (say, a senate and an assembly), and the second relies on four agencies (e.g., such as for policies on security, economics, health, and infrastructure).

The symmetrical opposite is a **parallel-serial system**, which has first-order parallelization, second-order serialization, and overall probability equation

$$Y_{ps} = 1 - \left(1 - P^{\Theta}\right)^{\Gamma}. \tag{6.14}$$

The origin of chiefdoms (sociogenesis) provides an excellent example of hybrid social complexity. Within the overall formative process, a first-order structure of the compound event $\mathbb{P}$ ("the potential for sociogenesis occurs") is given by the following conjunction of necessary causal events:

$$\mathbb{P} = \Psi(\mathbb{X}_{kin}, \mathbb{X}_{com}, \mathbb{X}_{norm}, \ldots, \mathbb{X}_{ca}), \tag{6.15}$$

$$\Leftarrow \langle \mathbb{X}_{kin} \wedge \mathbb{X}_{com} \wedge \mathbb{X}_{norm} \wedge \cdots \wedge \mathbb{X}_{ca} \rangle, \tag{6.16}$$

where $\mathbb{X}_i$ denote various necessary conditions for chiefdom formation, such as existence of kinship knowledge $\mathbb{X}_{kin}$, communicative ability $\mathbb{X}_{com}$, normative knowledge $\mathbb{X}_{norm}$, and collective action ability $\mathbb{X}_{ca}$, among others as examined in the next chapter. Thus, the first-order probability equation is simply

$$P = X_{kin} \cdot X_{com} \cdot X_{norm} \cdots X_{ca} = \prod_{i=kin}^{ca} X_i \tag{6.17}$$

$$= X^{\Theta}, \tag{6.18}$$

consistent with earlier notation. In turn, collective action ability is satisfied through a variety of Γ strategies (e.g., providing incentives, exercising authority, among others), not in just one unique way.[3] Accordingly, the second-order probability equation in terms of Γ strategies is:

$$P = X^{\Theta-1} X_{ca} \tag{6.19}$$

$$= X^{\Theta-1} \cdot \left[1 - (1 - Q)^{\Gamma}\right], \tag{6.20}$$

[3]We will examine collective action theory more closely in the next chapter.

where Q now represents the probability of individual collective action strategies being known.

A more contemporary example consists of modeling the probability of crisis management policies in issue domains such as humanitarian disasters, financial crises, or cybersecurity. First-order complexity is typically serial,

$$P = X_1 \cdot X_2 \cdot X_3 \cdots X_n \tag{6.21}$$

$$= \prod_{i=1}^{n} X_i \tag{6.22}$$

$$= \prod_{i=1}^{n} \left[1 - \prod_{j=1}^{m} (1 - Z_j) \right]_i, \tag{6.23}$$

because n requirements (e.g., accurate intelligence, available capacity, implementation plans, among others) must occur in conjunction. In the case of humanitarian disaster response, supply chain management is also a prominent serialized structure, as are lines of communication. In the case of financial crisis management, passage of legislation and other regulatory procedures have similar serialized structures. However, second-order complexity is often parallelized, as each requirement is ensured through m different approaches or strategies. Alternative locations are often used for dropping humanitarian relief in affected zones, whereas financial crisis policies employ multiple interventions, rather than a single act of government.

From a computational perspective, hybrid social complexity is modeled with code that makes extensive use of functions as subprograms. For example, separate functions can be defined for computing each structural component. This also results in a program being more modular, which is almost always a desirable feature and a real necessity when dealing with algorithmic complex.

6.3.2 Distributional Laws: Scaling and Non-equilibrium Complexity

Social complexity is also characterized by statistical and probability distributions, specifically by **non-equilibrium distributions** and power laws. As suggested earlier in this chapter by the historical overview of milestones and pioneers, over the past century power laws have been shown to exist across multiple domains of social complexity. In almost all cases these distributions are about *size* variables, not durations, which is a intriguing feature that remains somewhat of a scientific mystery. To better appreciate and understand this area of CSS it is best to begin by defining a power law.

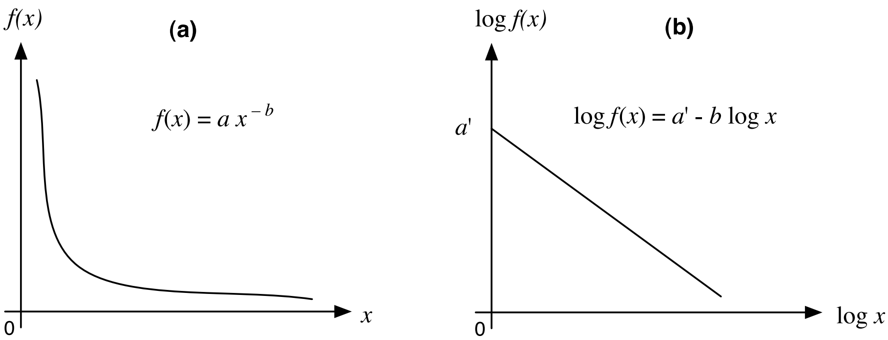


Fig. 6.3 The power law in (**a**) untransformed hyperbolic form and (**b**) linearized or log-linear form in log-log space

Definition 6.1 (Power Law) Let X be a real variable with a set of values $x \in \Re$. A power law is a function of x that is inversely proportional to x itself. Formally,

$$\begin{aligned} f(x) &\propto x^b \\ &= ax^b, \end{aligned} \tag{6.24}$$

where $a > 0$ and $b > 0$.

In purely mathematical terms, a power law refers to any equation of the form

$$y = ax^b, \tag{6.25}$$

where constants a and b can assume any value, such that $f(x)$ in Eq. (6.24) can be either increasing ($b > 0$), decreasing ($b < 0$), or constant ($b = 0$) in x, as well as positive ($a > 0$) or negative ($a < 0$). However, within the context of social complexity theory the term "power law" always implies a negative exponent ($b < 0$) and a positive function ($a > 0$), which in algebraic terms makes Eq. (6.25) the same as a hyperbolic function that is asymptotic in both Cartesian axes, as in Fig. 6.3(a).

For reasons that will become apparent in Sect. 6.3.2.1, the general functional equation (6.25) can be and often is *linearized* by applying a base-10 logarithmic transformation to both sides of the equation, which yields

$$\log f(x) = a' + b \log x, \tag{6.26}$$

where $a' = \log a$ and b now represent an *intercept* and a *slope*, respectively (Fig. 6.3(b)), in log-log space. Note that the slope b is an *elasticity* in log-log space, since

$$\eta_{y,x} = \frac{\partial \log y}{\partial \log x} = \frac{\partial y}{\partial x}\frac{x}{y}.$$

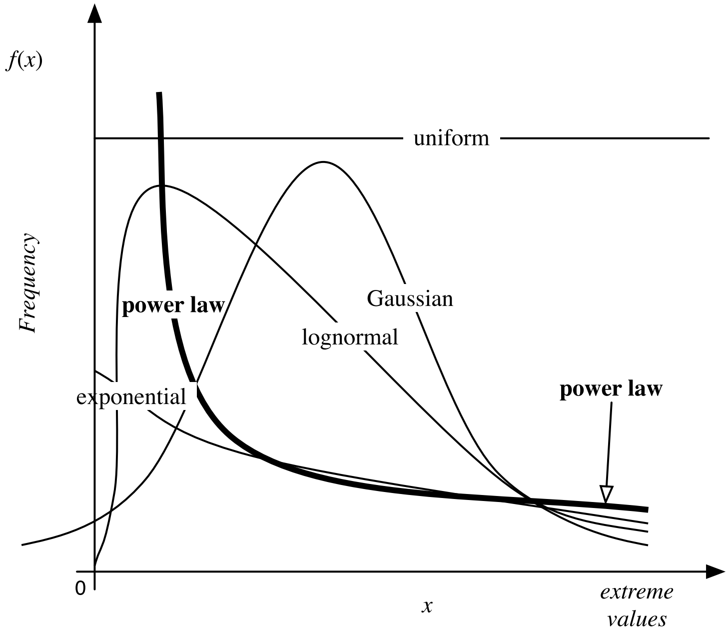


Fig. 6.4 The power law and other distribution models

The log-linear form of Eq. (6.26) is useful from an empirical perspective, because values of x can be plotted on log-log space to examine the form of the distribution, although strictly speaking the term "power law" refers to Eq. (6.25) (with $a > 0$ and $b < 0$), not Eq. (6.26) in log-linear form. For reasons shown below, Eq. (6.25) is the more theoretically relevant equation.

Social scientists familiar with regression analysis will readily recognize Eq. (6.26) as a log-linear regression equation, where both dependent (y) and independent (x) variables have been log-transformed using base 10. In power law analysis the main purpose of log-linearization is *not* to be able to apply ordinary least square (OLS) methods, but to observe how linear the resulting empirical x-y scattergram is and how constant the value of an observed slope $\hat{b}$ is.

Each form of a power law—linear or non-linear, in log-log or linear Cartesian space, respectively—highlights different properties of social complexity, similar to the way in which different forms of the same game in a game-theoretic model (i.e., normal or extensive forms) highlight different features of strategic interaction, or different probability functions (density, cumulative, intensity) provide different views on the uncertainty properties of the same random variable. In addition, each power-law function can also be related to other probability functions, as we shall examine.

Figure 6.4 shows a power law in the context of other distributions. Compared to the so-called normal, Gaussian, or bell-shaped distribution, a power law distribution has *many* small values, *some* (fewer) medium-range values, and a few *rare* extreme

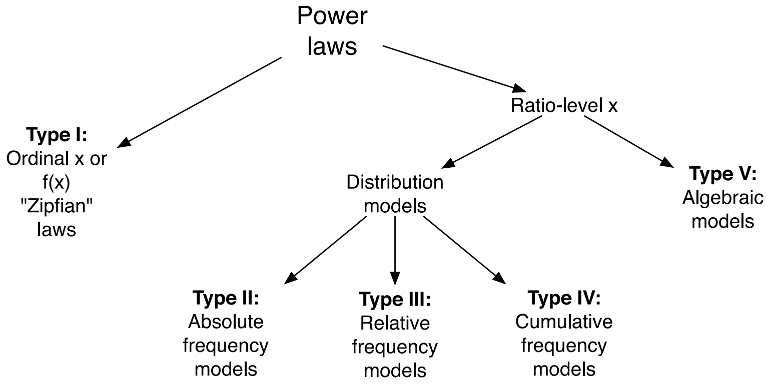


Fig. 6.5 Taxonomy of power law models according to types of dependent variables

values. By contrast, in a Gaussian distribution both smallest and largest values are extremely rare (with vanishingly small probability) and mid-range values are *the norm*.

Crucially, in terms of understanding complexity, extreme events are many times more "normal" in a power law distribution than in a Gaussian distribution. There are also other significant differences with respect to other major types of distributions, such as exponential, uniform, and lognormal, as examined in the next sections.

6.3.2.1 Systematics of Social Power Laws

It would appear from the preceding formalization that power law models are all analytically or formally similar (Eq. (6.25)), in the same sense that all hyperbolas are similar, in that they would differ only by the numerical value of the coefficients a and b. However, that is not the case, because the term on the *left* side of a power law—the function $f(x)$ that is inversely proportional to a given variable x—often denotes widely different quantities when examined in different disciplines and different empirical domains. In addition, as in the case of Zipf's Law, the independent variable can sometimes assume rank-ordinal values, such that the independent variable is not ratio-level.

Given such confusing practices in the literature, it is useful to identify and systematize the most common types of power laws, because the (seemingly) simple form of the linear log-log plots that are commonly reported in publications often conceal interesting subtle differences that stem from quite different quantities being plotted in vertical and horizontal axes, i.e., dependent and independent variables. Similarities and differences among various types of power laws of social complexity are meaningful and should be understood. The taxonomy shown in Fig. 6.5 spans five types of power laws across various social and natural phenomena.

As illustrated in Fig. 6.5, power law models are a class composed of two distinct—albeit related—subclasses or sets of models according to the level of measurement of the independent variable x (ordinal or ratio).[4] In turn, ratio-level power laws comprise several subtypes, as explained in the next sections. In spite of these differences, it must be stressed that all power law models are mathematical representations of extreme skewed variability that are **scale-free**, in the sense discussed below.

6.3.2.2 Type I: Rank-Size or Zipfian Models

The first (and oldest) type of power law model is **Zipf's Law** of harmonic sizes, also known as a **Rank-Size Law** (geography, linguistics) or **rank-size rule** (anthropological archaeology). Given an ordered set of values $\langle x_1, x_2, x_3, \dots, x_n\rangle$ of a variable X, where the subscript i denotes rank from highest ($i = 1$ or first) to lowest ($i = n$ or last), the power law for values of X with respect to rank i of each value $x_i \in X$ is given by the equation

$$x_i = \frac{a}{i^b} \quad \text{(Type I power law)}, \tag{6.27}$$

where $a = x_1$ (the largest value) and $b \approx 1$. Note that from Eq. (6.27) it also follows that for this type of distribution the product of any value $x_i \in X$ times its rank i always equals (or approximates) the constant a (the largest value x_1). Therefore, the largest value determines all other values of the distribution. Such a decreasing series of values is also known as a **harmonic series**, wherein the second largest value is $1/2$ the size of the largest, the third largest value is $1/3$ the size of the largest, $\dots$, and the last (the nth value) is $1/n$ the size of the largest. From Eq. (6.27) it also follows that

$$\log x_i = a' - \log i, \tag{6.28}$$

which is commonly used for analyzing empirical data with log-log plots. By definition, therefore, this type of power law has elasticity equal to 1.

Felix Auerbach was the first to discover this type of power law in the harmonic frequency of population concentrations. Perhaps somewhat unfairly, the model is commonly named after the Harvard linguist George Kingsley Zipf [1902–1950] because it was he who popularized it. This type of power law may be of unique interest in the social sciences and the life sciences (laws of so-called "allometry" or proportion), and perhaps they remain undiscovered in the physical sciences.

As shown in Fig. 6.5, the next three types of power laws consider different distributions of values of X in terms of various frequency measures: absolute frequency (Type II), relative frequency (Type III), and cumulative frequency (Type IV). All three distribution types of power laws—which are canonical variations on the common theme of modeling scale-free inequality—occur in both the social sciences and the natural sciences.

[4] Using the **Stevens level of measurement** as a classification criterion is useful for distinguishing formally different mathematical forms that are analyzed through different statistical and mathematical methods (discrete vs. continuous). The same classification might be less useful in physical power laws, where ranks and ordinal variables are not as common as they are in social science.

6.3.2.3 Type II: Absolute Frequency Models

In the second type of power law the *absolute frequency* ϕ of a given value $x \in X$ is inversely proportional to x. Thus,

$$\phi(x) = \frac{a}{x^b} \quad \text{(Type II power law).} \tag{6.29}$$

From Eq. (6.29) it follows that

$$\log \phi(x) = a' - b \log x, \tag{6.30}$$

where $a' = \log a$ is the intercept and b is the slope (exponent in Eq. (6.29)). Recall that b is also in this case the elasticity η of $\log \phi(x)$ with respect to $\log x$.

In the social sciences this type of power law has been frequently reported for variables as diverse as the size of archaeological sites in a given region, personal income, number of Internet routers, network links, and the number of fatalities that have occurred in warfare on all scales in modern history. Lewis Fry Richardson's Law of War Severity, describing the skewed distribution of fatalities generated by conflicts of all magnitudes, is a power law of this type. In the natural sciences, this type of power law has been reported for the size of species, the lifespan of genera, earthquake energy releases, meteor diameters, and the relative sizes of avalanches in Conway's Game of Life (a cellular automata model examined in Chap. 7).

The next two types of power laws are somewhat similar, since they are both based on probability functions, but different in several interesting, crucial details that are easy to overlook.

6.3.2.4 Type III: PDF Models

The third and closely related type of power law is stated in terms of *relative frequency*, which in the statistical limit approximates a **probability density**. Formally, this is the **hyperbolic probability density function** (p.d.f.)

$$p(x) = \frac{a}{x^b} \quad \text{(Type III power law).} \tag{6.31}$$

(In physics, Eq. (6.31) is often called a "distribution function," which is a mathematical misnomer that can cause confusion. The term "distribution function" refers to the cumulative density function $\Phi(x)$, or "mass function," as in the next section.)[5]

[5]For example, Bak (1996), Jensen (1998), and Barabasi (2002) misname these functions repeatedly—c.d.f., p.d.f., and complementary c.d.f.—as if they were synonymous, whereas each function refers to the probability of a different event: $\mathbf{Pr}(X \leq x)$, $\mathbf{Pr}(x < X \leq x + dx)$, and $\mathbf{Pr}(X > x)$, respectively. The obvious but important point is simply that probability functions that refer to different events should be named differently and consistently.

The log-linear form for the Type III power law is easily derived, from Eq. (6.31), as

$$\log p(x) = a' - b \log x, \tag{6.32}$$

with $a' = \log a$, and, again, b is the elasticity of $\log \phi(x)$ with respect to $\log x$.[6]

This type of power law also has strong empirical support across social domains. It has been reported for the size of firms in terms of employees (Simon's Law), the number of publications by scholars (Lotka's Law), the number of collaborations by movie actors, the size of commodity price fluctuations (Mandelbrot's Law), and other social variables. In the natural and engineering sciences, this same Type-III power law has been reported for the size of species, the connectivity of the US power grid, the size of forest fires (Turcotte's Law), and the size of sandpile avalanches (Bak's Law).

6.3.2.5 Type IV: Log-Survival or Log-CCDF Models

A fourth type of power law is based on the *complementary cumulative density function*, or $1 - \Phi(x) = \mathbf{Pr}(X > x)$, abbreviated as CCDF. When X denotes time T, the CCDF is called a *survival function*, or $S(t)$.[7] In a log-log linear graph this model has the form

$$\log\big[1 - \Phi(x)\big] = a' - (b-1)\log x, \tag{6.33}$$

with $a' = \log a$, which yields the c.d.f.

$$\Phi(x) = 1 - \frac{a}{x^{(b-1)}} = 1 - ax^{1-b} \tag{6.34}$$

and corresponding p.d.f. given by

$$p(x) = \frac{a(b-1)}{x^b} \quad \text{(Type IV power law)}. \tag{6.35}$$

Note that in this type of power law the elasticity in Eq. (6.33) is $\eta = (b-1)$, *not* just b as in previous models—a critical difference to remember! Table 6.1 provides a comparison of the defining probability functions of a Type IV power law model (top row) with respect to other distribution models of social phenomena. Note that the negative exponential p.d.f. also corresponds to a Poisson process, which is common in many social phenomena such as riots, onsets of warfare, and organizational

[6] Note that Type II (absolute frequency) and Type III (relative frequency) yield the same slope b, although the functions on the left side are not mathematically identical.

[7] Also, strictly speaking, the event "$X \geq x$" makes more sense than "$X > x$" when X is a discrete (count) variable. This is because 0.99999... is not computable and 0 is mathematically impossible, so 1 is the base count for social processes such as events, riots, wars, and other social count processes.

Table 6.1 The Type IV power law model of social complexity compared to other common social processes and distributions

Model	p.d.f. $p(x)$	c.d.f. $\Phi(x)$	h.f.f. $H(x)$	Mean $E(x)$
Power law	$\frac{a(b-1)}{x^b}$	$1 - ax^{b-1}$	$\frac{b-1}{x}$	$\frac{a(b-1)}{2-b}x^{2-b}\vert_{x_{\min}}^{\infty}$
Exponential	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	λ	$\frac{1}{\lambda}$
Weibull	$\lambda\gamma x^{\gamma-1}\exp(-\lambda x^{\gamma})$	$1 - \exp(-\lambda x^{\gamma})$	$\lambda\gamma x^{\gamma-1}$	$\lambda^{-1/\gamma}\Gamma(\frac{1}{\gamma}+1)$
Lognormal	$\frac{1}{\sigma x\sqrt{2\pi}} \times$ $\exp[-(\ln(x/m))^2/(2\sigma^2)]$	$1 - \frac{1}{\sigma\sqrt{2\pi}}\int_x^{\infty}\frac{p(u)}{u}du$	$\frac{p(x)}{1-\Phi(x)}$	$\exp(0.5\sigma)$
Gaussian	$\frac{1}{\sigma x\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$	$1 - \frac{1}{\sqrt{2\pi}} \times$ $\int_x^{\infty}\exp[-\frac{1}{2}(\frac{u-\mu}{\sigma})^2]du$	$\frac{p(x)}{1-\Phi(x)}$	μ

turnover. The intensity or hazard force functions (h.f.f.) corresponding to power law, exponential, and Weibull models are of major interest in practical applications. The lognormal and Gaussian cases are also computed as $p(x)/[1-\Phi(x)]$ but are omitted from the table due to space constraints and infrequent use. The graphs of probability density functions in Table 6.1 were shown earlier in Fig. 6.4.

Equation (6.35) looks deceptively similar to a Type III power law (compare with Eq. (6.31)), with the crucial difference that the proportionality constant is partially dependent on the exponent (b) or slope ($b-1$). This fourth type of power law, based on the complementary c.d.f., has been reported for the size of firms in terms of revenue, for fatalities that occur in warfare (both civil wars and international wars), as well as for a variety of natural phenomena including the magnitude of earthquakes (Gutenberg-Richter Law).

An important result that links this type of power law model to other classical distributions models (e.g., Weibull) is given by the following theorem:

Theorem 6.1 (Intensity Function of a Power Law) *Given a Type IV power law with p.d.f. as in Eq.* (6.35) *and c.d.f. as in Eq.* (6.34), *then the associated intensity function or hazard force function* $H(x)$ *is given by*

$$H(x) = \frac{b-1}{x}, \tag{6.36}$$

where $H(x)$ *is defined as* $p(x)/[1-\Phi(x)]$, *which is*:

1. *linear in* b
2. *hyperbolically decreasing in* x *with power law exponent* 1 (*scale-free*),
3. *independent of* a
4. *a special case of the Weibull distribution for* $\gamma(\text{shape}) = -1$ *and* $\lambda(\text{scale}) = b-1$, *or slope of the CCDF in log-log space*
5. *has an associated stress or load function* $\Lambda(x)$ *given by*

$$\Lambda(x) = \int_0^x H(u)du = (b-1)\ln x. \tag{6.37}$$

Proof By substituting Eqs. (6.34) and (6.35) into the definition of $H(x)$ and simplifying the resulting expression to obtain Eq. (6.36). □

Theorem 6.1 is interesting because it provides a simple and direct link between social complexity theory on the one hand, and risk analysis and uncertainty on the other. The principle says that *all complex social phenomena are generated by inverse intensity*. The Weibull model includes one such instance of an inverse function, as do other stochastic processes with hyperbolically decreasing intensity or hazard rate. Conversely, using Eq. (6.36), the intensity function theorem allows us to express a power law as a function of the many features associated with $H(x)$, such as moments and other characteristics.

Types III and IV power laws should never be referred to as "Zipf's Law for $b = 1$," because such terminology implies that these models contain ranked variables; they do not.

6.3.2.6 Type V: Algebraic Models

Finally, a fifth type of power law model found in the literature is based on the linear plot of two ordinary ratio-level variables, so

$$\log y(x) = a' - b \log x \tag{6.38}$$

and

$$y(x) = \frac{a}{x^b}. \tag{6.39}$$

Note that in this case there is no difference between the log-linear slope and the hyperbolic exponent—a property that differs from the previous cases. Although most social scientists do not think of ordinary algebraic expressions such as Eq. (6.39) as a power law, in the natural sciences (and in elementary mathematics) the study of power laws includes these models as well. For example, the relation between the number of routers y and the number of nodes x in the Internet is governed by Eq. (6.38) with $b \approx 1.9$ (Faloutsos's Law). If the class of power laws includes these algebraic relationships or hyperbolic models (type V), then all inverse empirical relationships that are linear in log-log space also qualify as power laws (e.g., Polachek's Law of international conflict and trade, and social gravity models in human geography and regional economics).

It should be reiterated that the preceding five types of power laws share a great deal in common—the right side of the equation is always a term inversely proportional to a given variable x—but the mappings are different because what is modeled on the left side of each equation varies across types. Such variations are sometimes relatively minor, as between Type II (absolute frequencies) and Type III (relative frequencies). Other times they are more significant, as between Type III (p.d.f.-based) and Type IV (c.d.f.-based), or between ratio variables, frequency-based, and variable-based models.[8] Beyond the formal differences highlighted by the preceding taxonomy, all power laws are susceptible to empirical analysis, as discussed in the next section.

6.4 Power Law Analysis

Power laws of social complexity are susceptible to various forms of empirical, data-oriented analysis, as well as theoretical, mathematically-oriented analysis. Both approaches are necessary and synergistic for understanding complexity in social phenomena.

6.4.1 Empirical Analysis: Estimation and Assessing Goodness of Fit

Suppose a given data sample or set of observations $\{x\}$ of a variable X yields a power law of some type (I–IV). From an empirical perspective a review of current practices in the extant literature shows that there are two common procedures for assessing the goodness of fit of a power law model in relation to empirical data: (1) visual inspection of the log-log plot to see if it approximates a straight line, and (2) judging goodness of fit on the basis of a high value for the R^2 statistic. These procedures deserve close scrutiny, because they can be misused, resulting in false inferences.

6.4.1.1 Visual Assessments

Visual assessments are useful, informal, and always subjective. A common problem that is often highlighted by data plotted on log-log scales is "bending" away from the log-linear model at lower and upper ranges of the distribution (see Fig. 6.6).

Bending of an empirical distribution at *lower* quantiles can occur because there might be missing observations for small values that are lost or hard to measure. For example, in a dataset of war magnitudes the smallest wars may not be recorded. This is a form of measurement error that can arise for many reasons. Bending at the lower quantiles can be acceptable if the claim that the smallest observations are incomplete can be supported; otherwise, lower quantile bending presents a serious problem with accepting the research hypothesis that the observed data conforms to a power law.

Bending can be found in empirical data that approximate a power law, but can also be diagnostic of an exponential or lognormal tail. Also, a *uniform distribution* (which is far from being a power law!) plotted on log-log space yields a curved pattern with both lower and upper quantile bending, so the problem in such cases may not be due to missing observations or finite size—it may be because the distribution is close to uniform, not at all a power law or even exponential.

[8] The basic point is that care must be taken to specify which type of power law model is being discussed or presented; this should not have to be deciphered from poorly labeled plots or misnamed equations.

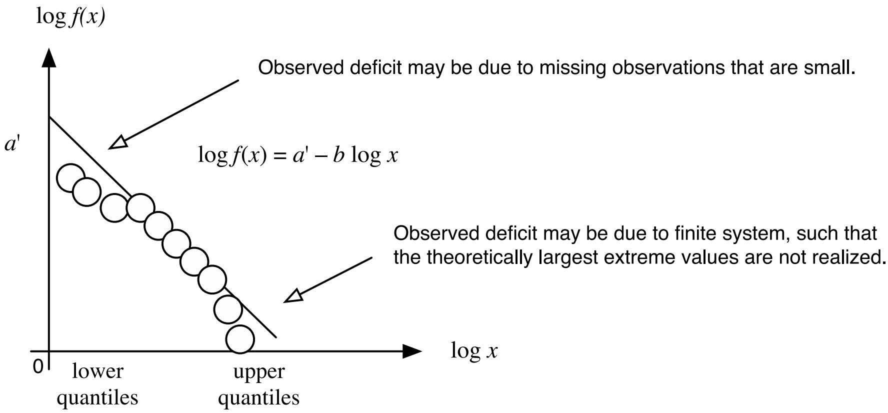


Fig. 6.6 "Bending" is frequently observed in visual assessment of empirical power law distributions

6.4.1.2 R-Squared

In much of the extant literature, goodness of fit is often assessed using the coefficient of determination, R^2. However, R^2 is best avoided as a measure of goodness of fit and the most recent specialized statistical works on size distributions do not discuss it. Other statistics and methods, such as the standard error of the coefficients or the Anderson-Darling test, are preferable when necessary. Still, a good use of the R^2 statistic is for comparing different empirical models that have the same functional form but are estimated using different data samples.

6.4.1.3 Good Practices: Multiple Lines of Evidence

As is normally the case for various estimators, goodness of fit also should be assessed on the basis of multiple methods that provide diverse lines of evidence: small standard errors, large t-ratios, the Kolmogorov-Smirnov test, the Anderson-Darling test, among other methods. The estimation of power law models using *maximum likelihood* methods is recommended, such as based on the **Hill estimator**. Table 6.2 compares various statistical assessments for power laws.

By way of summary, some good practices in the empirical analysis of power laws with statistical data include the following:

1. Use disaggregated data values $\{x\}$ of the observed variable X to construct the relevant frequency distribution plots ensuring that all axes and units of measurement are properly labeled. Report the standard errors of all coefficients when conducting an estimation. Specifically:
   (a) For the Type I power law (Eq. (6.27)), data values are ordered from largest to smallest and the resulting plot should resemble a simple harmonic function with a long upper tail. In log-log space the same data should approximate a straight line with slope value of 1.
   (b) For Type II (Eq. (6.29)), the data values should be used directly to construct a histogram of value frequencies and the results plotted in log-log space. The plot should approximate a straight line, as in Eq. (6.30). Note that in this case

Table 6.2 Goodness of fit statistics used for assessment of an empirical power law

Statistic	Pros	Cons	References
Hill estimator	MLE	Can be unstable for small sample size	Alfarano et al. (2008); Hill (1975)
Anderson-Darling	Sensitive to upper tail values	Rarely used; not well known; Type I error risk	Anderson and Darling (1954)
Kolmogorov-Smirnov	Widely known	Insensitive to upper tail values; Type II error risk	Chakravarti et al. (1967, pp. 392–394)
R^2	Commonly used; good for comparing samples	Not a proper goodness of fit statistic	King (1986)

the estimated slope $\hat{b}$ in Eq. (6.30) is exactly the value of the exponent b in Eq. (6.29)—i.e., without the $(+1)$ transformation that is necessary with the Type IV law.

(c) For Type III (Eq. (6.31)), the procedure is the same as for the Type II power law, except that it is necessary to compute relative as opposed to absolute frequencies.

(d) For Type IV (Eq. (6.35)), which is arguably the most important case, the data values are again used directly, this time to construct the normalized complementary cumulative frequencies—i.e., the values of the function $[1 - \Phi(x)]$, without binning.[9] The log-log plot should then approximate a straight line with slope $(b + 1)$. Accordingly, a *slope* of $(b + 1)$ for the distribution of the complementary c.d.f $[1 - \Phi(x)]$ in log-log space yields an *exponent* of b in the Type IV power law (Eq. (6.35)).[10] That is: *slope* $(b + 1) \rightleftarrows$ *exponent* b.

2. Inspect the upper and lower quantiles for excessive bending. Significant bending should be accounted for (e.g., are there missing observations? is finite size somehow involved?). Otherwise, the power law model simply may not fit the data and other models should therefore be considered (e.g., lognormal?).
3. Inspect the number of orders of magnitude (sometimes called "decades") covered by the domain of values. In general, the larger the number of orders of magnitude the more interesting the model because the scale-free property (discussed in the next section) will extend over several orders. Ensure that the range of orders of magnitude is not an artifact of the units of measurement.

[9] "Binning" refers to the procedure of classifying values into equal and finite intervals, which creates problems when the distribution of the underlying population is unknown. It is unnecessary in power law analysis that uses raw data. The direct construction of the histogram of normalized cumulative frequencies is often feasible and always preferable because no binning is necessary. However, sometimes binning is unavoidable when using official statistics such as provided by government agencies.

[10] The exact value of the exponent b is of great theoretical relevance, as explained below in Sect. 6.4.2.1, so reporting the standard error of b is another good practice.

4. Rely on the most valid and reliable data available, especially when N is not very large, because other issues such as bending and goodness of fit can be greatly affected by data quality.
5. Use the standard errors to assess the coefficient estimates, as well as other methods for assessing goodness of fit, such as the Hill estimator. (Ignore significance tests for the slope estimates of Type IV models, since, by definition, cumulative data will always yield slopes greater than zero.)
6. Avoid the R^2 for purposes of assessing goodness of fit, but use it to compare models that have the same functional form—as a comparative measure.[11]
7. Develop familiarization with standards and methods in various fields where power laws are used to gain a better perspective and improve the quality of empirical analysis in social power law modeling.

These good practices—based on multiple lines of evidence and complementary approaches demonstrated over the past century—are susceptible to improvement as social scientists and other modelers gain experience with empirical applications of power law models. Important scientific goals will be achieved as good practices emerge.

6.4.2 Theoretical Analysis: Deriving Implications

A power law is important, *inter alia*, because of the set of intriguing theoretical implications it can generate, not just because it establishes an empirical regularity based on empirical evidence. This is increasingly relevant as social scientists gain experience in the exploitation of synergies between formal models and empirical data. Among the theoretical implications that can be drawn from finding a power law in a given set of data, the following are especially significant in terms of understanding social complexity.

6.4.2.1 Average Size

The first moment (average or mean value) of a power law distribution exhibits some unusually interesting behavior. This is given by

$$E(x) = \int_{\min\{x\}}^{\infty} x p(x) dx = a(b-1) \int_{\min\{x\}}^{\infty} x^{1-b} dx \tag{6.40}$$

$$= \frac{a(b-1)}{2-b} x^{2-b} \Big|_{\min\{x\}}^{\infty} = \frac{x_{\min}(b-1)}{b-2}, \tag{6.41}$$

which goes to infinity when $b \leq 2$. In other words, there is no mean size (no expected value $E(x)$ exists) for social phenomena that are governed by a power law with exponent in the range $0 < b < 2$, or $(b-1) < 1$ (below unit elasticity). This is an insightful theoretical result for social patterns such as organizational sizes,

[11] However, recall that the standard error of estimates contains essentially the same information.

fatalities in warfare, and terrorist attacks. The threshold $b = 2$ is therefore theoretically critical, as it marks the boundary between social phenomena that have a finite average and computable size ($b > 2$) and those phenomena that lack an expected value or mean size ($b \leq 2$). This is a theoretical insight derived directly from the empirically estimated value of the power law exponent b.

6.4.2.2 Inequality

By definition, a power law is a model of *inequality* (the "many-some-rare" pattern discussed earlier in this chapter), so every power law model has an associated **Lorenz curve** given by:

$$L(\Phi) = 1 - \left[1 - \Phi(x)\right]^{1-1/(b-1)} \tag{6.42}$$

and a corresponding **Gini index** given by

$$G(b) = 1 - 2\int_0^1 L(\Phi)d\Phi = \frac{1}{2b-3}, \tag{6.43}$$

which can be estimated by the empirical equation (Kleiber and Kotz 2003: 35):

$$\hat{G} = \frac{1}{n^2 E(x)} \sum_{i=1}^{n}\sum_{j=1}^{n} |x_i - x_j|. \tag{6.44}$$

These interesting and insightful theoretical links between the exponent b of a power law and its corresponding Gini index G of inequality can be summarized by the following two relations in reference to the tail of a distribution:

$$\begin{array}{c} \text{heavy tail} \\ (b \to 0) \end{array} \iff \left\{ \begin{array}{l} \text{more inequality} \\ \text{less equality} \end{array} \right\} \iff \left\{ \begin{array}{l} \text{smaller } b \\ \text{larger } G \end{array} \right.$$

$$\begin{array}{c} \text{thin tail} \\ (b \to \infty) \end{array} \iff \left\{ \begin{array}{l} \text{more equality} \\ \text{less inequality} \end{array} \right\} \iff \left\{ \begin{array}{l} \text{larger } b \\ \text{smaller } G \end{array} \right.$$

6.4.2.3 Entropy

By extension, the greater inequality of a heavy tail also implies greater **Shannon entropy** in the distribution of values, or

$$U(b) = \ln\left(\frac{b-1}{\min\{x\}}\right) - \frac{1}{b-1} - 1, \tag{6.45}$$

where $\min\{x\}$ is the smallest value in the distribution of X. This last expression establishes a direct connection between complexity theory and information theory by linking Shannon's entropy U to the power law exponent b. Equation (6.45) guarantees the existence of as yet unknown information-theoretic properties of social power laws.

6.4.2.4 Self-Similarity

When a given variable X obeys a power law, a recurring pattern of constant proportion occurs across the entire range of values of X, as highlighted earlier by the linear graph in Fig. 6.3(b). The graph of the transformed function $f^*(x) = \log f(x)$ is as linear in the low range of values as it is in the high range and everywhere in between. This type of global symmetry is called **self-similarity** in complexity theory. Self-similarity is also said to be an "emergent" property, because it applies to a whole set of values, not to individual values or elements.

Self-similarity is also a property of structural laws of social complexity. For example, a system of first-order conjunctions (or disjunctions) embedded by higher-order conjunctions (or disjunctions) is self-similar. A policy process is a classical example of self-similar structural social complexity in terms of overall policy response (first-order), programs (second-order), activities (third-order), down to the smallest required events (nth-order) that produce policy results.

6.4.2.5 Scaling

The property of self-similarity is also known as **scaling**, which has prompted the term "scale-free phenomena." Vilfredo Pareto discovered that wealth and income scale. Lewis F. Richardson discovered in the late 1940s (possibly earlier) that warfare ("deadly quarrels") scales with respect to magnitude μ. Since then, it has been shown that not just international wars but civil wars also scale, as do certain features of terrorism. "Artificial" wars generated by agent-based models also scale. Do other dimensions besides war fatalities, such as time of onset and conflict duration, scale? The answer is: generally, no. Time durations are more often exponentially or Weibull-distributed, as we will discuss in Chap. 9.

Scaling is empirically demonstrated for numerous other dimensions of social phenomena, but remains a deep theoretical notion. Scaling means that dichotomies of small versus large wars are false, because of the scale invariance given by the global power law. Scaling also means that it is a misconception to think that small and large wars share little or nothing in common; they are all—small and large—part of the same overall pattern, just different ranges of a power law governed by an identical set of parameter values. Note that scaling occurs if and *only* if a variable obeys a power law. (Most biological organisms do *not* scale.)

6.4.2.6 Fractal Dimension

If the exponent b of a power law equation were allowed to assume only integer values $(1, 2, 3, 4, \dots)$ then the frequencies associated with each value would decrease inversely by the power of such integer proportions. However, when b assumes fractional values (as many exponents reported in the empirical literature) the range of proportions is itself continuous and no longer discrete as in Euclidean space. This is why the b-value in a power law is often called Mandelbrot's **fractal dimension**. Note that scaling vanishes as $b \to 0$, because all values of X assume the same frequency when $b = 0$, so from a scaling perspective a uniform random variable exists in a 0-dimensional space. A Zipfian power law ($b = 1$) yields a 1-dimensional space. A quadratic power law ($b = 2$ or critical value) yields a 2-dimensional space.

In general, a b-power law yields a b-dimensional space and fractional values of b yield fractal dimensions embedded within Euclidean space. Thus, for $0 < b < 1$ the fractal dimensionality is between a point and a line; for $1 < b < 2$ it is between a line and a plane; for $2 < b < 3$ it is between a plane and a solid; and so on. Thus, the fractal dimension also offers another new classification scheme for social phenomena, an idea that physics has begun to exploit with intriguing insights (e.g., Sornette 2003).

6.4.2.7 Criticality and Driven Threshold Systems

Scaling phenomena can be produced by an underlying process that is driven to a phase of **criticality** by slowly evolving input forces that stress the system. Although the input driving the system can behave continuously, the state variables can change abruptly inside a critical region known as a **bifurcation set**, producing scaled phenomena. A precursor to this important insight was contributed over three decades ago by **Catastrophe Theory**, pioneered by mathematician René Thom [1923–2002]. Complexity theory supports and extends Catastrophe Theory by providing a new interpretation of bifurcation dynamics and metastability. For instance, when a power law is reported for a given social phenomenon, such a finding should prompt a set of catastrophe-theoretic questions that would otherwise not arise:

- Is the phenomenon governed by a **driven threshold** system in the sense of Complexity Theory?
- How is the bifurcation set of critical, metastable states to be interpreted?
- What is the form of the associated **potential function** $P(x)$ defined over the state-space?

The demonstration of extensive scaling in warfare, demography, and economics provides significant support for the idea of criticality and related insights on social complexity, such as metastability, long-range interactions, and universality.

6.4.2.8 Metastability

Social events never "come out of the blue"—they must develop potential before they can occur. Another important theoretical inference that can be drawn from the empirical demonstration of a power law in a given social domain is the complexity-theoretic condition known as "metastability." A system (or, more precisely, a given state $x \in X$ of a system) is said to be **Lyapunov-stable** if it is able to maintain its equilibrium under a range of perturbations. For instance, a positive social relation (e.g., a marriage, a friendship, an alliance) is stable in this sense if it is able to endure in spite of stresses that commonly affect social relations. By contrast, a social system is unstable it if falls apart when stressed, such as a polity or an alliance that ends under the pressure of conflict or unresolved issues. A broad range of social system theories—such as in the work of Pareto, Parsons, Samuelson, Deutsch, Easton, Flannery, Dahl and other social systems theorists—employ this Lyapunov concept of stability.

By contrast, a system is said to develop **metastability** when there exist one or more *potential states* $x' \in X$ or potential operating regimes (with $x \neq x'$), *other than the extant state*, to which the system could transition, given the realization of certain conditions. Metastability is common in many social systems, given their capacity for change. For example, a domestic political system or polity becomes metastable during an election or, even more dramatically, during a constitutional convention. *State failure* occurs when a polity that has first become metastable then loses governance capacity relative to accumulated or unresolved stresses. Similarly, an international system becomes metastable—sometimes increasingly so—in a time of crisis, because an alternate state of overt hostility or actual violence grows as the potential for war increases. In economics, financial markets become metastable when they develop a "bubble" capable of bringing about a market crash. Similarly, from a more positive viewpoint, a state of warfare becomes metastable when the potential for a return to peace increases; domestic turmoil and civil unrest also become metastable—as in state-building operations—as the state potential for governance (capacity) increases relative to stresses. Power laws are diagnostic of metastability because they model social situations where a broad range of states—not just the extant equilibrium or observed status quo—has the potential of being realized. Theories of social change should leverage the concept of metastability inherent in power laws.

6.4.2.9 Long-Range Interactions

Scaling phenomena are produced by systems that evolve into a critical phase where **long-range interactions** become possible and sometimes occur. A system governed by only nearest-neighbor interactions will tend to produce mostly normal or Gaussian-distributed phenomena, or other non-power law phenomena with significantly shorter or thinner tails in the upper (and lower) quantiles.

By contrast, a "globalized" system governed by long-range *spatio-temporal* interactions is subject to non-equilibrium dynamics and processes that produce power laws. In such systems the occurrence of extreme events is orders of magnitude higher (not just greater) than in "normal" (Gaussian) equilibrium systems. The spatial dimension of long-range interactions is fairly straightforward in terms of social or physical distance among social actors. Temporal long-range interactions refer to persistent memory of the past as well as future expectations, as already seen for the Hurst parameter in Sect. 5.5.2.2, Fig. 5.2.

The main purpose of these theoretical observations has been to alert readers to several significant potential implications that go beyond the demonstration of an empirical power law. This is not to suggest that each one of these theoretical implications is valid in every instance of an empirical power law, so these potential implications should be seen as a theoretical heuristic for discovering properties of social phenomena, not as proven properties.

6.5 Universality in Laws of Social Complexity

The social sciences have evolved from an initially unified tradition seeking to uncover universal scientific principles of human and social dynamics—which was the original spirit of the Age of Enlightenment and the rise of modern positive science in recent centuries—to today's condition of significant fragmentation along multiple dimensions: differences in empirical domains, disciplinary cultures, methodologies, even epistemologies. For those intrigued or motivated by the prospect of a unified science of the social universe, structural laws and power laws examined in this chapter offer robust and encouraging grounds for uncovering further universal principles to better understand human dynamics and social complexity based on a common set of empirical and theoretical features, such as those discussed in this chapters.

Self-similarity, scaling, fractal dimensionality, self-organized criticality, metastability, long-range interactions, and universality are all new perspectives surrounding power laws of social phenomena, based on Complexity Theory. These properties and insights were unknown at the time when the first power laws were discovered by Pareto, Zipf, Richardson, and other pioneers. Complexity Theory contains other properties of power laws that may prove insightful for the social sciences. In turn, discovery of power laws in the social sciences may contribute new insights for Complexity Theory and non-equilibrium dynamics.[12]

Recommended Readings

S. Alfarano, T. Lux, F. Wagner, Time variation of higher moments in a financial market with heterogeneous agents: an analytical approach. J. Econ. Dyn. Control **32**(1), 101–136 (2008)

G.A. Almond, S.J. Genco, Clouds, clocks, and the study of politics. World Polit. **29**, 489–522 (1977)

O.M. Ashford, H. Charnock, P.G. Drazin, J.C.R. Hunt, P. Smoker, I. Sutherland (eds.), *The Collected Papers of Lewis Fry Richardson. Volume 2: Quantitative Psychology and Studies of Conflict* (Cambridge University Press, Cambridge, 1993)

B.J.L. Berry, *Geography of Market Centers and Retail Distributions* (Prentice-Hall, Englewood Cliffs, 1967)

L.M.A. Bettencourt, The origins of scaling in cities. Science **340**, 1438–1441 (2013)

C. Cioffi-Revilla, *Politics and Uncertainty: Theory, Models, and Applications* (Cambridge University Press, Cambridge, 1998)

C. Cioffi-Revilla, I. Midlarsky Manus, Power laws, scaling and fractals in the most lethal international and civil wars, in *The Scourge of War: New Extensions on an Old Problem*, ed. by P.F. Diehl (University of Michigan Press, Ann Arbor, 2004), pp. 3–27

R.A. Dahl, *Polyarchy: Participation and Opposition* (Yale University Press, New Haven, 1972)

C. Gini, Sulla misura della concentrazione e della variabilità dei caratteri [On the measure of concentration and variability of traits]. Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti [Annals of the Royal Venetian Institute of Sciences, Letters, and Arts] **53**(2) (1913)

J.W. Kingdon, *Agendas, Alternatives and Public Policies*, 2nd edn. (Harper Collins, New York, 1995)

C. Kleiber, S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences* (Wiley, New York, 2003)

[12]Other important and insightful theoretical extensions of the power law functions discussed in this chapter consist of the gradient ∇f associated with several of the functions. For instance, the field $\mathbf{E}_b = -\nabla f(x; a, b)$ associated with the power law exponent has intrinsic interest for its social interpretation and relation to criticality, metastability, and other complexity concepts. Theoretical and empirical implications of these and other advanced extensions lie beyond the scope of this introductory textbook.

M. Landau, Redundancy, rationality, and the problem of duplication and overlap. Public Adm. Rev. **29**(4), 346–358 (1969)

M. Landau, *Political Theory and Political Science: Studies in the Methodology of Political Inquiry* (Humanities Press, New Jersey, 1979)

M.O. Lorenz, Methods of measuring the concentration of wealth. J. Am. Stat. Assoc. **9**, 209–219 (1905)

A.J. Lotka, The frequency distribution of scientific productivity. J. Wash. Acad. Sci. **16**, 317 (1926)

V. Ostrom, C. Tiebout, R. Warner, The organization of government in metropolitan areas. Am. Polit. Sci. Rev. **55**(4), 831–842 (1961)

J.F. Padgett, Bounded rationality in budgetary research. Am. Polit. Sci. Rev. **74**(2), 354–372 (1980)

V. Pareto, *Cours D'économie Politique* (Editions Rouge, Lausanne, 1897)

J.L. Pressman, A. Wildavsky, *Implementation: How Great Expectations in Washington are Dashed in Oakland* (University of California Press, Berkeley, 1973)

L.F. Richardson, Frequency and occurrence of wars and other fatal quarrels. Nature **148**, 598 (1941)

L.F. Richardson, The distribution of wars in time. J. R. Stat. Soc. A **107**(3–4), 242–250 (1945)

L.F. Richardson, Variation of the frequency of fatal quarrels with magnitude. J. Am. Stat. Assoc. **43**(244), 523–546 (1948)

L.F. Richardson, Is it possible to prove any general statements about historical fact? Br. J. Sociol. **3**(1), 77–84 (1952)

L.F. Richardson, *Statistics of Deadly Quarrels* (Boxwood Press, Pacific Grove, 1960)

W.H. Riker, *The Theory of Political Coalitions* (Yale University Press, New Haven, 1962)

H.A. Simon, On a class of skew distribution functions. Biometrika **42**(3/4), 425–440 (1955)

H.A. Simon, C.P. Bonini, The size distribution of business firms. Am. Econ. Rev. **48**(4), 607–617 (1958)

N.D. Singpurwalla, *Reliability and risk: a Bayesian perspective* (Wiley, New York, 2006)

A. Wohlstetter, The delicate balance of terror. Foreign Aff. J. **37**(1), 211–234 (1959)

G.K. Zipf, The unity of nature, least-action, and natural social science. Sociometry **5**(1), 48–62 (1942)

G.K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Reading, 1949)

Social Complexity III: Theories

7

7.1 Introduction and Motivation

This chapter takes a more formal approach to social complexity ideas introduced in earlier chapters, as required by theoretical analysis. The focus in this chapter is on explanatory theories of social complexity, given the empirical evidence and patterns discussed in Chaps. 5 and 6, respectively. To do this in a systematic way, this chapter highlights elements of causal explanation that are necessary for supporting viable theoretical explanations, in Sect. 7.3. These foundations are then used as a common framework for presenting theories that explain initial social complexity, in Sect. 7.4, as well as more general theories of social complexity that have universal application, in Sect. 7.5.

Based on what has been covered in the previous two chapters, it is essential to recall the primary function of theories: to explain observed phenomena. Laws describe, lines of evidence measure, concepts provide building blocks, and so forth. Hence, each argument that claims to be a theory must conform to a pattern of scientific explanation. A theory is a causal account of observed phenomena based on antecedents or precursor events.

7.2 History and First Pioneers

Contemporary models and theories of social complexity have roots in the 18th century, when social science began to formalize accumulated knowledge through the medium of mathematics. Elementary probability, decision models, and graph-theoretical models were among the earliest mathematical structures used, soon to be followed by dynamical systems of differential equations, game theory, difference equations, stochastic processes, fuzzy sets, and computational models for conducting social simulations and developing theory.

Since the formation and development of polities and social systems has been a subject of intense interest across the social sciences, it is not surprising to learn

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_7,

that "theories of the origin of government" or even "theories of the origin of civilizations" have been appearing since the 18th century. In fact, until the 1950s the first introductory chapter of many political science textbooks focused on origins questions. However, since the behavioral or quantitative revolution of the early post-World War II years, it was mostly anthropology that continued to examine the causes of origins of government, not by design, but by default. Nonetheless, the subject matter remains distributed across the social sciences, so the integrative approach of CSS has acquired increased salience in recent years, especially in the light's of Simon's paradigm.

1762 Political philosopher Jean-Jacques Rousseau [1712–1778] publishes one of the earliest theories of the origin of social complexity in his classic treatise, *Du Contrat Social; Ou Principes du Droit Politique*.

1961, 1963 Robert Dahl of Yale University publishes *Who Governs?* and the first edition of *Modern Political Analysis*, providing foundations for the current standard model of a polity.

1962 Political scientist William H. Riker of the University of Rochester, New York, publishes *The Theory of Political Coalitions*, the first mathematical theory of alliances, based on N-person game theory.

1965 Lofti Zadeh publishes his seminal paper on fuzzy sets, creating a new mathematical approach for formalizing ambiguity in complex systems, including human reasoning and decision-making.

1965 Political scientist David Easton of the University of Chicago, another leader of the Behavioral Revolution, publishes the first edition of *A Systems Analysis of Political Life*, the first systems theory of a polity.

1967 Anthropologist Morton Fried [1923–1986] highlights the significance of asserting elite property rights in the theory of chiefdom formation.

1968 Mathematician and mathematical biologist Nicolas Rashevsky publishes the first mathematical model of chiefly formation in an appendix to his pioneering monograph, *Looking at History Through Mathematics*.

1969, 1996 Herbert A. Simon proposes his Artifactual Theory of Social Complexity for the first time in the first edition of his classic work, *The Sciences of the Artificial*.

1969 In the same year Martin Landau publishes his first pioneering paper on redundancy in organizational complexity, demonstrating the so-called hyperprobability effect.

1971 Robert Dahl publishes *Polyarchy*, the first theory to explicitly account for contending political authorities within the same polity.

1972 Anthropologist Robert Carneiro proposes his influential Theory of Circumscription for explaining the origin of early states.

1977 Anthropologist Timothy Earle begins contributing to the theory of chiefdom formation, based on control over sources of power; Henry Wright publishes his influential paper on "Recent Research on the Origin of the State."

1978 Simon is awarded the Nobel Memorial Prize in Economics "for his pioneering research into the decision-making process within economic organizations."

1983, 1989 Archaeologist Joyce Marcus proposes the Dynamic Theory of Chiefdom Cycling for explaining the origin of early states.

1987 Carol Crumley introduces the concept of *heterarchy* in anthropology, meaning the same as polyarchic (Dahl 1971) and polycentric systems (Ostrom et al. 1961) in political science.

1994 The EOS Project on modeling Upper Paleolithic social change is published in the UK by computer scientist Jim Doran and collaborators.

1996 Nobel laureate Herbert A. Simon publishes the third and last edition of *The Sciences of the Artificial*, adding a new chapter on social complexity and near-decomposability.

1997 Timothy Earl publishes his synthesis of social complexity theory and case studies in *How Chiefs Come to Power*.

1997 Computational social scientists Lena Sanders and Denise Pumain of the University of Paris-Sorbonne publish the SIMPOP model, the first hexagon-based cellular automata model of early urbanization, in the journal *Environment and Planning B: Planning and Design*.

1998 Archaeologist Charles S. Spencer publishes a paper on "A Mathematical Model of Primary State Formation" in the journal *Cultural Dynamics*.

2002, 2005 The Canonical Theory of Social Complexity is proposed (2002) and published (2005) in the *Journal of Mathematical Sociology* as a general theory for explaining original emergence and historical development of social complexity.

2003 American computational social scientist Peter Turchin publishes one of the first cellular automata models of a system of chiefdoms in his seminal book *Historical Dynamics*.

2007 The first empirically calibrated agent-based model of early states in ancient Mesopotamia is published by Tony Wilkinson, John Christensen, and collaborators from the University of Chicago and Argonne National Laboratory. Charles Stanish and collaborators at UCLA publish the first agent-based model of social complexity in ancient Peru and Bolivia.

2009 Behavioral scientists David Lewis-Williams and David Pearce publish *Inside the Neolithic Mind: Consciousness, Cosmos and the Realm of the Gods*, a theory explaining the role of shamans and religion in the origins of social complexity.

2010 A formal model for the cycling of complexity in early societies appears in the online journal *Cliodynamics: Journal of Theoretical and Mathematical History*.

2011 Political scientist Francis Fukuyama of Stanford University publishes a theory of political complexity based on the rule of law.

7.3 Theories of Social Complexity: Elements of Explanation

A defining characteristic of a scientific theory is that it must always contain a story or causal narrative that links antecedents (causes) to consequences (effects).[1] Theories of social complexity must explain its emergence through a causal process or mechanism, the hallmark of which is the ability of the process to account for observed facts or empirical patterns in available data.

The object of explanation—what is being explained, or *explanandum*—is the emergence of social complexity. The explanation, or *explanans*, is a theory. A more formal definition of emergence of social complexity, one that is mathematically tractable, is therefore desirable in order to develop theoretical explanations and understanding.

Definition 7.1 (Emergence of Social Complexity) The emergence of social complexity is a *compound event* $\mathbb{C}$ at a given *macro-level of reference* consisting of a specific combination of more *elemental events* (*sample points*) at a lower *micro-level* in a *sample space* Ω produced by *human decisions* $\square$ and *natural lotteries* $\triangledown$ involved in adaptation via the creation of artifacts.

Emergence of social complexity is well defined if the following two components—what constitutes a compound event—are specified: (a) a set of more elemental micro-level events (sample points consisting of decisional outcomes and states of nature associated with adaptation) and (b) an operational rule that causally links such events. The use of decisional outcomes and states of nature as elementary occurrences grounds theory on micro-foundations.[2] Based on elementary probability theory, the sample points that are used to define an event are axiomatic, left undefined. Similarly, at some point, the elemental events composing the emergence of social complexity are left undefined. At which point? The answer is: at some point beyond which we do not care. Given that social complexity emerges as a result of human decisions (as opposed to being mostly the result of states of nature), a natural resting place for modeling and explaining the occurrence of social complexity is at the level of decisional outcomes. In turn, the elements of a choice situation are generally, albeit not always, considered to be states of nature, no longer decisional outcomes. This approach also allows the theory to rest on micro-foundations of decision-making performed by agents.

What explains the emergence of social complexity is a causal logic that makes the event occur, based on how other causal events from the background sample space occur or fail to occur. For example, for a state to be created from a pre-existing system

[1] Charles A. Lave and James G. March explain the character of theories as causal "stories" in their social science classic, *An Introduction to Models in the Social Sciences* (1993).

[2] By convention, events are written in uppercase hollow letters (e.g., $\mathbb{C}$); variables are in uppercase italics (e.g., C). Event $\mathbb{C}$ is defined on the sample space Ω, variable C is defined on the set of values. Each realization of a variable constitutes an event; a variable is a set of realizations. These conceptual distinctions are critical for developing a unified theory linking macro and micro levels of analysis.

of rival chiefdoms, prior related events connected with strategic decision-making, leadership, procurement of capabilities, enlistment of allies, and so on, must occur or fail to occur in a given combination, or sometimes in one of several equally effective combinations. Causal events must occur in non-arbitrary ways in order for social complexity to emerge. For collective action to take place, a critical combination of causal events must occur in a specific way; otherwise collective action will not occur. Today, as was true thousands of years ago, the process of state formation is caused by specifiable causal events; it does not just happen. In general, the emergence of social complexity is caused by more elementary and sometimes unobservable states of nature and decisional outcomes. The next tool we need to explain social complexity is a way of mapping causal events onto its emergence.

Definition 7.2 (Event Function) Given a compound event $\mathbb{Y}$ and a set of other events $\{\mathbb{X}\}$ causally connected to the occurrence or failure of $\mathbb{Y}$, the mapping $\Psi : \{\mathbb{X}\} \to \mathbb{Y}$ is called the event function of $\mathbb{Y}$. Thus, $\mathbb{Y} = \Psi\{\mathbb{X}\}$.

An event function $\Psi(\cdot)$ defines any causal explanation, which in practice means modeling a function of functions of functions ... to whatever desired depth in a theory's causal argument $\{\mathbb{X}\}$. From a computational perspective, this means writing code with many embedded functions down to the desired resolution. Based on this definition, the event function for emergence of social complexity can be defined as follows.

Definition 7.3 (Event Function for Emergence of Social Complexity) Given a compound event $\mathbb{C}$ of emergent social complexity and a set of other events $\{\mathbb{X}\}$ causally connected to the occurrence or failure of $\mathbb{C}$, the mapping $\Psi : \{\mathbb{X}\} \to \mathbb{C}$ is called the event function of $\mathbb{C}$. Thus, $\mathbb{C} = \Psi\{\mathbb{X}\}$.

Formally, the argument of an event function spells out in specific detail the exact causal logic explaining how a compound event is produced. Which event functions exist and how do different event functions explain the occurrence of a compound event? How does an event function determine the probability of a compound event? To answer these questions, and others like them, we must now examine the logic of social complexity at the micro level to identify two causal modes of explanation based on *sequential logic* and *conditional logic*.

7.3.1 Sequentiality: Modeling Processes. Forward Logic

In **sequential logic mode**, the emergence of social complexity as a compound event is explained by providing a temporal succession or path of prior events that leads to emergence as an outcome.[3] In this mode the emergence of social complexity $\mathbb{C}$ is

[3] The popular idiom according to which "nothing simply comes out of the blue" provides an apt description of so-called forward logic.

explained as an outcome—one among several possible events—which takes place in the sample space Ω of a branching process P that passes through several lotteries and decisions. Sequential logic generally places most of the explanatory emphasis on a process-oriented causal argument with several intervening contingencies, looking toward the future from the vantage point of the past—hence the term **forward logic**. The occurrence of a compound event in sequential logic mode is explained more as a possible outcome, among several alternative outcomes, rather than as a given that must occur.

Example: Polity formation. Polities at any level of complexity cannot form without prior occurrence of necessary events, such as certain kinds of shared knowledge and sets of skills, including leadership-related events. Polity formation is only one of several outcomes; others may be continued disaggregation or warfare.

Example: Hazards and humanitarian disasters. Hazards are natural, anthropogenic, or technological occurrences that can cause damage to humans, especially when people fail to prepare for them or actually increase risk by ignoring warnings or increasing exposure, such as settling in seismically or volcanically active zones.

Example: Financial crises and recessions. Severe economic conditions originate with earlier events, such as irresponsible policies, institutional failures, abusive legal practices, fraud, over-consumption, indebtedness, and similar antecedents.

Example: Contentious crises and war. Conflicts of all kinds result from escalation of violence that originates in antecedent events such as unresolved grievances, adversarial decisions, and other root events.

Example: Political crises and collapse. Polities do not simply collapse for no reason. They do so when earlier events begin to detract capability and other factors increase stress to a point where the polity is no longer viable.

Forward logic is reminiscent of extensive form games, including the use of sequential event trees to describe the causal process. The **initiating event** $\mathbb{I}$ marks the start of a sequential process $\mathscr{P}_N(\mathbb{I} \to \mathbb{C})$ leading to some event $\mathbb{C}$ after N **branching nodes**, where $\mathbb{I}$ is chosen as a base state, such that the occurrence of $\mathbb{C}$ is remote or even impossible unless a number of future contingencies occur. For example, in the previous examples, initiating events are given by base states such as a pre-complex polity, society unaffected by significant hazards, an economy in good health, a peaceful society with insignificant risk of warfare, and a viable polity with surplus capacity, respectively. Branching nodes between $\mathbb{I}$ and outcomes in the space Ω are given by **decisional outcomes**, generated by **human choices**, and **states of nature**, generated by **lotteries**, where both choices and lotteries are cases of contingencies.[4] Thus, based on sequential, forward logic, social complexity oc-

[4]Note the dichotomous taxonomy of events as either "decisional outcomes" or "states of nature." The former are generated by human decisions, whereas the latter are produced by lotteries. "Inflation increases by 1.2 percent" is a *state of nature*, because it is not an event that is decided by anyone; so, its generative mechanism is called a *lottery*. "Humanitarian assistance will be provided to Kenya" is a *decisional outcome* generated by a human *choice*, not a product of any lottery.

curs through a contingent, evolutionary sequence of prior events initiating from a base state.

Assumption 7.1 (Sequential Causal Logic of Social Complexity) *Social complexity $\mathbb{C}$ emerges as a future outcome at time τ in the sample space Ω of a branching process that begins at $\tau - n$. Formally,*

$\mathbb{C}$ occurs iff "$\mathbb{X}_{\tau-1}$ | all necessary events since $\mathbb{X}_{\tau-n}$," *so* $\mathbb{C} \Leftarrow \mathbb{X}_{\tau-1} \Leftarrow \mathbb{X}_{\tau-2} \Leftarrow \cdots \Leftarrow \mathbb{X}_{\tau-n+1} \Leftarrow \mathbb{X}_{\tau-n}$.

Theorem 7.1 (Sequential Probability of Social Complexity) *The emergence of social complexity $\mathbb{C}$ with event function given by*

$$\mathbb{C} = \mathbb{X}_{\tau-1} \wedge \mathbb{X}_{\tau-2} \wedge \cdots \wedge \mathbb{X}_{\tau-n+1} \wedge \mathbb{X}_{\tau-n}, \tag{7.1}$$

where the time index τ denotes time before the occurrence of $\mathbb{C}$ and each event is dependent on the previous event, has sequential probability given by the product of conditional probabilities

$$\Pr(\mathbb{C}) = p_{-n} \cdot p_{-n+1} \cdot p_{-n+2} \cdots p_{-1} = \prod_{i=0}^{n-1} p_i \tag{7.2}$$

$$= P^{\Lambda}, \tag{7.3}$$

where $\Lambda = 0, 1, 2, 3, \ldots, n-1$, *and*:

$$p_{-n} = \Pr(\mathbb{X}_{-n}), \text{ for the first (initiating) event} \tag{7.4}$$

$$p_{-n+1} = \Pr(\mathbb{X}_{-n+1} \mid \mathbb{X}_{-n}), \text{ for the second event} \tag{7.5}$$

$$p_{-n+2} = \Pr(\mathbb{X}_{-n+2} \mid \mathbb{X}_{-n} \wedge \mathbb{X}_{-n+1}), \text{ for the third event} \tag{7.6}$$

$$\vdots \tag{7.7}$$

$$p_{-1} = \Pr(\mathbb{X}_{-1} \mid \textit{all prior events}), \text{ for the last event prior to } \mathbb{C} \tag{7.8}$$

$$\Lambda = \textit{number of prior events leading to } \mathbb{C}, \textit{ or length of the process}, \tag{7.9}$$

and $P = p_n = p_{-n+1} = p_{-n+2} = \cdots = p_{-1}$, *when the individual probability of each event is taken as the same.*

In general, all events prior to an outcome of interest $\mathbb{C}$, such as the sequential priors $\mathbb{X}_i$ in Eq. (7.1), constitute a **potential** for $\mathbb{C}$, or "a potential for the realization of $\mathbb{C}$."

Theorem 7.2 (Sequential Hypoprobability of Social Complexity) *When prior events of an emergent social complexity outcome $\mathbb{C}$ have not yet occurred, the a priori probability of $\mathbb{C}$ (the "out-of-the-blue probability") is always*: (i) smaller *than* *the individual probability* P, *and* (ii) *smaller than the* smallest *of the probabilities of the prior events. Formally*,

$$\Pr(\mathbb{C}) < \min\langle p_{-n}, p_{-n+1}, p_{-n+2}, \ldots, p_{-1}\rangle \tag{7.10}$$
$$< P. \tag{7.11}$$

Looking at the probability of any of the Λ prior causal events leading to $\mathbb{C}$ is always misleading, because such probabilities always overestimate the objective value of $\Pr(\mathbb{C})$. Moreover, in sequential logic a compound event such as $\mathbb{C}$ always occurs with probability *lower* than the *least* probable of the priors.[5]

Another interesting theoretical property of social complexity, from a forward logic perspective, has to do with different effects of changes in prior events and in the length of the branching process. Which of the two has greater effect? In other words, what has greater effect on $\Pr(\mathbb{C})$: changes in $p_i \in P$ or changes in Λ? The precise answer is developed by the following principles.

Theorem 7.3 (Dependence of Sequential Probability on the Probability of Priors) *The rate of change in the sequential probability of a social complexity outcome* $\mathbb{C}$ *with respect to change in the probability of prior events* P *is given by the expression*

$$\frac{\partial \Pr(\mathbb{C})}{\partial P} = \Lambda P^{\Lambda - 1}, \tag{7.12}$$

which is always positive, *so* $\Pr(\mathbb{C})$ *is concave with respect to P.*

Theorem 7.4 (Dependence of Sequential Probability on Length of Process) *The rate of change in the sequential probability of a social complexity outcome* $\mathbb{C}$ *with respect to change in the number of prior events* Λ *is given by the expression*:

$$\frac{\Delta \Pr(\mathbb{C})}{\Delta \Lambda} = P^{\Lambda + 1} - P^{\Lambda}, \tag{7.13}$$

which is always negative, *so* $\Pr(\mathbb{C})$ *is convex with respect to* Λ.

Both dependence equations are non-linear, consistent with the complexity of emergent compound events, such as $\mathbb{C}$. These theorems serve as building blocks for answering the previous question.

[5]Hypoprobability has nothing to do with incomplete information. The effect emerges from the fundamental character of uncertainty as expressed by the sequential probability theorem. No amount of additional information or intelligence can narrow the gap between the probability of prior events and the sequential probability of a compound outcome.

Theorem 7.5 (Sequential Dominance Principle) *The sequential probability of social complexity outcome $\mathbb{C}$ is more sensitive to the probability P of prior causal events in the branching process than to the number of events Λ. Formally,*

$$s_P > s_\Lambda, \tag{7.14}$$

because

$$\frac{\partial \Pr(\mathbb{C})}{\partial P} \frac{P}{\Pr(\mathbb{C})} > \frac{\Delta \Pr(\mathbb{C})}{\Delta \Lambda} \frac{\Lambda}{\Pr(\mathbb{C})}. \tag{7.15}$$

For many, the dominance principle is counter-intuitive, because intuition would have us place greater causal attention on cardinality than on probability—exactly the opposite is true. Informally, we might say something like "prior causal events leading up to $\mathbb{C}$ count more individually than in their total number" or "it matters more to change the probability of causal priors than to alter their total number." This answer is not straightforward without formal analysis, which can be verified computationally. In terms of social complexity, this is often good news, because policies can affect probabilities whereas the cardinality of prior causes usually depends on nature:

Example: Polity formation Changes in the probability of polity formation antecedents matter more than changes in the total number of antecedents.

Example: Hazards and humanitarian disasters The probability of experiencing a disaster is influenced more by the probability of hazards, preparations, and other antecedents than by their total number.

Example: Financial crises and recessions Averting a financial crisis depends more on ensuring the quality of policies than on increasing their total number.

Example: Contentious crises and war Conflict prevention is more sensitive to the probability of escalation, retaliation, and other interactions, than it is to the length of the road to war.

Example: Political crises and collapse The collapse of polities is more affected by the probability of critical failures, such as significant losses in human capital, state resources, infrastructure, and other debilitating failures, than by the total number of possible failures.

7.3.2 Conditionality: Modeling Causes. Backward Logic

The emergence of social complexity $\mathbb{C}$ as a compound event in **conditional logic mode** is explained by providing necessary or sufficient conditions, where the "or" means "and/or." Conditional logic places explanatory emphasis on the Boolean structure of a causal argument, looking toward background conditions from the vantage point of the present—hence the term **backward logic**. The occurrence of a compound event in conditional logic mode is best explained as a given that must somehow be accounted for, not as a possible process outcome.

In Chap. 6 we examined descriptive laws of social complexity by introducing serial, parallel, and hybrid structures for compound events such as emergence of social complexity $\mathbb{C}$. We also introduced the dual concepts of hypo- and hyper-probability as emergent properties of conjunctive/serial and disjunctive/parallel structures, respectively. Here we look at these more closely.

Assumption 7.2 (Conditional Causal Logic of Social Complexity) *Social complexity* $\mathbb{C}$ *emerges in dual causal modes*: *either by joint occurrence of* necessary *conditions* (intersection *of events* $\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \dots, \mathbb{X}_n$, *by Boolean logic conjunctive* AND); *or by occurrence of one among several* sufficient *conditions* (union *of events* $\mathbb{Z}_1, \mathbb{Z}_2, \mathbb{Z}_3, \dots, \mathbb{Z}_m$, *by Boolean logic disjunctive* OR). *Formally*,

$$\mathbb{C}_X = \Psi_\cap(\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \dots, \mathbb{X}_n) \tag{7.16}$$

$$= \mathbb{X}_1 \wedge \mathbb{X}_2 \wedge \mathbb{X}_3 \wedge \cdots \wedge \mathbb{X}_n \tag{7.17}$$

for a conjunctive (AND-*caused*) *event* $\mathbb{C}_X$, *and*

$$\mathbb{C}_Z = \Psi_\cup(\mathbb{Z}_1, \mathbb{Z}_2, \mathbb{Z}_3, \dots, \mathbb{Z}_m) \tag{7.18}$$

$$= \mathbb{Z}_1 \vee \mathbb{Z}_2 \vee \mathbb{Z}_3 \vee \cdots \vee \mathbb{Z}_m \tag{7.19}$$

for a disjunctive (OR-*caused*) *event* $\mathbb{C}_Z$.

The fundamental theoretical reason why the conditional logic assumption on dual causality is true for social complexity events $\mathbb{C}$, as it is for all compound events, is because the sample space Ω of causal events can always be partitioned in logically orthogonal but causally equivalent ways to generate the same compound event $\mathbb{C}$.

Backward logic explanations of social complexity are universally based on two conditional operators (conjunction/intersection/AND and disjunction/union/OR) and a variation on each of them (sequential and exclusive extensions, respectively). These four backward logic operators are examined next.

7.3.2.1 Serialized Complexity: Logic Conjunction ∼ Set Intersection ∼ Boolean AND

The following principle of Social Complexity Theory follows from application of the fundamental theorem for the probability of a compound event.

Theorem 7.6 (Conjunctive Principle of Social Complexity) *The probability of social complexity* $\mathbb{C}$ *by conjunction is given by the product of probabilities of its n necessary events*. *Formally*,

$$\Pr(\mathbb{C}_X) = \Pr\Big(\bigwedge \mathbb{X}_i\Big) = \prod_{i=1}^{n} \Pr(\mathbb{X}_i) \tag{7.20}$$

$$= p_1 p_2 p_3 \cdots p_n = P^{\Theta}, \tag{7.21}$$

where Θ denotes the number of necessary causal events for $\mathbb{C}$ to occur $(2 < \Theta < n)$ *and P is the probability of these events*.

Theorem 7.6 is the cornerstone of Social Complexity Theory when emergence is understood as *a macro-level compound event generated by micro-level causal events*. Comparing Eqs. (7.21) and (7.11) we can see immediately that the sequential logic mode was a special case of conjunction, also called **sequential Boolean AND**, or conjunction by sequential conditionality. Hence, some of the properties of sequential forward models of social complexity also apply to conditional backward models based on conjunction. Hypoprobability, dependence, and dominance principles are among the most significant. (They are not repeated here in the interest of space.)

7.3.2.2 Parallelized Complexity: Logic Disjunction ∼ Set Union ∼ Boolean OR

The next principle follows from Theorem 7.6.

Theorem 7.7 (Disjunctive Principle of Social Complexity) *The probability of social complexity $\mathbb{C}$ by disjunction is given by the following equations*:

$$\Pr(\mathbb{C}_Z) = \Pr\Big(\bigvee \mathbb{Z}_j\Big) = 1 - \prod_{j=1}^{m}\big[1 - \Pr(\mathbb{Z}_j)\big] \tag{7.22}$$

$$= 1 - (1-q_1)(1-q_2)(1-q_3)\cdots(1-q_m) \tag{7.23}$$

$$= 1 - (1-Q)^{\Gamma}, \tag{7.24}$$

where Γ denotes the number of sufficient causal events for $\mathbb{C}$ to occur $(2 < \Gamma < m)$ *and Q is their probability.*

The proof of Theorem 7.7 is easily seen by noting that the disjunctive failure of social complexity to emerge, event $\neg\mathbb{C}_Z$, has probability $1 - \Pr(\mathbb{C}_Z)$, which is

$$\big[1 - \Pr(\mathbb{Z}_j)\big] = (1-Q)^{\Gamma}. \tag{7.25}$$

The following principle follows from Theorem 7.7.

Theorem 7.8 (Hyperprobability Principle) *When emergence of social complexity $\mathbb{C}$ occurs by disjunction of other causal events*, *the probability of $\mathbb{C}$ is always*: (i) larger *than the individual probability Q of individual causal events*, *and* (ii) larger *than the largest of the probabilities of the causal events*. *Formally,*

$$\Pr(\mathbb{C}) > \max\{q_1, q_2, q_3, \ldots, q_m\} \tag{7.26}$$

$$> Q. \tag{7.27}$$

Hyperprobability and hypoprobability principles highlight precise and symmetrically opposite properties of social complexity generated by disjunctive and conjunctive causal structures, respectively.

How is the probability of disjunctive social complexity affected by changes in the probability and number of redundancies? Which effect is dominant? The following dependence, sensitivity, and dominance principles for disjunctive social complexity follow from Theorem 7.7, with similar multivariate analysis as for the conjunctive mode.

Theorem 7.9 (Dependence on Probability of Redundancies Q) *The rate of change in the probability of a disjunctive social complexity event $\mathbb{C}$ with respect to change in probability of causal events Q is given by the expression*

$$\frac{\partial \Pr(\mathbb{C})}{\partial Q} = \Gamma(1-Q)^{\Gamma-1}, \tag{7.28}$$

which is always positive, so $\Pr(\mathbb{C})$ *is concave with respect to Q.*

Theorem 7.10 (Dependence on Number of Redundancies Γ) *The rate of change in the probability of a disjunctive social complexity event $\mathbb{C}$ with respect to change in the number of causal events Γ is given by the expression*

$$\frac{\Delta \Pr(\mathbb{C})}{\Delta \Gamma} = Q(1-Q)^{\Gamma}, \tag{7.29}$$

which is always positive, so $\Pr(\mathbb{C})$ *is concave with respect to Γ.*

Both dependence equations are non-linear, as expected by the compound event probability theorem, consistent with previous results. However, note that redundancy/sufficiency Γ has an opposite effect from necessity Λ. These theorems serve as building blocks for answering the previous question on the dominant effect of redundancies on overall disjunctive probability of a social complexity event $\mathbb{C}$.

Theorem 7.11 (Dominance Principle for Disjunctive Social Complexity) *The probability of a disjunctive social complexity event $\mathbb{C}$ is more sensitive to the probability Q of redundant/sufficient causal events than to the number of events Γ. Formally,*

$$s_Q > s_\Gamma, \tag{7.30}$$

because

$$\frac{\partial \Pr(\mathbb{C})}{\partial Q}\frac{Q}{\Pr(\mathbb{C})} > \frac{\Delta \Pr(\mathbb{C})}{\Delta \Gamma}\frac{\Gamma}{\Pr(\mathbb{C})}. \tag{7.31}$$

All previous results for parallelized social complexity are valid under the standard Boolean OR, also called **inclusive disjunction**, meaning "and/or" in common language. Logically, $\mathbb{X} \vee \mathbb{Z} = (\mathbb{X} \veebar \mathbb{Z}) \wedge (\mathbb{X} \wedge Z)$, where the latter conjunction is inclusive. A variation on this is the **exclusive disjunction**, also known as Boolean XOR, which is defined as $\mathbb{X} \veebar \mathbb{Z} = (\mathbb{X} \vee \mathbb{Z}) \wedge \neg(\mathbb{X} \wedge Z)$.

7.3.3 Hybrid Bimodal Social Complexity: Several-Among-Some Causes

The previous two causal situations—conjunction and disjunction—represent pure causal modes, in the sense that social complexity $\mathbb{C}$ is modeled as requiring either necessary or sufficient causes. However, between these two causal extremes lie many cases of social complexity caused by partial necessity or partial sufficiency. This happens when several causes (more than one) must occur from among a broader set. An example of this occurs in collective action, which is initiated not by the totality of individuals in a society, or by a single individual acting alone, but rather by some core subgroup—which, in turn, may consist of a single leader plus a few close followers.

Another example is in public policy for addressing complex issues. Typically, a set of programs is prepared and implemented, hoping that some measures will work, knowing that all will probably not work, and that one alone is insufficient to obtaining desired results.

Many voting bodies also share this form of social complexity. For instance, this is the case when unanimity is not required but some minimal set of votes is prescribed for approving a decision. In the United Nations Security Council, for example, 5 out of 10 non-permanent members must vote with all five permanent members to pass a resolution.

The **several-among-some** structure of social complexity is generalized by the binomial combination of a number ν of minimally necessary requirements among m that are available, where $m > \nu > 1$. This means that the number of causal combinations that can support $\mathbb{C}$—even if not all are equally feasible—is given by

$$\binom{m}{\nu} = \frac{m!}{(m+\nu)!\nu!}, \tag{7.32}$$

where $m! = m(m-1)(m-2)\cdots 1$ is the factorial of m. Several-among-some complexity is significant because, formally, it reduces to

(i) the pure conjunctive case as $\nu \to m$ (by Theorem 7.6) and
(ii) the pure disjunctive case as $\nu \to 1$ (by Theorem 7.7).

The cardinal number ν is therefore a *critical modal variable*: toward the upper bound $(\nu \to m)$, complexity is caused by conjunction of necessary causes (with hypoprobability), whereas toward the lower bound $(\nu \to 1)$ causation is disjunctive (with hyperprobability).

An oversized coalition experiences hyperprobability when excess members belong to the coalition. If members begin to leave and the coalition reaches minimal winning size, then hypoprobability begins to set in until the critical threshold is reached, beyond which the coalition collapses.

Theorem 7.12 (Several-Among-Some Principle) *The probability of a social complexity event $\mathbb{C}$ caused by a minimum ν conditions from among a set of m that are possible or available, with a ν-out-of-m event function, is given by the equation*

$$\Pr(\mathbb{C}) = \sum_{i=\nu}^{m} \binom{m}{\nu} P^{i}(1-P)^{m-i}, \tag{7.33}$$

where P is the probability of the causal events and $i = 1, 2, 3, \ldots, \nu, \ldots, m$.

Numerous aspects of political complexity are due to combinatorial complexity. This principle of partial necessity/sufficiency reduces to the simpler conjunctive principle (7.6) and to the disjunctive principle when $\nu \to m$ and $\nu \to 1$, respectively. This is a powerful result in Social Complexity Theory, because it encompasses both conjunctive and disjunctive causal structures. The principle is strongly non-linear in P, as each binomial term induces hypoprobability as determined by the exponents i and $m - i$ in Eq. (7.33).

The preceding theoretical principles provide foundations for explaining initial and subsequent emergence of social complexity, as seen in the remaining sections of this chapter.

7.4 Explaining Initial Social Complexity

Amoebae, mammals, and entire biomes are living systems that form through different processes, just as planets, moons, stars, and galaxies are generated by different processes of formation. Different formative processes are explained in terms of different theories. At the same time, some general theories also exist to account for cross-level or multi-scale phenomena, such as gravitational theory, relativity theory, and the theory of evolution.

The same is true of social systems: different human aggregates require different theories to account for their formation. Chiefdoms, states, markets, trade networks, empires, and world systems are characterized by different formative processes for emergence of social complexity, some of which are better known than others.

In each case it is essential to understand exactly what is being explained: the **explanandum**. Chiefdoms, states, empires, and global systems are all instances of the class of complex social entities known as *polities*. Specifically, they are not "societies" or "cultures" (which are other, quite different, social entities), but *specific types of political systems* with distinct patterns of authority and government. In Chap. 2 we introduced the concept of a polity and examined it in some detail using UML diagrams to specify its constituent entities and associations. Now it is necessary to formalize some earlier definitions in order to provide a more rigorous theoretical explanation of *initial* social complexity (in this section) and its subsequent development (next).

Definition 7.4 (Polity) A polity is a complex adaptive system consisting of a society and a system (or subsystem) of government for managing collective issues that affect members of society in the normal course of history. Management of collective issues is done through public policies prepared, implemented, and monitored by government.

Understanding how and why a polity forms for the first time—i.e., **politogenesis**—requires what anthropologists call an **etic approach** and other social scientists call a **nomothetic approach**: a precise understanding of what a polity consists of (as well as what it is *not*)—including all main component entities and relations among components—and how it operates under a range of conditions or operating modes (stable, unstable, failing, recovering, failed). The etic approach has a universal, *erga omnes* orientation. Understanding any real-world polity also requires an **emic approach**, for mapping or "fitting" the theoretical model onto empirical data. Entities, variables/attributes are etic; instances and values are emic. The simplest polity is already complex, because of the presence of goal-seeking and adaptation, both nonlinear, not simple processes, as we have seen and will re-examine in greater detail below. Now, we need a more formal understanding grounded on etic-based theoretical principles.

From an emic perspective, polities in the initial epochs of social complexity, in all four regions discussed in Chap. 5, had the complete set of features in Definition 7.4—although many proper nouns and details remain unknown.

- *Mesopotamian* polities consisted of Sumerian, Elamite, and neighboring societies (Amorites, Gutians, among others) governed by assemblies of elites and rulers, who dealt with public issues such as flooding of the Tigris and Euphrates rivers, trade regulation, religious worship, industrial-scale textile production, and protracted border conflicts, among others. Some of the early capitals included 'Ubaid, Uruk, Susa, Choga Mish, and Arslantepe, among many others.
- In *northeast Asia*, Shang society and neolithic predecessors were ruled by elites who resided in superior dwellings and managed issues such as irrigation and salt mining, production of refined jade and, later, bronze artifacts, which required collective action. Early capital centers included Erlitou and Angyang.
- The earliest *South American* polities consisted of societies composed mostly of fishermen and later also farmers and artisans governed by leaders who managed public issues such as disasters caused by natural hazards (El Niño, earthquakes, flooding, and mudslides, among the most common). Aspero, Caral, and El Paraíso were among the earliest polity centers.
- In *Mesoamerica* the earliest polities consisted of pre-classic societies—such as Zapotec, Olmec, and Maya—in several regions of present-day Mexico, Guatemala, and Honduras, governed by chiefs and ruling elites who dealt with public issues such as endemic conflict (internal and external), natural hazards

(flooding, earthquakes, wildfires), and infrastructure systems (canals, terracing, among the earliest, followed by roads and urban sanitation infrastructure). San José Mogote, Monte Albán, San Lorenzo, La Venta, El Mirador, Copán, and Kaminaljuyu were among the earliest polity centers.

In summary, all four politogenic regions had identifiable societies, public issues, governments, and policies—all the components of the standard model of a polity—based on lines of evidence discussed earlier in Sect. 5.5.1. Collective action (examined more closely later in this chapter) for monumental works (agricultural, funerary, military, civic, or religious monumental structures), specialized production (initially fine pottery, jade, and bronze) requiring surplus production, trade networks, and increasingly organized conflict, with formal armies by the time of the first state formations, emerged in all four regions, as well as elsewhere in less complete form.

Another social science term for a polity is a **political system**, in the same sense as:

$$polity \equiv political\ system$$

$$society \equiv social\ system$$

$$economy \equiv economic\ system$$

In turn, each main component of a polity needs an explicit definition that is universally applicable across time and space. The following etic-based definitions—as the definition of a polity—are made empirically specific, or emic-based, as necessary.

Definition 7.5 (Society) A society is a collectivity of persons that interact through social relations and share one or more identities in common. Attributes of a society include its size, location, composition, identities, authorities, stratification, wealth, and associated statistics and distributions, including social network features.

Computationally, the state of society is defined by the tuple of societal attributes. In particular, the **level of stress** of a society is given by the effect of public issues, as defined below. Social **identity** (which can be kin-based, ethnic, linguistic, or geographic, among most common forms) determines **authority** or, in common language, "whom people listen to/obey." In any given society, multiple identities map onto multiple authorities, as in a bipartite graph, because identities and authorities are disjoint sets. The social entity "society" consists of individuals, groups, social relations, and norms; it does not include other entities, such as institutions of authority or government, which form part of a different component of a polity.[6]

[6] Advanced polities, such as democracies, also include intermediary structures (e.g., political parties, lobbying groups, labor unions) located between society and government. These are not required for explaining initial social complexity, so we examine them later.

The society of most early polities was rather uniform, but neighboring polities were often populated by culturally different societies. Sumerians dominated early Mesopotamian polities in West Asia, but neighboring polities to the East were populated by Elamite societies. In East Asia, the ancestors of what later became the Han people, as well as other neolithic cultures (e.g., Xinle, Yangshao, Dadiwan, Longshan, Dawenkow, Daxi, among others), composed the society of early polities. Names for pre-Moche societies of the earliest South American polities (Aspero, Caral, El Paraíso, among many others) are unknown. Zapotec, Olmec, Maya, and Teotiuacano societies populated the earliest Mesoamerican polities. The powerful polity of Teotihuacán had a heterogeneous, multi-cultural society, consisting of local and foreign residents (Maya, Zapotec, Otomi, Mixtec) in segregated neighborhoods. All early empires (Akkad, Shang, Moche, Teotihuacán) comprised heterogeneous societies.

Definition 7.6 (Public Issue) A public issue is a collective concern that affects members of a society in some consequential way, which can be positive (opportunities) or negative (threats, hazards).

Issues are *public*, as opposed to *private*, when they affect a collectivity of persons in a given society, as opposed to individual or internal household matters. The main effect that public issues have on society is to cause stress on one or more groups, which is a situational change that must be dealt with to eliminate or mitigate the stress. Public issues define the realm of the political and provide causal motivation for generating systems of government. Examples of public issues vary with epochs. Security, leadership succession, transportation, migrations, technological innovations, public health, and trade standards are among the oldest public issues that the earliest polities engaged with in all primary politogenic regions and elsewhere. Education, consumer protection, and management of the economy are more recent. *The need to solve public issues—to enjoy a better life—is the main generative source for first emergence and subsequent long-range evolution of social complexity*. Public issues justify government, which produces policies for managing issues.

Definition 7.7 (Government) The government of a polity consists of the organizational system of institutions and procedures for managing societal issues through public policies.

The association between society and government is known as **regime** or, more specifically, constitutional regime, because the relationship between society and its respective government is defined by constitutional code or custom. Democracy, dictatorship, and monarchy, are modes of regimes. From a computational object-oriented perspective, regime is an *association class* with encapsulated attributes such as

- `typeOfRegime` [string]
- `dateOfFormation` [tuple]
- `constitutionSource` [string]

- `legislativeInstitutions` [list]
- `executiveInstitutions` [list]
- `judiciaryInstitutions` [list]

among others, and operations (implementation, amendment, suspension, abrogation, and other).

From a governmental and computational information-processing perspective, chiefdoms have undifferentiated institutions of government (the chief or paramount leader carries out all functions of governance, with maximum centralization of information-processing), whereas states have specialized institutions (federated information-processing). Early forms of government in Sumerian polities included *assemblies* and *authoritative rulers*, and later, *bureaucracies* comprising *systems of public administration*. Similar forms emerged in East Asia, South America, and Mesoamerica. Governance and information-processing in all four areas, with the exception of South America, were supported by systems of writing (cuneiform, glyphs, and other logographic writing systems). Andean polities used a recording system for storing information called *quipu* since ca. 2000 BC (thus, *quipu* was invented much earlier than the Inca empire), consisting of sets of chords with knots denoting various base-10 values for encoding information. From a computational perspective, a system of writing provides much greater information-processing capacity, as well as memory, which explains the emergence of states concurrent with the invention of writing.

Definition 7.8 (Policy) A policy is a program of actions intended to manage (i.e., define and resolve or mitigate) a public issue.

Computationally, a policy is an *association class* with encapsulated attributes such as

- `targetIssue` [string]
- `targetSocialGroup` [string]
- `dateOfFormulation` [tuple]
- `dateInitialImplementation` [tuple]
- `cost` [int]
- `effectiveness` [float]
- `efficiency` [float]
- `popularity` [float]
- `implementingActors` [list]

among others, and operations such as `fundingRate()`, `changePopularity()`, and others. Trade policy was among the earliest forms of policy in primary polities, used for regulating commerce and possession of luxury items (precious and semi-precious stones and metals) and intended to provide rulers with unique control. Territorial deterrence and defense policies, first putatively under chiefdoms and later much more reliably under states, were also among the first policies to be enacted by rulers. Fiscal policies provided tax revenue and other forms of income to pay for other policies and government programs (e.g., infrastructure construction and maintenance, dating back to the earliest chiefdoms), including the cost of government itself.

As a social complexity event $\mathbb{P}$, policy requires conceptualization based on need ($\mathbb{X}_1$), design ($\mathbb{X}_2$), implementation ($\mathbb{X}_3$), monitoring ($\mathbb{X}_4$), and (optionally) adjustments ($\mathbb{X}_5$). This minimal, first-order, five-event ($\Lambda = 5$ causal requirements) generative process has event function $\Psi(\cdot)$ and probability equations given by

$$\mathbb{P} = \mathbb{X}_1 \wedge \mathbb{X}_2 \wedge \mathbb{X}_3 \wedge \mathbb{X}_4 \wedge \mathbb{X}_5 \tag{7.34}$$

$$\Pr(\mathbb{P}) = p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5 = \prod_{i=1}^{5} p_i \tag{7.35}$$

$$= P^5 \tag{7.36}$$

$$< \min\langle p_1, p_2, p_3, p_4, p_5\rangle \tag{7.37}$$

$$< P, \tag{7.38}$$

where P denotes the probability of sequentially conjunctive causal events taken across stages of the policy process. Chiefdoms and states have relatively low and high values of P, respectively, because of differences in policy-making capacity and reliability. For example, chiefdoms struggle to defend territory because they lack many of the attributes that states have: state rulers have access to more reliable intelligence, and a bureaucracy and system of public administration to support policy design, implementation, monitoring, and adjustments, to name a few. In developmental terms, chiefdoms and states are "rudimentary" and "mature" forms of complex adaptive systems, respectively.

States also have capacity to build redundancy into policies to increase their overall reliability. A state policy $\mathbb{P}^*$ with 2nd-order, ν-out-of-m partial redundancies for, say, implementation, monitoring, and adjustments, has event function $\Psi(\cdot)^*$ and probability equations given by

$$\mathbb{P}^* = \mathbb{X}_1 \wedge \mathbb{X}_2 \wedge \left(\bigvee_{i=\alpha}^{m} \mathbb{Z}_i\right) \wedge \left(\bigvee_{j=\beta}^{n} \mathbb{Z}_j\right) \wedge \mathbb{X}_5 \tag{7.39}$$

$$\Pr\big(\mathbb{P}^*\big) = \prod_{k=1,2,5} p_k \cdot \left[\sum_{i=\alpha}^{m} \binom{m}{\alpha} Q^i (1-Q)^{m-i}\right] \cdot \left[\sum_{j=\beta}^{n} \binom{n}{\beta} R^j (1-R)^{n-j}\right] \tag{7.40}$$

$$> \Pr(\mathbb{P}), \tag{7.41}$$

where Q and R are the disjunctive probabilities, and $\alpha < m$ and $\beta < n$ are binomial parameters for partial redundancies in implementation and monitoring, respectively. Note that the 1st-order conjunction in Eq. (7.40) still requires a product of all five policy process probabilities, which induces *overall hypoprobability*. However, implementation and monitoring redundancies, represented by the disjunctive-binomial expressions, induce some *local hyperprobability*, which is helpful and is entirely absent in the policies of chiefdoms due to their inferior capacity. This is one way in which greater complexity of a state produces higher policy performance (Eq. (7.41)). Imperial polities, characterized by quantum greater governance complexity, can attain extremely high levels of policy performance when operating at maximum capacities.

In the next sections we examine theories of social complexity pertaining to chiefdoms and states. Social complexity theories of empires represent an exciting but relatively undeveloped research frontier, especially from a CSS perspective.

7.4.1 Emergence of Chiefdoms

7.4.1.1 What Kind of Polity is a Chiefdom?

A chiefdom must be defined in sufficient scientific detail before explaining how one forms.

Definition 7.9 (Chiefdom) A chiefdom is a polity with stratified and ranked society (minimally elite members and commoners), public authority exercised by a chief (paramount leader, strongman) and subordinate village rulers (sub chiefs), and putative control over a regional territory comprising several villages.

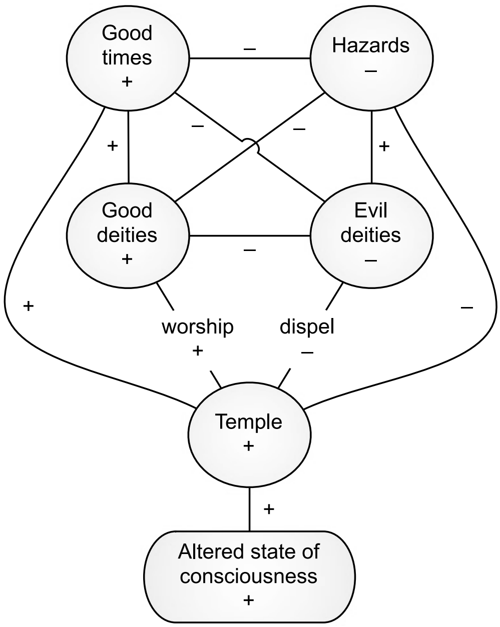
For commoners, chiefly authority is a function of local identity and ability to provide basic public goods (security, basic well-being). For elite members, it is based on rewards, as explained below. Territorial control by government (the chief) is putative, and not highly reliable (as for a state), due to lack of capacity to establish and defend boundaries. Chiefdoms lack permanently staffed institutions (public administration, judicial system, military forces, among others), but have specialized craftsmen that do not depend on elaborate supply chains for producing elite goods. Shamans or religious leaders (temple priests) specializing in spiritual life through private and public rituals are members of the non-commoner group.

Hunter-gatherer, pre-complex societies began building shrines—temporary, non-residential places of worship, often at remote locations—but not temples. Chiefly elites constructed temples for worshiping the community's deities, as in a three-way, "win-win-win" or mutually reinforcing triadic social relation:

1. Chiefs gain authoritative legitimacy and approval from commoners by building and dedicating temples; they also gain support from religious authorities as sponsors.
2. Commoners support temple projects because they provide a place of worship and a link to the afterlife, and because the community is reinforced and energized.
3. Priests play a key role as intermediaries between this life and the afterlife by constraining chiefs and elites, consoling commoners, and highlighting community identity.

Such a simple triad is also cognitively balanced, in the sense of Abelson (Sect. 4.8.1), and it supports the belief system in Fig. 7.1. The six nodes and eleven relations are all balanced, making this a powerful, stable, and shared belief system for the

Fig. 7.1 An Abelson-balanced belief system relating multiple aspects of communal worship



community—and the temple is the physical venue for the communal practice of worship.

As measured by the *Peregrine-Ember-Ember scale of social complexity* discussed earlier in Sect. 5.5.2, chiefdoms are empirically characterized by mostly sedentarism (nomadic chiefdoms do exist, as among steppe societies in Central and Inner Asia and desert regions of the Middle East, but are exceptional), inegalitarian (status or wealth) differences, population density greater than 1 person/mi^2, reliance on food production, and villages with population greater than 100 persons.

The *political economy* or *public finance* of a chiefdom has the following characteristics:

1. Coalition government: Government by the paramount chief depends on a *political coalition* with local chiefs who lead commoners at the village level. The paramount chief is not the sole ruler who governs in the polity. The governing coalition is the *main social artifact*, since a chiefdom lacks other institutions.
2. Side-payments: Every political coalition entails costs, both tangible and intangible. By Riker's theory, *side-payments* (gifts, bribes, rewards, honors, and other benefits) are used by the paramount chief to gain, maintain, or strengthen the allegiance of confederate subordinate chiefs and their villages.
3. Resource flows: Local village rulers exact taxes from commoners, keeping some for themselves and providing some to the paramount chief, while some is spent on local provision of private and public goods (e.g., building temples and defensive works).

4. Private property: Elites (secular and religious) have property rights over tangible (land, labor, animals, water wells, among others) and intangible (symbols, status, sacred attributes) resources, which they assert *vis-à-vis* commoners.
5. Interdependencies: Similar to the win-win-win triadic relationship noted earlier, a paramount chief depends on his ability to extract resources from subordinates, as well on his capacity to deliver public goods, such as defense and security against neighboring chiefdoms; on participation in periodic rituals and major events in the spiritual and social life of the community; and on administration of justice. Elite members and local chiefs depend partly on the paramount chief for their livelihood and prestige, and partly on local commoners. In turn, local commoners depend on rulers for defense against aggressors and for organizing other forms of collective action, including temple construction.
6. Monumental structures: Large-scale monumental structures in chiefdoms—such as construction of temples and spiritual structures that are perceived to provide rewards in the afterlife, or utilitarian infrastructure such as irrigation systems for agriculture or flood control systems—are financed by forced and voluntary labor.
7. Energy budget (energetics): A chiefdom must maintain a neutral (minimally) or positive (preferably) energy balance in order to be sustainable, just as in any other polity. In particular, food production (agricultural, maritime, or foraged) must yield sufficient surplus to support all persons who are not producing, such as rulers and all elite members, craftsmen, and clergy.[7]
8. More public structures: A corollary of this is the construction of communal artifacts such as storage facilities, and defensive structures to protect increased wealth coveted by neighboring chiefdoms.
9. Environmental conditions: Features of the natural environment, including natural hazards present in the region, provide costs/threats and benefits/opportunities that are an integral part of the overall political economy of a chiefdom. Some of these are fixed, others are variable; some are periodic, others are random.
10. Precious stones and metals: Control over exotic materials, such as precious stones (jade, turquoise, obsidian, lapis lazuli, carnelian, pearls) and metals (gold, silver, copper), is sought by chiefs because these materials provide distinction and are also used as rewards for obedient subordinates. Elaborate forms of these materials, such as jewelry and other status symbols, require provisioning of raw materials, specialized craftsmen, secure workshops, inventory control, and viable distribution.

A chiefdom, unlike a state, lacks a palace for rulers because while a chief can get commoners and allied elites to finance and build a temple, a chief lacks sufficient power to have them build a palace for himself and his entourage. Temples are dedicated to local deities, so they belong to the community, not to the chief. The palace

[7] The term energetics is used in archaeology to demote the caloric budget of a community in terms of energy produced and consumed. For example, a community of a given size, producing so many surplus tons of barley per year, is able to build during so many days of the year. Conversely, when archaeological excavation reveals a given number of structures, the total energy necessary to construct them must be accounted for in terms of population available and food to sustain the required labor force.

would be different, because as a private dwelling, in addition to being a place of public government business, it would belong to the paramount chief and his family and friends. That requires a state-level polity, as we shall see in the next section.

A **simple chiefdom** has a minimal version of all the features we have discussed so far: a few villages distributed in a relatively small territory, totalling around 1,000 inhabitants or less, with governance provided by a strong leader and subordinate confederates. Basic artifacts include rudimentary defensive structures (moats, berms, ditches, palisades), a temple in the paramount chief's village (smaller ones are also possible in other villages), and a small political coalition as the sole institution to support governance. A **complex chiefdom** will have an additional level of elite hierarchy, which acts as a multiplier of social complexity in the polity, while still lacking specialized institutions or permanent bureaucracy. Both kinds of chiefdom have temples; neither has palaces.

Finally, all chiefdoms are *unstable* polities, which cycle through integration and disintegration for multiple reasons:

- The paramount chief has to struggle constantly to secure resources necessary to provide side-payments for confederate chiefs; otherwise the coalition may fall apart.
- Subordinate chiefs decide strategically, so they may change allies, causing civil war.
- Reserves and other resource buffers for ensuring against inevitable and unpredictable natural hazards (droughts, floods, mudslides, earthquakes, El Niño) are unreliable, when they exist at all.
- Neighboring chiefdoms pose a constant threat through raids and attempted conquests.
- Rulers depend on the manipulation of the spiritual realm and communal deities, as well as priestly consent, to maintain authority.
- Absence of permanent institutions makes all governmental operations precarious at best, including the assertion of elite property rights.

From this theoretical perspective, chiefdoms are always in a **metastable state**, on the brink of either disintegrating, being conquered by a neighbor, or—in rare cases—undergoing a phase transition in a state-level polity by conquering other neighboring chiefdoms (discussed in the next section). In short, *in a chiefdom there is never sufficient reliable capacity for managing emerging public issues with high probability of success*.

7.4.1.2 How do Chiefdoms Emerge?

Politogenesis—the first emergence of chiefly social complexity—is explained by considering antecedent conditions and their realization through time, including systematic, specifiable sets of conjunctive and disjunctive events in the causal process of polity formation. Specifically, the dynamic phase transition from a simple, purely kin-based society to initial social complexity at time τ involves the *realization of a potential* that developed in such societies at time $\tau - \Delta\tau$.

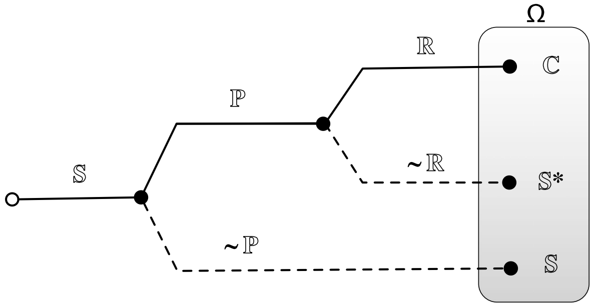


Fig. 7.2 Forward sequential causal logic tree for initial sociopolitical complexity, denoted as the contingent process $\mathscr{P}_3(\Omega)$ of politogenesis with three antecedents

Assumption 7.3 (Potential as Antecedent Condition) *Initial social complexity occurs (event $\mathbb{C}$) if and only if (i.f.f.) a prior* potential $\mathbb{P}$ *emerges or forms in the state space S of a previously simple society, such that $\mathbb{C}$ occurs when $\mathbb{P}$ is* realized. *Conversely, complexity cannot occur without prior formation and subsequent realization of an associated potential.*

Figure 7.2 shows the forward-sequential causal logic tree for occurrence of initial complexity $\mathbb{C}$ within the social outcomes space Ω, according to Assumption 7.3. Given a society in a simple state with only kin-based organization (event $\mathbb{S}$), the potential for sociopolitical complexity may or may not occur (events $\mathbb{P}$ and $\sim\mathbb{P}$, respectively). If $\sim\mathbb{P}$, then the potential cannot be realized (since it does not exist) and the outcome in Ω is that society does not change. If $\mathbb{P}$ occurs, in terms of knowledge and ability conditions 1–9 (examined below), then such a potential may or may not be realized (events $\mathbb{R}$ and $\sim\mathbb{R}$, respectively). If $\sim\mathbb{R}$, then the polity becomes and remains metastable (event $\mathbb{S}^* \in \Omega$). If $\mathbb{R}$, then the outcome is the occurrence of initial sociopolitical complexity (event $\mathbb{C}$).

Thus, $\Omega = \{\mathbb{S}, \mathbb{S}^*, \mathbb{C}\}$, where $\mathbb{C}$ is an outcome in the contingent process $\mathscr{P}_3(\Omega)$ of politogenesis with three antecedents. In causal logic form, initial complexity $\mathbb{C}$ at time τ implies an associated prior potential $\mathbb{P}$ at some prior time $\tau - \delta$ as a necessary condition, $\mathbb{C}(\tau) \Rightarrow \mathbb{P}(\tau - \delta)$, but not conversely, for some $\delta < \tau$.

Assumption 7.4 (Potential as Compound Event) *The emergence of potential for initial social complexity $\mathbb{P}$ is a compound event, not a singleton or elementary event, as specified by an event function $\Psi(\cdot)$ in terms of a set $\{\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \ldots, \mathbb{X}_n\}$ of more elementary events causally linked to the occurrence of $\mathbb{P}$. Formally,*

$$\Psi : \mathbb{P} \Leftarrow \{\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \ldots, \mathbb{X}_n\}, \tag{7.42}$$

so

$$\mathbb{P} = \Psi(\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \ldots, \mathbb{X}_n), \tag{7.43}$$

where $\mathbb{X}_i$ denotes the ith causal event for $i = 1, 2, 3, \ldots, n$.

Assumptions 7.3 and 7.4 lead to the following two key questions:

1. Exactly what constituted such potential for initial social complexity?
2. Under which conditions would such potential be realized?

The first question translates into: What knowledge and abilities did community members have *before* they formed the simplest chiefdoms? What did they *have* to know? The following minimal ensemble of necessary conditions (*conditio sine qua non*) possessed by members of simple bands in pre-complex societies created the *potential*—albeit not the certainty—for emergence of initial social complexity:

1. **Kinship knowledge**. People had knowledge of their kin, which supported extended households beyond a family nucleus, as well as enabling collective action based on deontic (obligation-based) norms or for advancing other goals.
2. **Communicative ability**. Humans began using language to communicate between ca. 100,000 and 50,000 years ago. Communicative ability was necessary for collective action (both planning and execution), such as in large-scale hunting.
3. **Normative sociality**. Cooperative social norms were known to people in pre-complex societies via biological evolution, specifically norms of kin selection and reciprocal altruism.
4. **Social identification ability**. The ability to classify others into in-group vs. out-group status was essential for detecting potential threats and opportunities, as well as for norm use or invocation. In-out group identification generated cognitive complexity and balancing.
5. **Environmental knowledge**. Awareness concerning the biophysical landscape was necessary for finding resources and detecting significant change, such as in local species, "normal" climate, and other aspects.
6. **Knowledge of normal vs. rare events**. Ability to detect situational change, such as emergent threats or opportunities, beyond the biophysical environment, was necessary for assigning levels of urgency, significance, or priority.
7. **Food procurement ability**. Hunting, gathering, fishing, herding, farming, or preying on others (stealing) was necessary for maintaining sustenance throughout seasons of the year and longer time spans, especially in temperate regions far from the Equator, where seasonal variations determine the basic food supply.
8. **Homicidal ability**. Originally derived from the hunting skill-set, homicidal ability was a necessity in some modes of collective action (while remaining a tabu among group members), such as when facing lethally aggressive adversaries. Deterrence also requires credible homicidal action.
9. **Collective action ability**. People knew how to organize for collective action (i.e., how to lead and how to follow, and other modes of collective action) before chiefdoms formed. Collective action was invented and perfected through ancient activities such as hunting large mammals.

None of these abilities or kinds of knowledge *per se* necessarily produced social complexity; they were merely abilities among others. Also, not all ancient societies met these conditions everywhere at the same time. In fact, in vast areas of the world these conditions were never met, or were met much later.

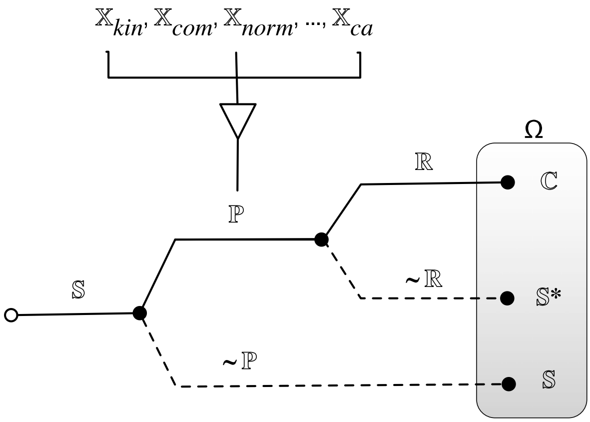


Fig. 7.3 Forward sequential causal logic tree for initial politogenesis $\mathbb{C}$ grafted with a first-order backward conditional causal tree for complexity potential $\mathbb{P}$ (Conditions 1–9; Sect. 7.4.1.2)

Assumption 7.5 (Specific Requirements for Chiefdom Formation) *The event function Ψ for the compound event $\mathbb{P}$ includes the following minimally necessary causal events $\mathbb{X}_i$ of required knowledge and abilities* (*conditions* 1–9 *detailed above*):
1. $\mathbb{X}_{kin}$ = *Kinship knowledge*,
2. $\mathbb{X}_{com}$ = *Communicative ability*,
3. $\mathbb{X}_{norm}$ = *Normative knowledge*,
4. $\mathbb{X}_{id}$ = *Social identity knowledge*,
5. $\mathbb{X}_{env}$ = *Environmental knowledge*,
6. $\mathbb{X}_{rare}$ = *Knowledge of normal vs. rare events*,
7. $\mathbb{X}_{food}$ = *Food procurement ability*,
8. $\mathbb{X}_{kill}$ = *Homicidal ability, and*
9. $\mathbb{X}_{ca}$ = *Collective action ability*.

Based on these assumptions, the potential $\mathbb{P}$ for chiefdom formation is given by the conjunctive event equation

$$\mathbb{P} = \Psi(\mathbb{X}_{kin}, \mathbb{X}_{com}, \mathbb{X}_{norm}, \ldots, \mathbb{X}_{ca}), \tag{7.44}$$

$$\Leftarrow (\mathbb{X}_{kin} \wedge \mathbb{X}_{com} \wedge \mathbb{X}_{norm} \wedge \cdots \wedge \mathbb{X}_{ca}), \tag{7.45}$$

which specifies the conjunction ($\bigwedge_i \mathbb{X}_i$) of causal events that generate $\mathbb{P}$. Equation (7.45) is used in Fig. 7.3, which extends Fig. 7.2 by specifying preconditions 1–9 for $\mathbb{P}$.

Similarly, event $\mathbb{R}$, which consists of the actual realization of $\mathbb{P}$, (see Fig. 7.2), is specified by the conjunctive event equation

$$\mathbb{R} = \mathbb{O} \wedge \mathbb{W} \wedge \mathbb{I}, \tag{7.46}$$

where $\mathbb{O}$, $\mathbb{W}$, and $\mathbb{I}$ denote the occurrence of willingness, opportunity, and implementation.

Theorem 7.13 (First-Order Probability of Chiefdom Formation) *Let* $X = \Pr(\mathbb{X})$. *The probability of initial social complexity* (*event* $\mathbb{C} \in \Omega$ *in Fig.* 7.3) *is given by*

$$C = S \cdot P \cdot R = \prod_{i=S}^{R} X_i \tag{7.47}$$
$$= c^3, \tag{7.48}$$

where c denotes some uniform probability on the closed interval [0, 1] *taken across causal events* $\mathbb{S}$, $\mathbb{P}$, *and* $\mathbb{R}$.

Theorem 7.14 (Probability of Potential for Chiefdom Formation) *The probability of potential for politogenesis* $\mathbb{P} \in \mathscr{P}_3(\Omega)$ *as a function of first-order causal events* $\mathbb{X}_i$ (*Assumption* 7.5) *is given by*

$$P = X_{kin} \cdot X_{com} \cdot X_{norm} \cdots X_{ca} = \prod_{i=kin}^{ca} X_i \tag{7.49}$$
$$= x^{\Theta}, \tag{7.50}$$

where x denotes some uniform probability taken across Θ *causal events*, *which are nine assuming* $\mathbb{X}_{kin}$ *to* $\mathbb{X}_{ca}$ (*causal necessary conditions* 1–9, *Assumption* 7.5).

Theorem 7.15 (Probability of Realization) *The probability of realizing a politogenic potential* $\mathbb{R} \in \mathscr{P}_3(\Omega)$ *as a function of first-order causal conditions for opportunity* $\mathbb{O}$, *willingness* $\mathbb{W}$, *and implementation* $\mathbb{I}$ *is given by*

$$R = O \cdot W \cdot I \tag{7.51}$$
$$= r^3, \tag{7.52}$$

where r denotes some uniform probability taken across $\mathbb{O}$, *willingness* $\mathbb{W}$, *and implementation* $\mathbb{W}$ *events*.

The following second-order principles extend previous principles. These are stated in terms of more specific causal events, which is useful because second-order conditions are closer to observation and operational events than the more abstract, theoretical first-order conditions.

Theorem 7.16 (Second-Order Probability of Chiefdom Formation) *The second-order probability of initial chiefdom formation is given by*

$$\mathbb{C} = \Psi(\mathbb{S}; \mathbb{X}_{kin}, \mathbb{X}_{com}, \mathbb{X}_{norm}, \ldots, \mathbb{X}_{ca}; \mathbb{O}, \mathbb{W}, \mathbb{I}), \tag{7.53}$$
$$\Leftarrow \langle \mathbb{S} \wedge (\mathbb{X}_{kin} \wedge \mathbb{X}_{com} \wedge \mathbb{X}_{norm} \wedge \cdots \wedge \mathbb{X}_{ca}) \wedge \mathbb{O} \wedge \mathbb{W} \wedge \mathbb{I} \rangle, \tag{7.54}$$

and

$$C = S\left(\prod_{i=kin}^{ca} X_i\right) O \cdot W \cdot I \tag{7.55}$$
$$= y^{\Gamma}, \tag{7.56}$$

where y is some uniform probability taken across the set of Γ second-order causal events for $\mathbb{C}$ and $\Gamma > \Theta$.

Note that $\Gamma = 13$ in Eq. (7.56), w.r.t. second-order conditions. In fact, $\Gamma \gg 13$, because much more conjunction is involved before reaching the operational behavioral level. For example, implementation $\mathbb{I}$ is itself a *compound event* (i.e., actually enforcing elite property rights, creating the chiefly coalition, building the temple, and other necessary and difficult collective action strategies that the chief, elites, and commoners must accomplish) produced by highly contingent processes. Theorem 7.16 explains why politogenesis was such a rare occurrence in history (early Holocene). Since $\Gamma \gg \Theta + 4$ (by Eq. (7.54)) and $\Theta = 9$, it follows that $C(y; \Gamma) \ll y^{13}$, which yields a relatively minuscule probability C of chiefdom formation for arbitrary values of y. If Assumption 7.5 is incomplete ($\Theta > 9$), then politogenesis, as well as its potential, are even rarer events!

The following sensitivity results follow from multivariate analysis of the preceding principles.

Theorem 7.17 (Gradient of the Potential for Chiefdom Formation) *The gradient of the probability P of potential for politogenesis is given by*

$$\nabla P = \Theta x^{\Theta-1}\hat{\mathbf{x}} + \left(x^{\Theta+1} - x^{\Theta}\right)\hat{\boldsymbol{\vartheta}}, \tag{7.57}$$

so P is increasing in x and decreasing in Θ; and

$$|\nabla P| \approx \Theta x^{\Theta-1}, \tag{7.58}$$

so ∇P points mainly in the direction of x.

Theorem 7.18 (Gradient of the Probability of Chiefdom Formation) *The gradient of the probability C of chiefdom formation is given by*

$$\nabla P = \Gamma y^{\Gamma-1}\hat{\mathbf{y}} + \left(y^{\Gamma+1} - y^{\Gamma}\right)\hat{\boldsymbol{\gamma}}, \tag{7.59}$$

so C is increasing in y and decreasing in Γ; and

$$|\nabla C| \approx \Gamma y^{\Gamma-1}, \tag{7.60}$$

so ∇C points mainly in the direction of y.

7.4.2 Emergence of States

We now turn attention to a more precise specification of the state polity and the theoretical explanation of its primary formation. This section parallels the earlier analysis of chiefdoms.

7.4.2.1 What Kind of Polity Is a State?

"A state is not a chiefdom on hormones," the American archaeologist Joyce Marcus (1992) once wrote.

Definition 7.10 (State) A state is a polity with a stratified and ranked society (elite members, civil servants, traders, military, and commoners), a system of government composed of specialized, differentiated institutions with authoritative decision-making, capacity to collect taxes as government revenue, and reliable control over territory and its resources.

Stately authority is a function of local identity, monopoly over the use of force, and ability to reliably provide public goods beyond defense and security. Government offices are held through ascriptive (hereditary) as well as achieved (meritocratic) modes. Territorial control by government is dependable, enforced by permanent standing military forces, and highly reliable (unlike a chiefdom), due to sufficient capacity to defend boundaries. States have permanently staffed institutions (public administration, judicial system, military forces, among others), and industrial organizations with specialized craftsmen that are dependent on supply chains for producing elite and utilitarian goods. Religious leaders (temple priests) are generally also part of the elite, non-commoner group, but play a less essential role than in a chiefdom, due to greater political autonomy of state rulers and institutions relative to chiefdom rules.

As measured by the *Peregrine-Ember-Ember scale of social complexity*, states are empirically characterized by metal production, social classes, towns with more than 400 persons, three or more levels of settlement hierarchy, population density > 25 person/mi^2, wheeled transport, writing of any kind, and money of any kind.

The *political economy* or *public finance* of a state has the following characteristics, which are fundamentally different from a chiefdom:

1. Public issues: Members of society who live in a state have an expectation that government policy will address public issues; this expectation generally increases over time (well-being has positive feedback), rather than decreases.
2. Policymaking: Problem-solving to address public issues through policies is a formal, institutionalized process.
3. Coalition: Rulers still depend on a support coalition, often called nobility, with side-payments provided to supporters, but on a scale greater than in chiefdoms.
4. Taxation: Government operations are financed primarily by tax revenues extracted from commoners and merchants, as well as by the spoils of warfare (e.g., used for paying military forces).

5. Bureaucracy: The system of public administration plays a critical role in the provision of public goods and in tracking state revenue streams.
6. Cost of government operations: Maintaining a ruling elite that decides policy and a bureaucracy that implements it is a permanent, recurring cost that somehow must be financed.
7. Private property: Right over private property is enforced by rule of law and judiciary institutions.
8. Interdependencies: The state leader (now a king, as opposed to a paramount chief) depends on his ability to extract resources from the nobility in exchange for titles and rights, as well as on the capacity of government to deliver an array of public goods (defense, justice, public sanitation, policing, markets, roads, port facilities). Members of the nobility depend partly on the state leader for their livelihood and prestige, and partly on compliance from local commoners. In turn, commoners depend on members of the local nobility for policing, defense against aggressors, and for organizing other forms of collective action, including public works.
9. Monumental structures: Large-scale monumental structures in states are created by paid and forced labor (including slaves, captives). These include of palaces and monumental tombs, road networks, aqueducts, military fortifications of many kinds (from sophisticated and massive city walls to regional frontier walls still visible from space), industrial factories (e.g., bronze, requiring complex supply chains and thousands of workers and specialized managers), among the most costly. Temples and spiritual structures are not neglected by the state; they are built bigger, since they are still perceived to provide rewards in the afterlife.
10. Energy budget (energetics): Food production in a state polity is organized to yield surplus on a large scale, because the number of persons who are not producing food is a much greater proportion of the population.[8]
11. More public structures: A corollary of the above feature is the construction of non-residential office space in palaces to support operations of public administration, judicial courts, and military barracks and forts.
12. Environmental conditions: The environment, including natural hazards in the state territory, has even greater significance, because of greater population size and increasingly complex infrastructure systems exposed to a broader spectrum of risks; some of them interdependent or "cascading," linked via infrastructure.
13. Precious stones, metals, textiles: Consumption of jewelry, all forms of elaborate ornaments, and sumptuous clothing by state elites (secular, military, and religious) is exuberant, compared to wares in chiefdoms. All of these must be financed.
14. Military expenditures: The cost of a permanent, on-demand military force (personnel, armor, weapons, facilities) is a major component of a state budget.

[8] Hence the significance of the invention of agriculture and related technologies (e.g., official seals, measures, laws).

Indeed, paraphrasing Marcus, a state is *quantum more* than a chiefdom! The palace of rulers is diagnostic of a state polity, as is the population settlement hierarchy in three or more levels, and other large-scale complex artifacts, such as government bureaucracy and infrastructure systems.

An **archaic state** generally refers to primary and secondary states, in a chronological sense, as well as subsequent feudal states. A **modern state** refers to state polities beginning during the early modern period of European history, or what is known in the World History tradition as the end of the Postclassical Period (500–1500) and the start of the Early Modern Period (1500–1800). Both kinds of states have palaces, bureaucracies, tax and legal systems, territorial control, and monopoly over use of force, unlike chiefdoms, which lack all of these.

Finally, all states *can* be *stable* polities (chiefdoms cannot, for reasons already discussed), but they can also cycle through integration and disintegration for multiple reasons:

- Growth of the bureaucracy can bankrupt the budget of the state.
- Mass movements can detract legitimacy of governmental authority, toppling a regime.
- Rebellion in one or more provinces can fragment a state.
- Natural disasters can cause irreparable damage to infrastructure and bring about regime collapse.
- Neighboring polities pose a constant threat through raids and attempted conquests.
- Invasion and conquest by more powerful rivals can end in subjugation.
- Corruption, failure in rule of law, and other institutional pathologies can bring about state failure.

From this theoretical perspective, a state can be either stable (avoiding the above hazards), unstable/metastable, failing/collapsing, or failed/collapsed. In short, *in a stable state there is sufficient reliable capacity for managing emerging public issues with high probability of policy success*.

7.4.2.2 How Do States Emerge?

Social science has produced more theories of the origin of the state (both archaic and modern) than of other ordinal ranks of complex polities, such as chiefdoms, empires, or world government. The following two theories of state formation are among the better known.

Carneiro's Theory of Circumscription (1970). This theory explains state formation as resulting from warfare among small villages, and eventually among chiefdoms, under conditions of *circumscription*.[9] Success in agricultural production enabled demographic growth, requiring more arable land, as in a positive feedback process producing increasing pressure for territorial expansion. When a circumscribed society is attacked by another seeking scarce land to

[9] A polity is said to be circumscribed when surrounding territories prevent migration in time of crisis. Circumscription may be caused by neighboring mountains (Peru's Andean coastal region), deserts (Near East west of the Tigris-Euphrates basin), and similar obstacles.

cultivate (a common condition among chiefdoms), the defender is either victorious or defeated, being unable to escape. The winner either destroys or subjugates the vanquished, through a process of fusion until "the political unit thus formed [*sic*] was undoubtedly sufficiently centralized and complex to warrant being called a state" (p. 736).

The idea of a positive feedback process between agricultural success (food surplus, wealth) and demographic growth had been formally theorized by N. Rashevsky since 1947. Carneiro's theory makes a valuable contribution by highlighting the role of circumscription in preventing migration. However, the theory is deficient in specifically explaining how "the political unit thus formed." A winner could just become a bigger chiefdom, not a state. How do the institutions of a state emerge? Carneiro's theory does not explain the critical organizational difference between a chiefdom and a state, but simply views the latter as a larger version of the former.

Marcus's Dynamic Model (1989, 1992, 1998). Prior to the formation of a state in a given region, there exist chiefdoms with local populations governed through two or at most three levels of hierarchy, corresponding to simple and complex chiefdoms (as discussed earlier in Sect. 7.4.1.1). Competition and rivalries among chiefdoms cause conflicts that result in some chiefdoms growing more than others. At some point this process leads to the largest complex chiefdom in a region annexing its weaker neighbors and creating an additional level of hierarchy to control the conquered chiefdoms. The new state—a four- (possibly five-) level regional system—is composed of provinces consisting of former simpler chiefdoms, and the state capital is the central place of the former complex chiefdom.

Marcus's theory, which has been demonstrated for multiple regions around the world, uses the same conflict-ridden overture as Carneiro's, but the theory explains more because it tells us why a state generates more levels of governance and public administration than a chiefdom. It is because the aggregation of former chiefdoms requires one or two new levels of government in order to reliably consolidate and regulate its functions. Attention to institutional development marks theoretical progress. A key aspect that remains unexplained by both Carneiro's and Marcus's earlier theories is functional differentiation in the institutions of a state.

Most theories assume that a set of neighboring chiefdoms exists in a given region prior to a state forming, consistent with the archaeological record. However, chiefdoms were challenged by many other public issues besides conflict, such as natural hazards and endogenous stresses. This and related ideas are examined in this section within the formal Theory of Politogenesis and later as part of the more general Canonical Theory.

All known cases of primary states have emerged from regional systems of unstable rival chiefdoms. Politogenesis of states—the first emergence of stately social complexity—is also explained by considering antecedent conditions and their realization through time, including systematic, specifiable sets of conjunctive and disjunctive events in the causal process of polity formation. In this case, the dynamic phase transition from chiefly to stately social complexity at a given time τ involves the *realization of a potential* that developed during the chiefdom phase at time

$\tau - \Delta\tau$. Accordingly, the same theoretical framework we have already discussed (Assumption 7.3) holds true for explaining and understanding state formation.

In a chiefly society at time $\tau - \Delta\tau$, the potential for state-level complexity may or may not occur (events $\mathbb{P}_s$ and $\sim \mathbb{P}_s$, respectively). If $\sim \mathbb{P}_s$, then the potential cannot be realized (since it does not exist) and the outcome in Ω_s is that the polity does not change. If $\mathbb{P}_s$ occurs, this time in terms of additional, state-relevant knowledge and ability conditions 1–15 (examined below), then the potential may or may not be realized (events $\mathbb{R}_s$ and $\sim \mathbb{R}_s$, respectively). If $\sim \mathbb{R}_s$, then the polity becomes and remains metastable as a chiefdom (event $\mathbb{C}^* \in \Omega$). If $\mathbb{R}_s$, then the outcome is the occurrence of a phase transition into state-level sociopolitical complexity (event $\mathbb{S}$). Note that in this section $\mathbb{S}$ denotes the event of *state formation*, not a simple, pre-chiefdom polity in regard to chiefdom formation.

Thus, now $\Omega = \{\mathbb{C}, \mathbb{C}^*, \mathbb{S}\}$, where $\mathbb{S}$ is an outcome in the contingent process $\mathscr{P}_3(\Omega)$ of state formation with three antecedents. In causal logic form, state formation $\mathbb{S}$ at time τ implies an associated prior potential $\mathbb{P}_s$ at some prior time $\tau - \delta$ as necessary condition, $\mathbb{S}(\tau) \Rightarrow \mathbb{P}(\tau - \delta)$, but not conversely, for some $\delta < \tau$.

Following Assumption 7.4, the potential for state formation is similarly assumed to be compound, not a singleton or elementary event, as specified by an event function $\Psi(\cdot)$ in terms of a set $\{\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \ldots, \mathbb{X}_n\}$ of more elementary events causally linked to the occurrence of $\mathbb{P}$. Formally,

$$\Psi_s : \mathbb{P}_s \Leftarrow \{\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \ldots, \mathbb{X}_n\}, \tag{7.61}$$

so

$$\mathbb{P}_s = \Psi_s(\mathbb{X}_1, \mathbb{X}_2, \mathbb{X}_3, \ldots, \mathbb{X}_n), \tag{7.62}$$

where $\mathbb{X}_i$ denotes the ith causal event and $i = 1, 2, 3, \ldots, n$.

So we must now ask

1. What constitutes the potential for state formation?
2. Under which conditions would such a potential be realized?

Again, the first question translates into: What knowledge and abilities did community members in a chiefdom have *before* they established the first states?

Life in a chiefdom, especially a successful one that eventually evolved into a state (very few of them did!), produced numerous quantum gains in knowledge, abilities, and institutions of community members, both rulers and commoners.

1. **Non-kinship knowledge**. Beyond kinship knowledge, people had knowledge of significant non-kin members of society and government, especially chiefs (paramount and local) and priests (or shamans).
2. **Strategic ability**. Based on their coalition-based government experience, leaders (paramount and local, as well as priests) possessed *strategic ability*; i.e., they understood the interdependent nature of outcomes, including strategic signaling, as we would say today in game-theoretic terminology.
3. **Commons sociality**. Living in a chiefly village community, people understood the basics of *The Tragedy of the Commons*, including the role of *sanctions* for maintaining cooperation in the use of *common pool resources* (pastures, rivers,

wells, defensive structures) and inter-personal record-keeping in some unwritten form, all of which contributed to public administration skills, even in embryonic or rudimentary form.

4. **Residential skills**. Life in a chiefly village was in permanent house dwellings (regardless of building quality, structure, or materials: round or square; sunken, level, or raised; poles or bricks), not temporary hunter-gatherer camps. This also implied knowledge of basic sanitation needs and related infrastructure, including communal systems for waste management, such as ditches and piped drainage systems.
5. **Conflict memory**. Having experienced conflicts with neighboring chiefdoms, people in pre-state societies knew how to classify friends and foes with some precision, which was a significant refinement beyond the simpler in-group vs. out-group classification of more primitive societies.
6. **Environmental engineering knowledge**. Beyond empirical environmental knowledge, chiefly societies possessed environmental engineering knowledge in the form of animal exploitation, agriculture, and related engineered structures (e.g., communal irrigation systems, terracing).
7. **Village security ability**. Ability to defend against raids, even if not always succeeding, nonetheless produced significant skills in military affairs. Planning, building, and maintaining permanent defensive structures—such as palisades, ditches, baffled gates, towers, berms, bridges, raised roads—was common in many chiefdom villages.
8. **Food-processing ability**. Village dwellers processed food by blending and cooking ingredients procured through hunting, gathering, fishing, herding, farming, or preying. Food-processing required portable as well as permanent utilitarian artifacts, such as sieves, and ovens, and the knowledge to design, build, and maintain them.
9. **Military ability**. Raiding was common in chiefdoms, and the most successful chiefdoms raided better and were capable of conquering and absorbing neighboring chiefdoms through superior strategy, tactics, and logistics, even if at a rudimentary level.
10. **Complex collective action ability**. Pre-complex forms of collective action, used for hunting large animals, were significantly perfected by chiefly societies. Large monumental construction, elaborate rituals, communal feasting, basic communal sanitation, and effective raiding, among other activities required in a sustainable village, all required increasingly sophisticated skill in planning and executing collective action—quantum more than for hunting large animals.
11. **Supply chains**. Some activities required supply chains, in addition to collective action, such as in the construction of large monumental structures. Supply chain management also required discipline, precision, and coordination in the public domain, as well as planning, execution, and maintenance.
12. **Political autonomy**. Village dwellers, both rulers and commoners, became accustomed to enjoying political autonomy as a whole society, with "home rule," so to speak. They did not answer to any higher polity authority beyond their own local and paramount chiefs and community priests.

13. **Political culture**. Village life also produced specific instances of political culture, which is a community's shared set of values, beliefs, expectations, and practices with regard to what is just, proper, and taboo in all aspects of private and (especially) public life. Human sacrifice, slavery, revenge, obedience to authority, cannibalism, gifting, the paramount as chief judge, and sumptuous feasting were features of chiefly political culture, with local (emic) variations.
14. **Private property**. Elites enjoyed private property, including slaves, land, buildings, and livestock, so villagers gained familiarity with the idea and practice of private property, including bargaining and negotiation in resolution of claims, adjudication, and compensation.
15. **Chronic stress**. Life in a chiefly village community was highly stressful, due to the unstable nature of the polity, constant warfare with neighbors, insufficient food surplus to support more needed collective activities, and unsolved collective action problems that required political solutions on a broader regional scale than rulers and commoners were able to provide (environmental degradation, endemic warfare, migrations, natural disasters, among others).

Societies with these and related kinds of knowledge and abilities did not automatically evolve into states, but all those who did possessed these capabilities *because they were necessary*. Creating a state, based on this prior potential, requires creative use of these and other conditions. The exact number of initial conditions is not essential; what matters is that they are multiple *and* finite.

Assumption 7.6 (Specific Requirements for State Formation) *The event function Ψ_s for the compound event $\mathbb{P}_s$ includes the following minimally necessary causal events $\mathbb{X}_i$ on required knowledge and abilities* (*conditions* 1–15 *detailed above*):

1. $\mathbb{X}_{nonkin} =$ *Non-kinship knowledge*,
2. $\mathbb{X}_{strategic} =$ *Strategic ability*,
3. $\mathbb{X}_{commons} =$ *Commons sociality*,

$\vdots$

15. $\mathbb{X}_{stress} =$ *Chronic stress condition*.

Based on these assumptions, the potential $\mathbb{P}_s$ for state formation is given by the conjunctive event equation

$$\mathbb{P}_s = \Psi_s(\mathbb{X}_{nonkin}, \mathbb{X}_{strategic}, \mathbb{X}_{commons}, \ldots, \mathbb{X}_{stress}), \tag{7.63}$$

$$\Leftarrow (\mathbb{X}_{nonkin} \wedge \mathbb{X}_{strategic} \wedge \mathbb{X}_{commons} \wedge \cdots \wedge \mathbb{X}_{stress}), \tag{7.64}$$

which specifies the conjunction ($\bigwedge_i \mathbb{X}_i$) of causal events that generate $\mathbb{P}_s$. Equation (7.64) yields a forward sequential logic tree similar to Fig. 7.2 by specifying preconditions 1–15 for $\mathbb{P}_s$.

Similarly, event $\mathbb{R}_s$, which consists of the actual realization of $\mathbb{P}_s$ for state formation, is specified by the conjunctive event equation

$$\mathbb{R}_s = \mathbb{O} \wedge \mathbb{W} \wedge \mathbb{I}, \tag{7.65}$$

where $\mathbb{O}$, $\mathbb{W}$, and $\mathbb{I}$ denote the occurrence of willingness, opportunity, and implementation of state formation, given a prior potential $\mathbb{P}_s$.

Theorem 7.19 (First-Order Probability of State Formation) *Let* $X = \Pr(\mathbb{X})$. *The probability of state-level complexity* (*event* $\mathbb{S} \in \Omega_s$) *is given by*

$$S = C \cdot P_s \cdot R_s = \prod_{i=C}^{R_s} X_i \tag{7.66}$$

$$= s^3, \tag{7.67}$$

where s *denotes some uniform probability on the closed interval* $[0, 1]$ *taken across causal events* $\mathbb{C}$, $\mathbb{P}_s$, *and* $\mathbb{R}_s$.

Theorem 7.20 (Probability of Potential for State Formation) *The probability of potential for state formation* $\mathbb{P}_s \in \mathscr{P}3(\Omega_s)$ *as a function of first-order causal events* $\mathbb{X}_i$ (*Assumption* 7.5) *is given by*

$$P_s = X_{nonkin} \cdot X_{strategic} \cdot X_{commons} \cdots X_{stress} = \prod_{i=nonkin}^{stress} X_i \tag{7.68}$$

$$= x^{\Theta}, \tag{7.69}$$

where x *denotes some uniform probability taken across* Θ *causal events*, *which are fifteen assuming* $\mathbb{X}_{nonkin}$ *to* $\mathbb{X}_{stress}$ (*causal necessary conditions* 1–15 *for state formation*).

Theorem 7.21 (Probability of Realization) *The probability of realizing a state formation potential* $\mathbb{R}_s \in \mathscr{P}3(\Omega_s)$ *as a function of first-order causal conditions for opportunity* $\mathbb{O}$, *willingness* $\mathbb{W}$, *and implementation* $\mathbb{I}$ *is given by*

$$R_s = O \cdot W \cdot I \tag{7.70}$$

$$= r_s^3, \tag{7.71}$$

where r_s *denotes some uniform probability taken across* $\mathbb{O}$, *willingness* $\mathbb{W}$, *and implementation* $\mathbb{W}$ *events*.

The following second-order principles extend previous principles of state formation.

Theorem 7.22 (Second-Order Probability of State Formation) *The second-order probability of initial state formation is given by*

$$\mathbb{S} = \Psi(\mathbb{C}; \mathbb{X}_{nonkin}, \mathbb{X}_{strategic}, \mathbb{X}_{commons}, \ldots, \mathbb{X}_{stress}; \mathbb{O}, \mathbb{W}, \mathbb{I}), \tag{7.72}$$

$$\Leftarrow \big\langle \mathbb{C} \wedge (\mathbb{X}_{nonkin} \wedge \mathbb{X}_{strategic} \wedge \mathbb{X}_{commons} \wedge \cdots \wedge \mathbb{X}_{stress}) \wedge \mathbb{O} \wedge \mathbb{W} \wedge \mathbb{I} \big\rangle, \tag{7.73}$$

and

$$S = C\left(\prod_{i=nonkin}^{stress} X_i\right) O \cdot W \cdot I \tag{7.74}$$
$$= y^{\Gamma}, \tag{7.75}$$

where y is some uniform probability taken across the set of Γ second-order causal events for $\mathbb{C}$ and $\Gamma > \Theta$.

Note that, in the case of state formation, $\Gamma = 19$ in Eq. (7.75), w.r.t. second-order conditions. In fact, $\Gamma \gg 19$, due to more conjunctions toward the operational level. For example, in this case implementation $\mathbb{I}$ requires further development of potential capacities into state-level forms and functionalities, such as creating another layer of public administration (supporting provincial government, villages, and a central capital) in the form of bureaucratic institutions, appointing and managing public officials (political, judicial, military), building elite palaces, and other necessary and difficult collective action strategies that state leaders, elites, and commoners must accomplish—all produced by highly contingent processes and exogenous shocks (e.g., environmental conditions over a greater territory). Theorem 7.22 explains why primary state formation was such a rare occurrence in world history (early Holocene), even rarer than formation of primary chiefdoms. Since $\Gamma \gg \Theta + 4$ (by Eq. (7.73)) and $\Theta = 15$, it follows that $S(y; \Gamma) \ll y^{19}$, which yields a vanishingly small (but > 0) probability S of state formation for arbitrary values of y. If Assumption 7.6 is incomplete (i.e., if $\Theta > 15$), then primary state formation, as well as its potential, are even rarer events.

7.5 General Theories of Social Complexity

Thus far we have seen theories of social complexity focused on chiefdoms and states, which are significant but particular instances of a much broader class of systems. In this section we expand the theoretical scope to explain emergence and development of social complexity in a more general way. These broader theories are universal in the sense of being applicable to explaining origin, development, and decay of social complexity in all organizational forms.

7.5.1 Theory of Collective Action

The Theory of Collective Action was first formulated by economist Mancur Olson in his 1965 classic monograph, *The Logic of Collective Action*. It has since undergone milestone developments, including:

- Ecologist Garrett Hardin's 1968 game-theoretic formulation of "The Tragedy of the Commons" (collective action as an N-person Prisoners' Dilemma game)

- Economist Albert O. Hirschman's 1970 classic trifurcation, *Exit, Voice, and Loyalty*
- Nobel laureate Elinor Ostrom's discovery of the role of local traditional governance for sustainable management of common pool resources (and public goods and services in general)
- Political scientist Mark I. Lichbach's 1996 generative mechanisms for collective action
- Economist Todd Sandler's 1992 comprehensive synthesis of Collective Action Theory

Paul Samuelson's seminal 1954 paper on "The Pure Theory of Public Expenditure" was a key scientific precursor that established the Theory of Public Goods. Hirschman's trifurcation of *Exit, Voice, and Loyalty* anticipated the Theory of Circumscription examined in the previous section: a circumscribed chiefdom population cannot escape (no exit), so it can only offer resistance (voice) or submit (pledge loyalty). Collective Action Theory ranks among the most important areas of theory and research in social science, integrating psychological, political, economic, cultural, and social dynamics.

Collective action theory seeks to explain why and how humans solve collective action problems, which is a core aspect of social complexity.

Definition 7.11 (Collective Action Problem) A condition where members of a group or society recognize a need to act in a coordinated way in order to overcome a situation, but collective action is hampered because no one perceives an individual incentive to cooperate.

Even if someone would want to solve a collective action problem on his own, it would be impossible for a single individual to produce what is needed for the group, hence the need for coordination of behavior with others. Producing a **public good** or **public service** for a given group or society presents a collective action problem. Classic examples are public sanitation, clean air and water, national defense, neighborhood safety, emergency health services, technical standards and measures, and systems of transportation and communication.

Why humans solve collective action problems is fairly easy to explain: because they recognize a need or desire. Safety from hazards, as well as improvements in quality of life, are universally recognized as desirable outcomes. No one wants to be worse off just for the sake of it.

How humans solve collective action problems, in specific, causal detail, is *not* so straightforward. Significant theoretical progress toward answering this question lies in the mechanisms for collective action problem-solving.

Definition 7.12 (Collective Action Coordination Mechanisms; Lichbach 1996) There are four mechanisms for generating coordinated behavior aimed at solving a collective action problem:

1. *Market:* Providing personal incentives to individual or group participants in collective action. Paramount chiefs provide payoffs to their confederate local chiefs, consistent with Riker's theory, and side-payments in ruling coalitions.

2. *Community:* Invoking norms of solidarity among community members. Deontic obligation based on shared values provides a powerful, intangible incentive that often trumps rational utilitarian choice.
3. *Contract:* Invoking agreements that obligate members to undertake collective action. Contracts can range from enforceable legal documents to private agreements.
4. *Hierarchy:* Exercising authority over community or group members. Besides authority in a narrow sense ("do X"), deterrence ("do not do X or else Y"), and compellence ("do X or else Y") are related forms of exercising power.

Each mechanism has significant implications for explaining social complexity. The market-based mechanism—or market solution, for short—requires significant capacity for providing rewards to participants in collective action. Community solutions require cognitive references (such as in the community shrine/temple worship discussed earlier (Fig. 7.1)) as well as social communication. Contract solutions require enforcement to have credibility. Hierarchy solutions require social capital and capabilities, both, in turn, requiring planning, acquisition, and maintenance to be effective.

The level of difficulty of a collective action problem can be measured by the number of mechanisms required for solution, which can be used to classify collective action problems into four classes:

Class I The simplest collective action problems are amenable to solution via a single mechanism. For example, tax compliance is generally ensured through state authority. Similarly, a social or humanitarian emergency can sometimes be overcome via a community solution.

Class II Two mechanisms are required for solving more challenging collective action problems. National defense is assured through community and market mechanisms.

Class III More difficult collective action problems require use of three mechanisms. Adding a third solution can add resilience, such as when compulsory military service is added through state authority.

Class IV All four mechanisms are required for the most difficult collective action problems. Examples include: adapting to climate change on spatial scales from local to global; carrying out certifiably valid elections in an emerging democracy; solving or mitigating major issues in peace and security, whether domestic, transnational, or international; responding to humanitarian assistance and disaster relief challenges; or managing large financial crises by engaging producers, consumers, lenders, and financial government institutions. The most difficult Class IV problems are called *wicked problems* in policy analysis and management science.

Simply choosing a mechanism and implementing it does not guarantee success in solving a collective action problem. The preceding scale suggests the following result.

Theorem 7.23 (Collective Action Via Several-Among-4 Mechanisms) *The probability of collective action* $\mathbb{C}$ *via a necessary number* ν *of mechanisms from among the total of* 4 *possible or available*, *with a* ν*-out-of-*4 *event function*, *is given by the binomial equation*

$$\Pr(\mathbb{C}) = \sum_{i=\nu}^{4} \binom{4}{\nu} M^i (1 - M)^{4-i}, \tag{7.76}$$

where M *is the probability of individual mechanisms solving the collective action problem and* $i = 1, 2, 3, 4$ *denotes each mechanism.*

Note that ν is the class.

Leadership plays a critical role in collective action, because it can leverage any and all of the above mechanisms for solving collective action problems. Depending on circumstances, leaders can provide incentives (Market), invoke norms (Community), remind others of existing agreements-in-force (Contract), or order them to coordinate behavior (Authority). Leaders known for their ability to enable collective action also develop *reputation*, which facilitates future collective action, as examined more closely through Canonical Theory.

Leadership *can* be a *sufficient* condition for collective action, but it is *not* always a *necessary* condition. This is because a collective action problem might be solved in a leaderless mode, spontaneously. For example, members of a community may be so norm-compliant that they coordinate behavior without requiring leadership. The iconic example is when neighbors help each other in a disaster. What does *not* occur is collective action *without* solution mechanisms; one or more is *always* operant (Lichbach's Law).

Collective action is a ubiquitous, significant, and uncertain phenomenon for understanding social complexity. Besides its intrinsic scientific interest, it also provides foundations for the general theories that follow in the next two sections.

7.5.2 Simon's Theory of Adaptation Via Artifacts

Simon's Theory of Social Complexity—his "Big CSS Idea" based primarily on *The Sciences of the Artificial* (1969, 1981, 1996) and related work—has been introduced and used from a conceptual perspective since the first chapter (Sect. 1.5.3). *Social complexity is the result of human adaptations to complex environments via artificial systems, not because we humans are intrinsically complex; we are not, it's the environment that is complex.* This theory explains social complexity in human civilization since ca. 10,000 years ago. In social science Simon's theory is an idea as big as Copernican Theory, the Big Bang, Relativity Theory, or Darwin's Theory of Evolution in natural science. The theory can be verified, tested, validated, analyzed, and extended to numerous domains across social science (anthropology, economics, political science, sociology, psychology) and allied disciplines (geography, communications, linguistics, management, history).

Simon never explicated his Big Idea in a formal sense, in spite of much other work in mathematical social science. To do so it is necessary to draw on his own concepts and as little else as possible. The key concepts of Simon's theory also reflect his main theoretical assumptions:

Environmental complexity Humans, both individual and collectively in groups or whole societies, are always situated in environments that are often challenging. Numerous natural environments are hazardous or even lethal to humans, even when they appear beautiful to human sensory experience. Environmental complexity, especially in natural systems, exists independently of humans and across the Universe. Climate change is today an iconic example of increasingly challenging environmental complexity. So is the broader policy environment of domestic and global public issues.

Goal-seeking behavior Humans seek goals; they don't just act. Goals, beliefs, desires, and intentions are related entities that are also used in implementing cognitive models of human actors—a framework known as BDI (beliefs-desires-intentions).

Bounded rationality Unlike earlier economic theories of human decision-making, we now know that humans decide using bounded, not perfect, rationality. This means, *inter alia*:

- Humans use imperfect information when making decisions. Noise, imprecision, ambiguity (Zadeh's fuzziness), and uncertainty are common. Bayesian updating is a valuable aid to human decision-making (not just for robots).
- Limited cognitive capacity, faulty information-processing, low band-width, small and imperfect memory, and multiple types of biases are characteristics of human decision-making.
- Satisficing is the principal heuristic used in human decision-making. Optimizing is intractable.
- Computing machines can help improve human bounded rationality by mitigating its limitations, but they cannot support perfect rationality due to intrinsic operating characteristics of human reasoning.

Adaptation Humans adapt to their environments by using whatever bounded rationality they have as they seek goals. Successful adaptation means that a chosen strategy works. Adaptation is therefore conditional upon the environment, goal, and strategy of the circumstance. Successful adaptation requires both implementation and maintenance.

Artifacts Humans build artifacts or artificial systems as interfaces to achieve satisfactory adaptation. Artificial systems are disjoint albeit connected with natural systems. Couplings occur through sensors and effectors. In turn, artifacts can be physical (tangible, built, engineered systems, up to the scale of the largest infrastructure systems) or social (beliefs, norms, institutions, procedures).

Near-decomposability The architecture of human social complexity relies on nearly decomposable structures. Such a design is based on modularity and hierarchy, with a formal network structure similar to a tree or star. The span of a nearly decomposable structure is the number of subsystems or modules into which the system is partitioned.

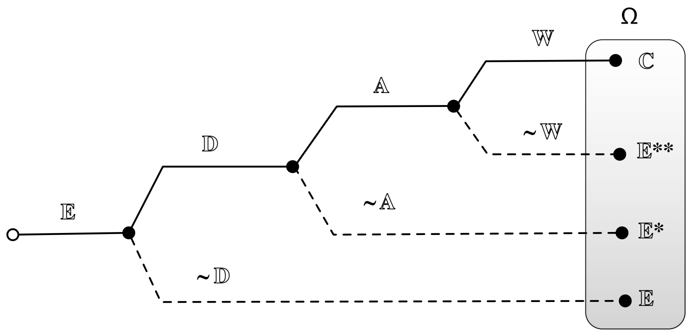


Fig. 7.4 Forward sequential causal logic tree for Simon's theory of adaptation and emergence of social complexity

Emergence Under some circumstances order can emerge through a multitude of local individual decisions, without any central planning.

Based on these concepts and assumptions from Simon's theory, the process of adaptation and complexity generation can be modeled by a sequential tree in forward causal logic, as in Fig. 7.4. At some initial time τ_0 a society is situated in a given environment (event $\mathbb{E}$). Given that the environment is challenging or difficult, at some subsequent time τ_1 humans may or may not decide to adapt (event $\mathbb{D}$), based on bounded rationality. If they do not decide ($\neg\mathbb{D}$) then they continue to endure the same environmental consequences as before at τ_0, whatever those may be (outcome $\mathbb{E}$). If they do decide to adapt, then at some time τ_2 they may or may not actually carry out their decision and implement an adaptive response (event $\mathbb{A}$) by means of some artificial system, which may be social or physical. If they fail to deploy the artifact (event $\neg\mathbb{A}$), then they still endure environmental consequences, only this time more time has passed (outcome $\mathbb{E}^*$). Arguably, $\mathbb{E} \approx \mathbb{E}^*$.

If they do respond via some artifact, then at some time τ_3 the response may or may not work. If it works (event $\mathbb{W}$) then the outcome is success and greater complexity, because now the artificial system has to be maintained (outcome $\mathbb{C}$). If the response fails (event $\neg\mathbb{W}$), then the outcome still entails enduring environmental consequences, this time after experiencing failure (outcome $\mathbb{E}^{**}$). Arguably, now $S(\mathbb{E}^{**}) \gg S(\mathbb{E})$, where $S(\mathbb{X})$ denotes stress or disutility associated with event $\mathbb{X}$.

The model in Fig. 7.4 provides a first-order representation of Simon's theory. The main result is that each outcome in the Ω-space is produced by conjunction. In particular, the emergence of social complexity $\mathbb{C}$ requires minimally four sequentially necessary conditions, implying significant hypo-probability; otherwise it fails to occur. The other outcomes (failures $\mathbb{E}$, $\mathbb{E}^*$, and $\mathbb{E}^{**}$) are relatively less hypo-probable, hence more probable.

A second-order model would also include conditional backward logic for causal occurrence of each event in the first-order model. Accordingly, the environment operates under some set of conditions that may or may not persist, causing it to become more or less challenging, depending on a structure function $\Psi_\wedge(\mathbb{E})$. Similarly, the decision to respond requires its own set of conditions (e.g., bounded rationality requirements), which is specified by a conjunctive structure function $\Psi_\wedge(\mathbb{D})$. Implementing the response via an artificial system is another highly conjunctive event (design, procurement of resources and components, site preparation, construction, initial operation) with its own structure function $\Psi_\wedge(\mathbb{A})$. Finally, whether or not the adaptive response works depends on a conjunctive structure function $\Psi_\wedge(\mathbb{W})$. Therefore, a second-order model would also be strictly conjunctive and, from this perspective, exponentially more hypo-probable.

These concepts and assumptions yield the following principles of Simon's Theory of Social Complexity.

Theorem 7.24 (Simon's Complexity-Simplicity Hypothesis) *"Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves." (Herbert A. Simon, The Sciences of the Artificial* 1996, *p.* 53)

Theorem 7.25 (Artifactual Complexity) *Every successful artificial system has complexity proportional to its associated environmental complexity, with some added complexity as a margin of safety. Symbolically*: $C_A \propto C_E + \delta$.

The following principle follows from application of the general conjunctive principle (Theorem 7.6).

Theorem 7.26 (First-Order Probability Principle for Social Complexity by Adaptation) *The probability of social complexity* $\mathbb{C}$ *by adaptation to a challenging environment is given by the product of probabilities of its four necessary events. Formally,*

$$\Pr(\mathbb{C}) = \Pr\big[\mathbb{E} \wedge (\mathbb{D} \mid \mathbb{E}) \wedge (\mathbb{A} \mid \mathbb{D}) \wedge (\mathbb{W} \mid \mathbb{A})\big] \tag{7.77}$$

$$= E \cdot D \cdot A \cdot W = P^4, \tag{7.78}$$

where P is the probability of these events.

The next principle follows from the structure of second-order events in Simon's theory, as described earlier.

Theorem 7.27 (Second-Order Probability Principle for Social Complexity) *The second-order probability of social complexity* $\mathbb{C}$ *in Simon's process (Fig.* 7.4*) is given by the equation*

$$\Pr(\mathbb{C}) = \Pr\Big(\bigwedge \mathbb{E}_i\Big) \cdot \Pr\Big(\bigwedge \mathbb{D}_j\Big) \cdot \Pr\Big(\bigwedge \mathbb{A}_k\Big) \cdot \Pr\Big(\bigwedge \mathbb{W}_l\Big) \tag{7.79}$$

$$= \prod_{i=1}^{m} \Pr(\mathbb{E}_i) \cdot \prod_{j=1}^{n} \Pr(\mathbb{D}_i) \cdot \prod_{k=1}^{r} \Pr(\mathbb{A}_i) \cdot \prod_{l=1}^{s} \Pr(\mathbb{W}_i) \tag{7.80}$$

$$= E^m \cdot D^n \cdot A^r \cdot W^s = P^{m+n+r+s}, \tag{7.81}$$

where P is a probability value taken across all second-order events.

If each of the four main events requires a minimum of two second-order causal events (i.e., $m = n = r = s = 2$), then $\Pr(\mathbb{C}) = P^8$, which makes emergence of social complexity quite hypoprobable and, consequentially, even more rare than w.r.t. first-order events.

These results, particularly the last two, imply that significant structures of *redundancies* must exist at third and higher causal orders; otherwise the probability of successful adaptations would be vanishingly small. Simon did discuss redundancy in *The Sciences of the Artificial*, but unfortunately not in the same depth as other social theorists (e.g., M. Landau and J. Bendor) who were not directly concerned with investigating social complexity.

Other results will no doubt follow from future analyses of Simon's theory. The results presented here facilitate computational analysis by highlighting agents (actors and environments), behavioral rules (adaptation and other patterns), and dynamics (interactions among main entities). Additional insights based on Simon's rich theory await implementation through variable-based and object-based social simulations. The theory can also be used in combination with others, to develop new theories, as examined in the next section.

7.5.3 Canonical Theory as a Unified Framework

The Canonical Theory of Social Complexity is based on elements from behavioral and collective action theory, Simon's theory, and related concepts on causes, origins, and evolution of social systems. It represents a reinterpretation and synthesis of earlier ideas, guided by the application of the General Theory of Political Uncertainty in the context of explaining social complexity.

The first distinction drawn by the Canonical Theory of Social Complexity concerns **dual time-scales** of social complexity, as stated by the following formal assumption.

Assumption 7.7 (Dual Time-Scales of Social Complexity) *Time has dual scales in social complexity processes: fast and slow modes. The* ***slow process*** *is marked by relatively low-frequency, long-term emergence and development of social complexity as observed by succession of polities and macro historical dynamics (e.g., rise and fall of polities), approximately on an annual to decadal or longer scale. The* ***fast process*** *is marked by relatively high-frequency, short-term events associated with problem-solving and adaptation and micro historical dynamics, approximately on an hourly or daily to weekly scale.*

Another way to understand these dual time-scales of social complexity is to view them as metrics for counting coarse- and fine-grained events that occur in history, using events data analysis terminology.

The precise theoretical relationship between dual time processes of social complexity is critical and given by the following premise.

Assumption 7.8 (Inter-Temporal Synchronization of Social Complexity) *The slow process of change in social complexity on a long-range, macro scale is generated by accrual of complexity-related consequences (externalities) of outcomes generated by fast process iterations.*

More specifically, as illustrated in Fig. 7.5, the fast process is a *sequential branching process* (event tree) spanned by **states of Nature** and **human acts** generated by **lotteries** (denoted by triangle-nodes) and **decisions** (square-nodes), respectively. The **outcome space** Ω of a fast process consists of all resulting compound events ($\mathbb{O}_j \in \Omega$) generated by the process. In this case, $n(\Omega) = 5$, so

$$\Omega = \{\mathbb{A}, \mathbb{Z}, \mathbb{X}, \mathbb{X}^*, \mathbb{E}^*\}, \tag{7.82}$$

as shown in Fig. 7.5. Specifically, social complexity changes—by increasing, decreasing, or remaining constant—as a direct result of outcomes realized in the fast process, as we will now examine in closer detail.

A fast process with potential (not certainty) for social complexity begins at an initial time τ_0, when a given society or social group is in some **ground state** x_0 (event $\mathbb{K}$ in the left of the graph). What happens next explains whether social complexity increases, decreases, or remains unchanged.

1. At some later time τ_1 a situational change may or may not occur (events $\mathbb{C}$ and $\neg\mathbb{C}$, respectively). If situational change does not occur, then the society persists without much change in social complexity (outcome $\mathbb{E}^*$ is generated, which is roughly comparable to $\mathbb{K}$ except for the passage of some time interval $\Delta\tau = \tau_1 - \tau_0$).
2. The interesting process begins when a situational change *does* occur—one that has significant effect on a society, whether immediate or potential. Such an occurrence may be a threat or an opportunity, corresponding to negative or positive consequences. Regardless, if $\mathbb{C}$ occurs, then societal members may or may not recognize a need for action at time τ_2 (events $\mathbb{N}$ and $\neg\mathbb{N}$, respectively). Since the situational change is societal, not private to an individual, then, by definition, the action required is collective, requiring coordination. If need for action is not recognized when it is objectively necessary, then the outcome will be detrimental consequences (outcome $\mathbb{X}^*$).
3. If $\mathbb{N}$ occurs, then societal members may or may not undertake action at time τ_3 (events $\mathbb{U}$ and $\neg\mathbb{U}$, respectively). If action need is not undertaken, then this outcome will entail detrimental consequences (outcome $\mathbb{X}$), even if the need was recognized.
4. If $\mathbb{U}$ occurs, then collective action may or may not work at time τ_4 (events $\mathbb{S}$ and $\neg\mathbb{S}$, respectively). If collective action fails, then this outcome will carry detrimental consequences (outcome $\mathbb{Z}$), even if action was undertaken.
5. If $\mathbb{S}$ occurs, then the outcome is successful adaptation at time $\tau_4 + \delta$ (outcome $\mathbb{A} \in \Omega$, which is a compound event).

The theory is called "canonical" because the same fast process cycles through unlimited iterations each time with only finite and identifiable variations.

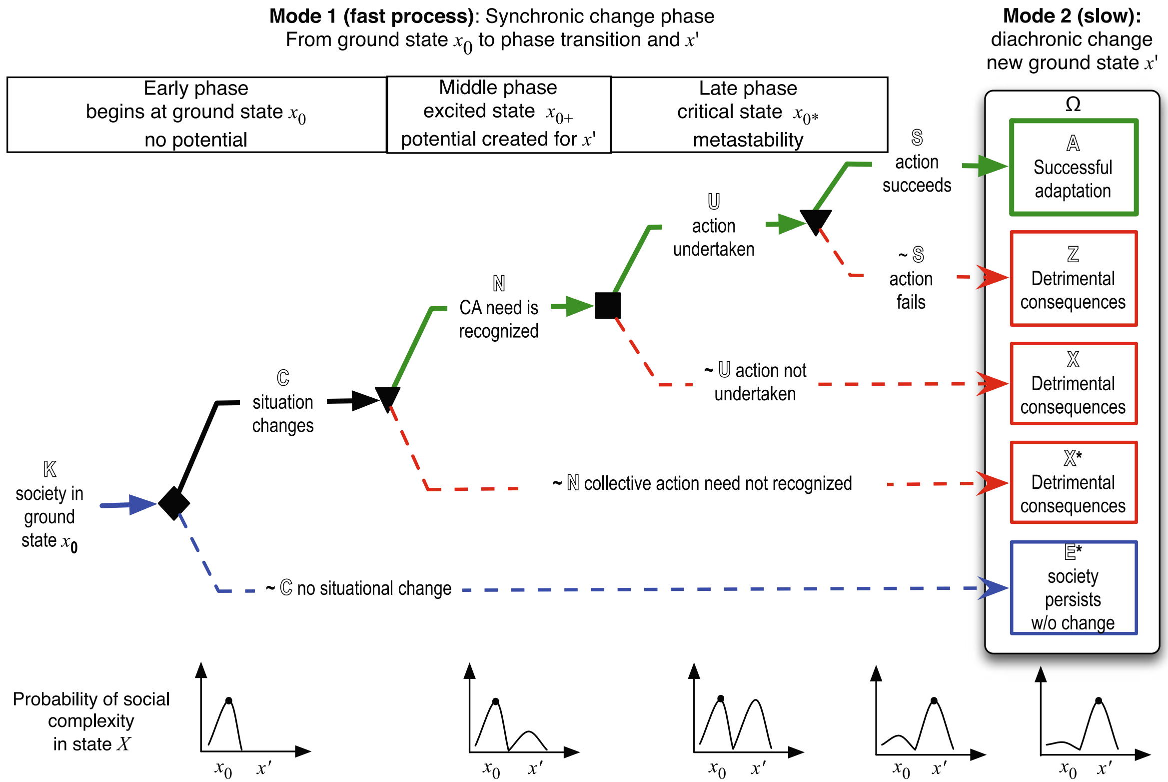


Fig. 7.5 Forward sequential causal logic tree for the Canonical Theory of emergence and development of social complexity. *The main upper part* of the graph illustrates the fast process. Change in the probability of social complexity is shown across the bottom. *Node notation*: decisions are denoted by *triangle-nodes*, lotteries by *square-nodes*, and hybrids by *diamond-nodes*

How do fast process iterations generate change in social complexity in the slow process? Formally, each fast process resulting in outcome $\mathbb{O}_i(\tau) \in \Omega$ generates a set of associated consequences $\kappa_\tau(\mathbb{O})$ relevant to social complexity (*externalities*, as they are called in economics). Obviously, not all outcomes generate the same complexity-related consequences, as shown by the Ω-space in Fig. 7.5. In turn, consequences of the outcome from time τ generate change in social complexity $C(\tau + 1)$.

Thus, on the long-range scale of the slow process, social complexity at time $\tau = m$, denoted by $C(m)$, is generated by the integration (summation, in discrete time) over all iterative fast process cycles

$$C(m) = \sum_{\tau=0}^{m} \kappa_\tau(\mathbb{O}) - L(\tau), \tag{7.83}$$

where $L(\tau)$ is a loss function representing some inevitable decay in complexity. Examples of the latter include faulty information-processing, imperfect or deficient learning, loss of memory, and similar individual or collective occurrences that, over the long-term, act to the detriment of social complexity. Bounded rationality has long-range societal effects on multiple spatial and temporal scales, not just local effects on individual decisions.

For example, consider successful adaptation (outcome $\mathbb{A}$ in Fig. 7.5), the most successful outcome of a fast process. This can have the following consequences $\kappa_\tau(\mathbb{A})$ in terms of capacity-building for further social complexity at time $\tau + 1$:

1. Members of the group enjoy success as a result of overcoming adversity, increasing their confidence in problem-solving.
2. Neighbors (local or distant) may take note of the group's success.
3. New values, beliefs, norms, procedures, or institutions emerge in the process of realizing each intermediary event and cumulatively with respect to the overall outcome (compound event).
4. New specific, practical experience in problem-solving is acquired, including ability in:
   - Recognizing need for collective action
   - Planning one or more actionable solutions to the problem
   - Implementing the plan by coordinating its execution
5. Leaders and followers experience each others' performance, learning whom to trust, who has which skills, who behaved well or dishonestly, and other valuations of actor attributes and behaviors.
6. Members' reputations and their perceptions are updated.

These cognitive and relational consequences to participants amount to increased social complexity in terms of larger and more informative belief systems, increased memory, development of social relations, and (sometimes) creation of new norms or institutions. Accordingly, $\kappa(\mathbb{A}) > 0$. The next time a situational change occurs at some τ_0' (i.e., at start of the next fast process iteration), the group or society will have greater complexity with new capacities for problem-solving.

Other fast process outcomes produce different sets of consequences. For example, when collective action need is not recognized ($\neg\mathbb{N}$) and situational changes remain completely unmanaged, or when action is not undertaken ($\neg\mathbb{U}$) or when it fails ($\neg\mathbb{S}$), all such outcomes ($\mathbb{X}^*$, $\mathbb{X}$, or $\mathbb{Z}$, respectively) have detrimental consequences ranging from mild to catastrophic, *resulting in short-term degradation of social complexity* ($\kappa < 0$). Failure, if not catastrophic, can define the new societal situation, generating a new fast process on the slow process time-scale, at $\tau + 1$, and iterating through the same canonical cycle. Successful adaptation often comes after an initial failure.

Success in a fast process may increase the probability of future success, provided societal members learn lessons from experience, future situational changes fall within the range of experience, and experience is properly used. Experience in problem-solving decays as a function of time, so the frequency of situational changes matters: high frequency can overwhelm society's capacity or ability to adapt successfully; low frequency can induce memory loss and decrease the probability of success.

Canonical Theory explains a significant range of complex phenomena of interest in basic and applied social science research. A more specific example pertains to explaining and understanding *disasters*, especially those affecting coupled, complex socio-techno-natural systems. Figure 7.6 illustrates this by applying Canonical Theory to the classical **hazards-disasters conundrum**, whereby disasters are not considered "natural," but instead are caused by failures to adapt or prepare for hazards. Hazards are natural or technological events; disasters are social consequences, which at least to some degree can be mitigated, if not entirely eliminated. The fast process in this case initiates with societal exposure to some set of hazards. Given such a ground state, preparedness may or may not occur, depending on awareness ($\mathbb{N}_P$), decisions ($\mathbb{D}_P$), and preparations taking place ($\mathbb{A}_P$). If preparedness fails, hazards may or may not occur, and other contingencies concerning incident response will determine a range of detrimental outcomes.

If preparedness takes place (event $\mathbb{P}$ in Fig. 7.6), a hazard may or may not occur, preparations may or may not work, incident responses may or may not be undertaken, and they may succeed or fail. These first-order events of the more complex fast process generate a larger but identifiable outcome space. Each outcome is a compound event, as before for the purely theoretical process, so each can be modeled by a probability equation given by the Sequential Conjunctive Principle.

As a society cycles through fast processes, the outcome of each iteration yields consequences directly determined by the path taken. This explains why social complexity is **path-dependent**: different paths generate different individual and collective consequences. Hazards-disasters fast processes are notorious for shaping the landscape of societies from a world history perspective.

Moreover, fast processes are multiple, concurrent, and asynchronous social processes, having parallel lanes and interdependent activity lanes on a Gantt chart or in Ganttian space. In the asymptotic limit the slow process is generated by the integration of fast processes over time, which explains how historical continuity emerges from statistical ensembles of discrete event-based fast processes.

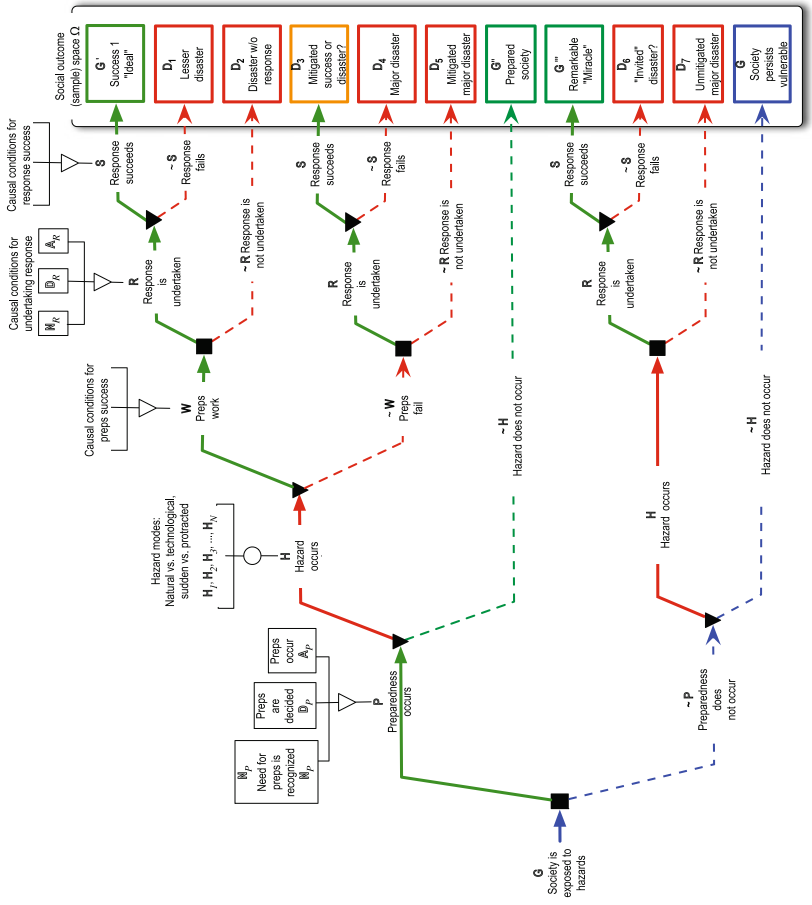


Fig. 7.6 Forward sequential causal logic tree for explaining risky hazards and societal disasters according to Canonical Theory

Some significant advantages of Canonical Theory can be summarized as follows:

1. The theory explains significant aspects of social complexity, including new phenomena and new links among previously unrelated ideas, going beyond earlier theories.
2. Canonical Theory includes elements of several valuable, earlier theories, such as Carneiro's, Marcus's, Dahl's, and Lichbach's, among others, as *special cases* within a broader and more general explanatory framework.
3. From a computational perspective, the explanatory mechanism of the theory is iterative and the fast process generates complexity, in the sense of von Neumann.
4. The theory is testable through a variety of approaches, including case studies, comparative analysis, and statistical assessment.
5. The theory is applicable from a long-range spatio-temporal perspective, in the sense that it explains social complexity in past, current, and future history.
6. Both the quest for survival and improvement in quality of life can serve as initiating events of a fast process, as reactive and proactive responses, respectively.
7. The dual time-scales allow the theory to provide integrated explanations of micro phenomena as well as macro trends in social complexity.
8. The fast process, in particular, offers a systematic template for conducting comparative research using cases across space and time as analyzed through a common framework.
9. The theory is applicable to social, socio-technical, socio-natural, and socio-techno-natural systems, including explanations of how and why natural, technological, and anthropogenic hazards cause disasters.
10. The event-based or discrete modeling approach allows the theory to be implemented in computational models, such as a multi-agent system, and makes full use of probability theory and related calculus for deriving analytical results.
11. By distinguishing between actors and other entities, as well as between decisions and lotteries, Canonical Theory leverages significant ideas and results from decision theory and game theory.
12. The theory can be improved by others, as further formal analyses and computational implementations uncover previously unknown formative and developmental processes.
13. Basic science questions as well as policy-oriented analyses can be addressed through Canonical Theory.

Social simulations provide computational implementations of social complexity theories, as examined in the next chapters.

Recommended Readings

R. Carneiro, A theory of the origin of the state. Science **169**, 733–738 (1970)

C. Cioffi-Revilla, *Politics and Uncertainty: Theory, Models and Applications* (Cambridge University Press, Cambridge, 1998)

C. Cioffi-Revilla, A canonical theory of origins and development of social complexity. J. Math. Sociol. **29**, 133–153 (2005)
R.A. Dahl, *Polyarchy: Participation and Opposition* (Yale University Press, New Haven, 1972)
T. Earle, *How Chiefs Come to Power: The Political Economy in Prehistory* (Stanford University Press, Stanford, 1997)
S.N. Eisenstadt, *The Political Systems of Empires* (Free Press, New York, 1963)
K.V. Flannery, J. Marcus, *The Creation of Inequality: How Our Prehistoric Ancestors Set the Stage for Monarchy, Slavery, and Empire* (Harvard University Press, Cambridge, 2012)
F. Fukuyama, *The Origins of Political Order: From Prehuman Times to the French Revolution* (Farrar, Straus and Giroux, New York, 2011)
J. Marcus, Dynamic cycles of Mesoamerican states. Res. Explor. **8**(4), 392–411 (1992)
J. Marcus, The peaks and valleys of ancient states, in *Archaic States*, ed. by G.M. Feinman, J. Marcus (School of American Research Press, Santa Fe, 1998), pp. 59–94
T. Parsons, *The Social System* (Routledge, London, 1991)
N. Rashevsky, *Looking at History Through Mathematics* (MIT Press, Cambridge, 1968)
C.L. Redman, *Human Impact on Ancient Environments* (University of Arizona Press, Tucson, 1999)
J.D. Rogers, C. Cioffi-Revilla, Expanding empires and a theory of change, in *Current Archaeological Research in Mongolia*, ed. by J. Bemmann, H. Parzinger, E. Pohl, D. Tseveendorzh (Bonn University Press, Bonn, 2009), pp. 445–459
H.A. Simon, The architecture of complexity. Proc. Am. Philos. Soc. **106**, 467–482 (1965)
H.A. Simon, *The Sciences of the Artificial*, 1st edn. (MIT Press, Cambridge, 1969)
H.A. Simon, *The Sciences of the Artificial*, 3rd edn. (MIT Press, Cambridge, 1996)
C.S. Spencer, A mathematical model of primary state formation. Cult. Dyn. **10**(1), 5–20 (1998)
H.T. Wright, Prestate political formations, in *Chiefdoms and Early Settlements in the Near East*, ed. by G. Stein, M.S. Rothman (Prehistory Press, Madison, 1994), pp. 67–84

Simulations I: Methodology

8

8.1 Introduction and Motivation

This chapter provides an introduction to social simulation as a major area of CSS research—independent, or almost independent, of the specific type of implementation. The core questions addressed in this chapter concern computer modeling and simulation in social science. Why use computer simulation as a methodology for scientific investigation of social complexity? The answer is—in brief—because formal theories of social complexity are sometimes more viable via computational modeling than through closed-form solutions. What unique insights on social complexity are gained through social simulation that are not available through other methodological approaches, such as statistical, mathematical, or historiographic? A major one is improved understanding of social complexity as an emergent phenomenon. What are the main limitations of social simulations? Full descriptions of social simulations are not as straightforward as thorough descriptions of other formal and statistical models, which sometimes can have significant consequences for replicating results. Another limitation is the relative shelf life of computer code as compared to mathematical models.

The main motivation for social simulation is based on the first two of these questions. Social simulations are capable of representing social systems and coupled socio-techno-natural systems in ways that other methodological approaches are not. Computer code in a well-chosen programming language or simulation system—such as those discussed in this and the next two chapters—provides a powerful formalism for theorizing, experimenting, and ultimately understanding social complexity.

8.2 History and First Pioneers

The following is a brief history of milestones and pioneers of social simulation research in CSS, with main emphasis on methodological concepts, principles, and practice—especially the founders' generation. Similar sections in the next two chap-

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_8,

ters focus more specifically on models. Some overlap between these summaries is unavoidable, since they are not completely disjunct.

1959 Oliver Benson at the University of Oklahoma pioneers the methodology of computer simulation in political science with his Simple Diplomatic Game Model.

1961–1971 Jay Forrester, founder of the System Dynamics Group at MIT, establishes the methodology of system dynamics theory and research through his classic monographs: *Industrial Dynamics*, *Principles of Systems*, *Urban Dynamics*, and *World Dynamics*.

1962 Psychologist and information science pioneer Harold Borko [1922–2012] publishes the edited volume *Computer Applications in the Behavioral Sciences*, possibly the first of its kind, including Julian Feldman's seminal chapter on "Computer Simulation of Cognitive Processes", Sydney and Beatrice Rome's computer simulation of large organizations, R. Clay Sprowls's "Business Simulation", and Benson's model.

1963 Political scientist Karl W. Deutsch [1912–1992] publishes *The Nerves of Government: Models of Political Communication and Control*, pioneering the information-processing paradigm of CSS, as a precursor to Simon's work. The same year Harold Guetzkow and collaborators publish the influential *Simulation in International Relations: Developments for Research and Teaching*, which soon becomes the new frontier.

1968 The Club of Rome, a major promotor of global carrying capacity modeling and simulation, is founded by Italian industrialist Aurelio Peccei and Scottish scientist Alexander King.

1969 Political scientists Hayward Alker and Ron Brunner publish the first comparative analysis of social simulation models in the journal *International Studies Quarterly*.

1970 Computer scientist James E. Doran publishes one of the earliest papers on the application of simulation methodology to archaeology, "Systems Theory, Computer Simulations and Archaeology", in the first volume of the journal *World Archaeology*.

1970s In Europe, social scientist Urs Luterbacher and collaborators at the Graduate Institute of International Studies in Geneva develop SIMPEST, the first numerical simulation model of political, economic, and strategic interactions based on a dynamical system of integral-differential equations, implemented in MINUIT. This model of the US-USSR-PRC triad correctly predicted the fall of the Soviet Union in late 1980s.

1970s In America, economist and strategist Thomas Schelling establishes foundations for a new methodological chapter in social simulations via cellular automata, and eventually agent-based modeling, through his study of racial segregation. John Casti, who later joined the Santa Fe Institute, coded the first implementation of Schelling's model while the two were at The Rand Corporation.

1972 Springer publishes the first edited volume on CSS in Europe, by Lucien Kern and collaborators, entitled *Simulation internationaler prozesse*, containing Jeffrey Krend's chapter on a replication of Oliver Benson's pioneering model.

1977 CSS pioneer Stuart Bremer [1943–2002] advances the methodology of social simulation with *Simulated Worlds: A Computer Model of National Decision Making*, published by Princeton University Press.

1980s Computer scientist Christopher Langton coins the term "artificial life".

1999 Computational social scientists Nigel Gilbert and Klaus Troiztch publish the first edition of the influential textbook, *Simulation for Social Scientists*.

2013 Computational social scientists Bruce Edmonds and Ruth Meyer edit the 754-page comprehensive handbook, *Simulating Social Complexity* by Springer. The same year both Springer and Wiley inaugurate specific series on Computational Social Science.

8.3 Purpose of Simulation: Investigating Social Complexity Via Virtual Worlds

The core scientific purpose of social simulation modeling and analysis is to investigate social complexity in ways that go beyond—often *way beyond!*—what is possible using other methodologies, such as historical, ethnographic, statistical, or mathematical approaches. This is accomplished by building a computer model of the social system or process under investigation—a virtual world representing relevant aspects of reality—and using that model to perform many kinds of analyses, as detailed in this and the next two chapters.

Reasons for using virtual worlds that simulate social complexity are numerous, including but not limited to the following:

Versatility: Many more complex social systems and processes can be investigated through simulation than through statistical or mathematical modeling. While every statistical or mathematical model can be simulated, the inverse is not true. Not every simulation model can be represented in mathematical form.[1]

High dimensionality: A common feature of social complexity, as we have seen in previous chapters, is having to analyze large numbers of variables, and interactions among them, a property called high-dimensionality. For example, emergence of collective action is a process involving numerous entities and variables, including situational parameters, goals, leadership characteristics, and resources, among numerous others. High-dimensional systems are common across domains of social complexity.

Non-linearities: Dynamic interactions among social entities are often nonlinear, independent of their dimensionality. Simple, low-dimensional systems are sometimes amenable to closed-form solutions, but that is generally not the case

[1] This is obviously not a blank criticism of statistical and mathematical models, which continue to play an essential role in CSS, as already shown in previous chapters.

for complex systems with high-dimensionality and nonlinear dynamics. Human perceptions, interaction as a function of physical distance, and patterns of cooperation and conflict are examples of nonlinear interactions. Social simulations can handle complex nonlinear dynamics, bound only by computational resources (which keep increasing).

Coupled systems: Another distinctive feature of social complexity is coupling among human, natural, and artificial systems, which virtually always implies high-dimensionality and nonlinear interactions. Computer simulation models provide an effective *and* efficient way of representing coupled socio-natural-artificial systems, as we will examine. For example, a computer model can be used to represent coupled dynamics among social institutions, the biophysical world of a society, and critical infrastructure.

Stochasticity: Randomness is ubiquitous and consequential in social systems and processes, as we have already examined. Stochasticity also comes in many forms, as defined by probability distributions. Examining the effects of diverse stochastic dynamics—how they generate patterns of social complexity—is another major reason for using simulations.

Incompleteness: Social science is incomplete, in the sense that not all parts of the social universe are known with the same degree of completeness. Social simulations are also used for testing alternative theories to advance our understanding of real-world social complexity.

Experimentation: The experimental method is a cornerstone of all science, but running experiments on complex social systems is not feasible for numerous reasons, including practical and ethical. Experimentation is rendered feasible through social simulations, including all classical features of this approach: treatments, control groups, and many different experimental designs. For example, computational experiments can be used to explore and test hypotheses concerning aspects of collective action, group dynamics, and governance under various assumptions of governance and public issues.

Policy analysis: Computer simulations of social complexity enable forms of policy analysis that are not available through other methodologies, including analysis of so-called "wicked problems"—the hallmark of hard challenges in policy analysis. For example, economic policies to mitigate inflation can be analyzed by modeling various actions such as wage subsidies or price controls.

These are powerful and compelling reasons! Interestingly, most of them are the same for scientists in other domains who use simulations—including astronomy, biology, and chemistry, among others—"Science in the 21st century is computational", as computer scientist Peter Denning once remarked.

8.4 Basic Simulation Terminology

Social simulation research employs a rich technical vocabulary that includes native CSS terms as well as terminology from computational science, such as object-oriented modeling and programming, UML, and related formal languages. For now

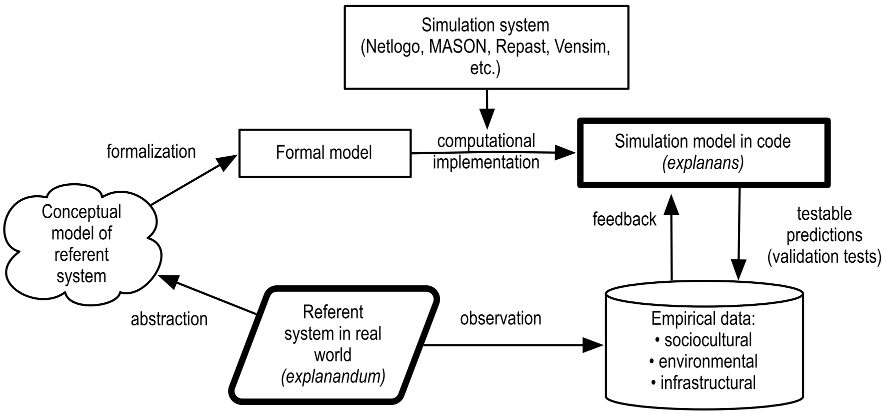


Fig. 8.1 *Basic terminology and general methodology of social simulation*. Social simulation methodology is an iterative process that begins with a referent system (*explanandum*) in the real world. Abstraction, formalization, programming, and appropriate data are used to develop a viable simulation model (*explanans*). This general process is independent of the specific kind of simulation model

we only need to clarify some initial terms; others will be presented as they are needed.

We shall use the following terms as *synonyms:*

- social simulation
- simulation model
- computer model
- machine simulation
- computational model
- simulated system

Hence, by "simulation", for short, we shall always refer to some kind of *computer model* of a social system or process, reserving the term "game" or "gaming" to human simulations solely based on role-playing.

The ontology of social simulation research includes the following basic terms, some of which are shared by other formal approaches, such as mathematical models. Consider Fig. 8.1, starting with the referent system (*explanadum*), in the bottom left and proceeding clockwise. Later we will use these initial building blocks to explain the methodology of modeling complex social systems as a systematic process.

Definition 8.1 (Referent System) A real-world system or process that is an object of investigation (*explanadum*) is called a referent system. Synonyms: target system, focal system, empirical or historical world.

Referent systems in CSS comprise the full *universe* of social entities, systems, and processes: the human mind, cognitive processes, decision-making, individual and group behavior, and societal, international, and global domains, including the World Wide Web. Some of the most complex referent systems in CSS are arguably coupled socio-techno-natural systems, although a referent system of any degree of complexity may focus on a purely human/social system, or pairwise combinations of socio-technical and socio-natural subsystems.

A referent system is *defined* or *specified* by the specific *research questions* being investigated; it is not open-ended or all-inclusive, simply because it is located in the real-world. "Reality" is infinitely detailed and vast, objectively speaking. Scientific research always focuses attention on some selected subset of reality—i.e., a given referent system defined by research questions.

The following definition uses the term "abstract" as a verb to describe a key modeling activity.

Definition 8.2 (Abstraction) The process of selecting a given set of features from a referent system for modeling purposes is called abstraction.[2]

Thus, abstraction produces a simplified *conceptual representation* of the referent system, consisting of elements such as entities, variables/attributes, associations, and other patterns that provide specificity to the referent system being investigated. Sometimes the conceptual model is formalized into an intermediate mathematical model to better understand some properties of interest—as is typical in formal social theory.[3] The conceptual model is actually formalized into a simulation model when it is rendered in code. A simulation model may be written in native code, using one or more programming languages, or using some pre-existing simulation system.

Definition 8.3 (Simulation System) A computational toolkit or code library for building simulation models is called a simulation system.

A simulation system is a highly sophisticated computational artifact for building *other* advanced computational artifacts (specific models), which can be highly complex and inefficient/ineffective to build in native code. Netlogo, DYNAMO, Stella, Vensim, Swarm, MASON, Repast, and their predecessors, among many others, are examples of computational simulation systems. A social simulation *model* is to a simulation *system/toolkit* as a car is to a car factory; the former is made using the latter. You can also build a car on your own (good luck!), rather than buying one made in a factory—which would be the equivalent of writing a social simulation model in native code—but its performance and reliability will probably not come even close to a factory-made car. An important reason for using one of the latest existing simulation systems (Vensim, MASON, Repast, among others) is to reach levels of model performance and reliability that are unattainable by relying exclusively on purely native code. This is, emphatically, *not* an argument against building

[2]Note that the term "abstraction" has a different meaning in the context of computation, where it means hiding information, as discussed in Chap. 2.

[3]The full and powerful family of mathematical structures is available for this, including continuous, discrete, and hybrid formalisms.

simulation models; sometimes they are the best solution to a given set of research questions.

Multi-purpose computational mathematical systems, such as Mathematica and Matlab, are also used as simulation systems, to build and analyze models.

Some common (albeit not universal) facilities of simulation systems include the following:

Frequently used primitives: Code library of common primitives or basic building blocks for building a model. Examples: mathematical functions, distributions, simple agents, landscapes, schedulers, common data fields, constructor methods.

Random number generator: Simulation models require random number generators to represent processes, either substantive or procedural, with various forms of randomness (uniform, Poisson, power-law, among many others).

GUI: A graphic user interface is standard in most simulation systems, especially those intended for beginners and intermediate programmers, such as Netlogo, Repast, and Vensim.

Visualization tools: Used to draw histograms, time-series graphs, network diagrams, maps, and other visual aids for understanding simulation output.

More specialized facilities are usually added by model developers. These might include, for example, autocorrelograms and spectral diagrams, difference maps, heat maps, dynamic networks, Lorenz-curve graphs, and various non-Cartesian coordinate systems (e.g., spherical, cylindrical). All major simulation systems today have active user communities and some hold regular conferences or workshops.

Finally, a simulation model is implemented in code (*explanans*), as highlighted in Fig. 8.1, in the upper right, diagonally opposite the referent system (*explanandum*).

Definition 8.4 (Simulation Model) A model of a referent system that is formalized by code written in a given computer programming language (native or toolkit) is called a simulation model.

In the next chapters we will discuss different types of simulation models and examples of each. To do so in a systematic way, however, it is necessary to develop a viable classification of simulation models, given how many exist.

Figure 8.1 and the preceding definitions provide a first, high-level pass through the general methodology of simulation research in CSS. A more in-depth presentation is necessary, but several other distinctions are needed before delving into methodological details of actual simulation development or model construction.

8.5 Fidelity of Representation and Implications

Social simulations differ by the fidelity with which the computational model attempts to replicate or resemble a given referent system. The following ordinal scale distinguishes social simulations by increasing level of empirical specificity, which approximately follows a pure-applied science continuum:

1. At one end of the basic-applied continuum are highly abstract simulations that bear only sparse qualitative resemblance to a referent system, without attempt to replicate any quantitative features at all. Theoretical analysis as basic science is the main use of these models, not operational policy analysis.
2. At the next level toward "the plane of empirical observation" of the referent system—as philosopher of science Carl G. Hempel would have said—are simulation models that show convincing qualitative fit and some quantitative calibration. These models are still mostly theoretical, but they are capable of providing some applied insights. Since policies should not ignore basic science, findings from this class of social simulations may have valuable implications that policymakers ignore at their own peril. A good example of this is the classical Schelling segregation model (examined in Chap. 10), which is a rather abstract theoretical model that nonetheless sheds significant light on emergent patterns of social segregation and contributes key insights for policymakers.
3. Next are models with extensive qualitative and significant quantitative fit. This class of social simulations is of maximal interest for conducting empirically grounded CSS research. We shall examine several examples of this.
4. Finally, we come to social simulations that "look closest at the plane of observation" (in the sense of Hempel), such that quantitative and qualitative fit between simulation output and empirical data is the closest. High-fidelity simulations are calibrated to a referent system along multiple dimensions, which can be spatial (including numerous and detailed geographic features, down to a given scale of resolution, rendered through GIS and remote sensing data), temporal (defined to small time increments, such as decades, years, seasons, months, weeks, days, hours, minutes, and so on, down to the smallest scale of interest), or organizational (matching detailed network patterns at node, subgraph, and graph levels of analysis), among the most universal. Relatively fewer of these models are found in an academic context, but they are abundant in business and governmental organizations.[4]

This scale is totally unrelated to the merits or value of a simulation model, which is a different matter that has to do with scientific quality.[5] The fidelity scale is merely a heuristic way to locate a simulation model along a realistic-abstract continuum in order to understand its value and limitations.

There are numerous implications that follow from a model's representational fidelity. Perhaps the most obvious is that a simulation at one level cannot be expected to perform well at a different level. Thus, operational, high-fidelity models may have

[4]Part of the reason for this is that operational, high-fidelity models often require sensitive or proprietary information not normally used in academic CSS research.

[5]DARPA—the Defense Advanced Research Projects Agency of the US Department of Defense—uses a scale for classifying projects, ranging from "basic science" (called "6.1 projects", named so after the section in the relevant law) to more applied and operational research, labeled 6.2, 6.3, 6.4, etc., all the way up to fully operational systems deployed in the field for combat or humanitarian missions. The 6.X nomenclature is helpful and commonly used by other agencies.

significant policy value, but have little or no theoretical interest. Conversely, theoretical models can provide deep scientific insights and understanding, but offer little by way of actionable results as far as policy contributions are concerned.

A somewhat less obvious implication of the fidelity scale is that CSS researchers must make an effort to clarify as best as possible the desirable resolution of a model, *given the research questions*.

8.6 Types of Social Simulation: From System Dynamics to Agent-Based Models

Social simulation models constitute several major superclasses, the two largest being **variable-oriented models** and **object-oriented models**, with a third superclass of **hybrid social simulations** at their intersection. In turn, each superclass encompasses several significant classes, which can be characterized as follows. (Each class is examined in the next two chapters.)

Variable-based social simulations use systems of mathematical equations to implement the conceptual model abstracted from the referent system of interest. Historically, these were the earliest forms of simulations in CSS. **System dynamics simulations** and **queuing models** constitute major classes, both based on variables and deterministic or stochastic systems of equations for representing dynamic interactions.[6] The most distinctive feature of a system dynamics model (or SD, for short) is the representation of the state and dynamics of the referent system in terms of *levels* and *rates*, or "stocks and flows", respectively, in the form of a system of difference equations in discrete time. Hence, social systems that are abstracted as networks of states and rates of change are eminently suitable to this kind of simulation model. An SD system may be completely deterministic or partly stochastic.

A queuing model is more appropriate for rendering a referent system that receives some stream of inputs and releases the entities after some processing. The iconic example of this is a commercial bank, where customers arrive and wait in line while those ahead get served and depart the bank when they are finished. These models are stochastic, because waiting time and service time are generally stochastic, not deterministic. Accordingly, *probability distributions* play a major role in this class of social simulations.

These two classes of models are called variable-oriented because the modeling orientation upon which the abstraction is based looks first at the identification of key variables, such as levels of some stock and waiting time in a queue. Neither of these two classes of simulation models makes an effort to render the social entities (actors) explicitly; they are simply implied by state equations.

By contrast, object-oriented simulation models are based on an abstraction strategy that looks first of all at entities in the referent system. **Cellular automata** social simulations (or CA models, for short) consist of cells related to each other by neighboring relations on a landscape, such as in a city grid consisting of blocks, or a

[6]Note the exact terminology: "system dynamics", *not* systems dynamics (both plural) or dynamical systems (which refer to systems of differential equations).

patchwork of farms in the country. CA models look first at entities—the cells and their topology—and then at attributes/variables. Agent-based models are somewhat similar, as detailed in Chap. 10.

8.7 Development Methodology of Social Simulations

All social simulations, whether simple or complex, abstract or empirical, variable-oriented or object-oriented, are developed by systematic steps that begin with some core research motivation and end with a viable model. Although the specifics of each class sometimes matter, in general all social simulations follow a similar developmental methodology.[7] This section provides a second pass (spiral) through the cycle in Fig. 8.1.

8.7.1 Motivation: What Are the Research Questions Addressed by a Given Model?

The first step in social simulation modeling consists of careful formulation of viable research questions. Every social simulation is intended to address one or more research questions defined in terms of the referent system. In fact, a referent system is in large part defined by research questions; there is a synergistic relationship between the two. In an abstract SD model of inter-group rivalry the research questions may concern phase portraits and qualitative dynamical features. The same kind of model calibrated with historical data would be able to address research questions on the timing and magnitude of real-world conflicts. Similarly, research questions in an agent-based model will vary by level of fidelity, ranging from abstract, theoretical questions that may have to do with thresholds, elasticities, gradient fields, and similar theoretical concepts, to empirically referenced questions that might concern specific locations, actors, parameter values, or historical epochs.

Since research questions are a major engine for scientific inquiry, they largely define the level of fidelity and, therefore, also the scope of the referent system to be investigated. That being said, practical considerations may affect decisions on exactly how research questions are formulated.

- The relevant social science may be incomplete, so research questions may require adjustment in order to gain scientific coherence. The same is true for incompleteness in natural science or technology when modeling coupled referent systems.
- Empirical data necessary for initial research questions may be incomplete, poor, or downright nonexistent. This is a common situation in CSS research because researchers often pose questions that are tractable through computational tools, but no one has collected data necessary to verify or validate the models, thereby requiring adjustments to obtain viable research questions.

[7] The same is generally true of mathematical social science models, and also to some degree of econometric and other statistical models.

- Computational resources may be insufficient for an original set of research questions. This is another common occurrence, especially for overly ambitious projects that fail to estimate the correct amount or types of computational resources. This too usually requires limiting the scope of research questions asked.
- Other practical considerations, such as deadlines, and available personnel, may also condition the formulation of research questions.

The non-computational literature in social science may or may not provide adequate guidance in terms of research questions. This is because the computational approach in general, and the social complexity paradigm in particular, offer different human and social dynamics that are invisible from the perspective of non-computational literature. For example, vast areas of social science are practically defined in terms of a single methodology, such as statistical multivariate models, or game theory models, or general equilibrium models. By contrast, social simulation models address research questions that require any combination of formalisms. That being said, CSS researchers would do well in seeking to address research questions that are recognized as significant by non-computational scientists, as well as other CSS researchers.

Failure to begin with clear and viable research questions guarantees that subsequent complications will require backtracking until proper research questions are posed. This is sometimes inevitable, especially when new territory is being explored. However, such false starts should be avoided when possible, because they can be wasteful along multiple dimensions: time, costs, personnel, and missed opportunities. Scientific discipline and experience are valuable assets in the formulation of research questions in CSS, as in all domains.

A remark on interdisciplinary research in CSS: Research questions addressed through social simulations are frequently interdisciplinary because of multiple reasons. Social complexity respects no disciplinary boundaries! Coupled systems are multidisciplinary by definition. Complex social simulations, in particular, require interdisciplinary research.

8.7.2 Conceptual Design: What Does the Abstraction Look Like?

Given a set of viable research questions, the next step in developing a social simulation is to conduct a process of abstraction that will yield a conceptual model of the referent system. The abstraction itself should be informed and guided by the research questions.

Ideally, the abstraction for producing a conceptual model of a referent system should be guided exclusively by research questions and conducted without regard to consideration of subsequent implementation.

In practice, the abstraction and resulting conceptual model will be influenced by the known implementation resources. This is the tyranny of a hammer looking only for nails. If you know or use only method M, then both abstraction and resulting conceptual model will be shaped (and perhaps completely determined) by M, rather than by research questions, as it should be.

This methodological pathology in CSS research is similar to what happens in non-computational social science when researchers conduct abstractions and produce conceptual models guided primarily by those methods they know or prefer, rather than by what the research questions actually require. This methodological error should be avoided by gaining familiarity with different simulation approaches and a broad range of human and social phenomena—not easy, but well worth it. The abstraction and resulting conceptual model should contribute to answering the research questions, no matter what tools are required.

There is a history lesson to be learned here. A *major* source of methodological innovation comes from *not* having the proper computational tools to answer research questions. Isaac Newton was led to the invention of infinitesimal calculus because he wished to answer research questions for which there were no tools. He refused to adapt the research questions to existing tools or provide only tool-driven answers (like everyone else was trying to do). Likewise, John von Newman did the same by inventing game theory; he wanted to answer research questions having to do with interdependent choices (strategic entanglement), and the extant theory of decisions established by Bayes for answering questions of choice against nature was insufficient. Like Newton and others before him, he became a mathematician, invented game theory as a novel branch of mathematics, and then returned to the social science of interdependent decision-making and formalized it through game-theoretic models. He also invented cellular automata, examined in Chap. 10, which we now use for developing a broad class of social simulations. Simulation systems—from DYNAMO to MASON—were invented with the same science motivation: to enable us to expand scientific frontiers by answering an increasing number of challenging questions.

Different graphic systems have been invented to facilitate specification of a conceptual model. Flowcharts, Forrester diagrams, and UML diagrams are some examples. These are useful for refining ideas and they are indispensable in interdisciplinary projects when specialists from various domains need to develop consensus and common understanding. They will be examined in the context of each model class. No doubt, others will be invented as CSS research increases demand to create clearer conceptual models.

8.7.3 Implementation: How Is the Abstracted Model Written in Code?

The third step in developing a social simulation involves implementing the conceptual model into code. This is where a major decision is made in terms of implementing the conceptual model using native code or a simulation system such as one of those mentioned earlier. The choice is based on multiple considerations, which should include:

Research question Again, research questions should inform implementation, not just the conceptual model. The character of research questions and the resulting conceptual model should first determine whether the simulation model should be variable-oriented (attributes are most prominent) or object-oriented (entities are most prominent) and, second, whether native code or a toolkit should be used.

Expertise Excellence in some implementation solutions may also bring novel answers to research questions. For example, a CSS team highly skilled in building SD models can make significant contributions to a given domain, even if alternative OO (object-oriented) models are possible. Different formalisms of the same referent system almost always bring to light different aspects that advance understanding.

Future use Consideration should be given to future uses that may be envisioned. Such uses include further research, use in teaching, or policy analysis or problem-solving.

The main result of this third step is an *initial version* of a simulation model, which will likely evolve through subsequent *versions*. By convention, the *initial version* of a simulation model is labeled 0.1 or lower. Relatively small, incremental changes prompt *decimal* increases in version numbers, whereas relatively large or major changes prompt *integer* increases—a protocol similar to numbering versions of "the same" software. In general, there are more decimal increases than integer increases.

A social simulation implemented in code should abide by all the principles discussed in Chap. 2 concerning best practices, such as commenting, modularity, defensive programing, multiple backups, and similar guidelines. Code that can no longer be understood even a year after it was written is useless.

In all cases, model code must be committed to some depository. Sourceforce, Googlecode, the Harvard-MIT Data Center (Dataverse), and OpenABM provide examples of online, open-source, code depositories. Besides code files, documentation must also be provided, including all supplementary supporting files. A great deal of effort goes into producing a high-quality model, as we will discuss later in this chapter. However, *simulation code is highly perishable, far more so than mathematical or statistical models*. Unfortunately, it is not uncommon for social simulations—even famous ones—to be lost within a relatively short span of time following their creation. Often all that remains is the conceptual model and some mathematical features.

8.7.4 Verification: Does the Simulation Perform as Intended?

The process of finding out whether a simulation model is working as intended by the conceptual model is known as **verification**, a procedure that also involves **debugging**. This is equivalent to what is traditionally called *internal validity* in non-computational social science formal methodology. An unverified model cannot be used for analysis. Verification is accomplished through multiple procedures, as detailed below. All of them typically unveil bugs hidden in the initial simulation code.

8.7.4.1 Code Walk-Through

Reading code line by line, commenting and refactoring it as necessary, is an indispensable procedure to ensure a simulation is working as intended by both model designers and programmers. Modularization facilitates this procedure, as well as providing other benefits. Code walk-through (also written as walkthrough) should be done while also consulting all relevant prior documentation, including conceptual narratives and diagrams. Again, good programming style resulting from best practices facilitates the code walk-through procedure.

8.7.4.2 Profiling

Another procedure for verifying code is to "profile" it. Profiling means to count the frequency with which key code elements are used, such as various methods or operations in OOP (object-oriented programming) code, or functions in other programming languages. In a sense, profiling is a form of quantitative, automated content analysis or information extraction procedure conducted on code—a means of *mining code* to detect possible errors. The result of profiling is a quantitative summary of findings, such as a frequency histogram of methods or functions called. Formally, the result of profiling code is a rank-size distribution, which resembles the idea behind a Type I Zipfian power-law model. Often it is impossible to draw inferences on the sole basis of profiling results; however, when added to other information from code walk-through, profiling can be a valuable procedure.

8.7.4.3 Parameter Sweeps

Social simulation models typically include large numbers of parameters. Such a large set of space parameters can be used for verification purposes by evaluating the model as a single parameter changes in values while others are held constant. Thus, results from a parameter sweep will provide a response surface which can be plotted and examined for possible anomalies indicative of bugs or other patterns that should not appear. Parameter sweeps can reveal special properties within a range, such as singularities, asymptotic behaviors, oscillations, or other quantitative and qualitative patterns.

8.7.5 Validation: Can We Trust the Results?

The process of finding out whether results from simulation model runs match what is known from empirical data is known as **validation**. Essentially, validation involves pattern matching between simulation output and observed patterns in the referent system.

There are a variety of ways in which simulation validation is conducted. Among the most important and common ones are:

Histograms: Frequency distributions obtained from simulation runs can be matched with empirical histograms—for example, income distributions, the size of spatial distributions, and similar.

Distribution moments: All distributions are characterized by moments, so matching moments generated by simulation runs with real data is another strategy.

Time series: Dynamic social simulations typically produce time-series data from simulation runs, which can be compared with empirical time series.

Special indices: Specific measures, such as the Gini coefficient, entropy, the Hurst coefficient, and similar indices can also be used.

Other: Results from simulation runs produce numerous statistics and patterns that are often characterized by the specific subject matter and can be used to compare with real-world data.

Sometimes an existing simulation system, such as Netlogo, MASON, or Repast, will already have some of these facilities for conducting model validation tests. However, it may be necessary to develop such facilities in the case of frequently used validation tests that are not provided by the simulation system being used.

Ideally, validating a social simulation model is facilitated by pre-existing empirical data that can be used to match results from simulation runs. This is often the case when data from simulation runs also exists in reference to actual empirical data. However, it is not uncommon to discover that simulation results produce data that has never been measured in the real world. In this case, there is no choice but to attempt to collect additional data as necessary. An interesting scientific situation arises when a social simulation produces results that no one has looked for before!

Validating a social simulation model also involves estimating and calibrating parameter values to their appropriate ranges. This is often done by beginning with existing empirical parameter values or informed guesses within a justifiable domain. In the end, validation always involves matching simulated, virtual data, with real, empirical data.

8.7.6 Virtual Experiments and Scenario Analyses: What New Information Does the Simulation Generate?

Earlier we discussed how virtual experiments are a major scientific contribution of social simulation models. Conducting virtual experiments, such as by analyzing alternative scenarios, is an intriguing and exciting use of computational modeling.

Computational experiments using social simulation models can be based on basic scientific research, as well as on applied policy analysis. Analyzing virtual experiments and alternative scenarios is a social simulation tradition that goes back to the earliest days of computer simulation modeling in the social and behavioral sciences. For example, the earliest system dynamics global models were used to analyze industrial development policies and global environmental trends under a variety of future scenarios. While many of the assumptions used in these initial models during the 1970s proved to be incorrect, the methodology itself was powerful and continues to develop to this day.

Conducting virtual experiments through simulation models is also common in other computational disciplines ranging from biology to astronomy. The reason for this affinity between CSS and computational biophysics and the earth and space sciences is the common problem of being unable to conduct real experiments on the referent systems of interest. The only way to understand what happens when two galaxies collide is to conduct computational experiments, much the same as is the case for conducting virtual experiments in computational biology.

8.8 Assessing the Quality of a Social Simulation

Social simulation methodology has begun to generate proposals for assessing and promoting quality across diverse and related areas.[8] For instance, proposals exist in the area of communicating social simulation models, assessing complex projects that involve large interdisciplinary teams (Sect. 8.9), and comparing models (see Sect. 8.10). A strong consensus on a universal set of quality standards in social simulation research has not yet emerged, but such a debate has already begun in the global CSS community.

8.8.1 General Principles for Social Modeling Assessment

The criteria of “Truth”, “Beauty”, and “Justice” have been proposed by Charles A. Lave and James G. March in the classic *Introduction to Models in the Social Sciences* (1993). These criteria are widely used for discerning quality in *social science formal models*, mainly mathematical in kind. The three terms “Truth”, “Beauty”, and “Justice” (or TBJ, for short) are labels for quality dimensions referring to fundamentally good—i.e., normatively desirable—features of social science modeling. Accordingly, the TBJ terms must be interpreted not literally but as labels.

Truth refers to the empirical explanatory content of a model—i.e., its contribution to improving causal understanding of social phenomena—in the sense of developing positive theory. For example, truth is normally judged by internal and external validation procedures, corresponding to axiomatic coherence and empirical veracity, respectively. Truthfulness is the main, classical criterion for evaluating empirical science, whether a model is statistical, mathematical, or computational. Truth must be a constituent feature in a social science model; without it, a model has no overall quality contribution.

Beauty refers to the esthetic quality of a model, to its elegance in terms of properties such as parsimony, formal style, syntactical structure, and similar features. Beauty is about art and form. For example, the mathematical beauty of some equations falls within this criterion, including features such as the style of a well-annotated system of equations where notation is clear, well-defined, and elegant.

[8]This section focuses on social simulations, so the broader field of CSS (e.g., social data algorithms or socioinformatics, complexity models, social networks, social GIS, and related areas of social computing) lies beyond the scope of this section. Quality research in those other areas is subject to its own standards, as discussed in previous chapter.

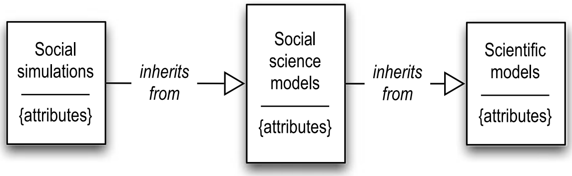


Fig. 8.2 UML class diagram illustrating the hierarchy of scientific models (*left*), social science models (*center*), and social simulations (*right*), each having increasingly specific standards for judging quality (moving from *left* to *right*). *Source*: Cioffi-Revilla (2013)

Unlike truth, beauty is not necessarily a constituent attribute, but is certainly a desirable scientific quality.

Justice refers to the extent to which a model contributes to a better world—to improvement in the quality of life, the betterment of the human condition, or the mitigation of unfairness. Justice is a normative criterion, unlike the other two that are positive and esthetic. For example, a model may improve our understanding of human conflict, inequality, refugee flows, or miscommunication, thereby helping to mitigate or improve social relations and well-being through conflict resolution, poverty reduction, humanitarian assistance, or improved cross-cultural communication, respectively. Policy analysis can be improved by social simulation models that are properly validated.

These Lave-March criteria of truth, beauty, and justice are useful for evaluating the quality of social simulation models. For example, in the classic Schelling model of segregation all three criteria are well-recognized. This is a fundamental reason why Schelling's model is so highly appreciated.

However, a further challenge exists because social simulations have features that render truth, beauty, and justice insufficient as criteria for assessing quality. This is because social simulation models are instantiated or rendered in code (a computer program in some language), so one can easily imagine a social simulation that would be of high quality in terms of truth, beauty, and justice, but fail in overall quality because simulation models pose additional challenges beyond other social science models (i.e., beyond the features of statistical or mathematical models).

As illustrated in Fig. 8.2, social simulations have properties that are shared with all models in science generally and social science in particular, based on inheritance as a specialized class, in addition to having other features of their own. For example, the specific programming language of an agent-based model (Java, C++, or other), or that of a system dynamics model, would be a defining feature.

The inheritance relation between social science models and social simulations readily suggests several key features that distinguish the latter from the former, as illustrated in Table 8.1.

Additional criteria for social simulations—i.e., criteria beyond classical standards for social science models—should allow us to judge quality in terms of "The Good, The Bad, and The Ugly".

Table 8.1 Quality criteria for evaluating models in domains of science

Models in ...	Truth	Beauty	Justice	Additional criteria
Science	Yes	Yes	No	No
Social science	Yes	Yes	Yes	No
Social simulation	Yes	Yes	Yes	Yes

Source: Cioffi-Revilla (2013)

Common required practices, such as verification and validation, are well-known quality control procedures for assessing scientific models in general. However, verification and validation are insufficient criteria for assessing the quality of social science models, specifically for social simulations. An important implication is that current emphasis on model verification and validation is warranted, but *verification and validation are insufficient by themselves for judging the quality of a social simulation model* (agent-based or other).

Therefore, a key methodological question concerning quality is: which additional criteria—i.e., beyond truth, beauty, and justice—could or should be used to assess the quality of a social simulation model? We shall now address this question based on a set of dimensions for evaluating the quality of a given social simulation model.

8.8.2 Dimensions of Quality in Social Simulation Models

The quality of any complex artifact—whether a social simulation model or the International Space Station—is a multifaceted property, not a single dimension. Dimensions of quality can be used for evaluation and can also provide a master checklist of desirable attributes for building and developing a social simulation model. Arguably, there are two levels of quality assessment for computational social simulations corresponding to the concepts of *a model* and *modeling*, respectively.

First, from a model's perspective, any set of quality dimensions for evaluating a social simulation must be based on its specific attributes or uniquely constituent features as a computational artifact in the sense of Simon. Moreover, whether the overall quality of a given model should be an additive or a multiplicative function of individual qualitative features is less important than the idea that overall quality depends on a set of dimensions or desirable features beyond the Lave-March criteria, not on some single preeminent feature (e.g., simulation environment or programming language).

Second, from a modeling perspective, quality assessment should cover the broader modeling or model-building process as such, beyond the social simulation model that is produced in a narrow sense. This is because a computational model in final (i.e., committed) instantiated code is the result of a sequence of earlier modeling stages that precede the model itself, such as the critical stage of model design prior to implementation. *Quality in design affects quality in the product of implementation*, even when implementation *per se* is carried out in a proper manner (i.e., competently, with effectiveness and efficiency).

The following Lifecycle Framework for quality assessment combines both perspectives—the model and its developmental process—by focusing on the classical methodological stages of social simulation modeling, as we discussed earlier in this chapter, with only minor modifications:

1. Formulation
2. Implementation
3. Verification
4. Validation
5. Analysis
6. Dissemination

Such a framework provides a viable checklist of quality dimensions to consider, based on the preceding methodological principles for social simulation research. Note that verification and validation constitute only two contexts for assessing quality and, as shown below, some of the others involve quite a number of *additional aspects* regarding quality evaluation.

1. **Formulation**. Quality can be assessed starting from the formulation of a research problem that a given social simulation is supposed to solve. A first set of quality assessments regards research questions. Is the research question or class of research questions clearly formulated? Is the focal or referent empirical system well-defined? Beyond clarity, is the research question original and significant? Originality should be supported by complete and reasoned surveys of prior, extant literature to assess scientific progress. Every computational simulation model is designed to address a research question, so clarity, originality, and significance are critical. Motivation is a related aspect of problem formulation. Is the model properly motivated in terms of relevant extant literature? Or, is the simulation model the very first of its kind? If so, are there prior statistical or mathematical models in the same domain? Literature reviews in published social simulation research should not be incomplete, poorly argued, or totally missing.
2. **Implementation**. Rendering an abstracted model in code involves numerous aspects with quality-related implications, starting with aspects of instantiation selection. Does the code instantiate relevant social theory? Is the underlying social theory instantiated using a proper program or programming language? Code quality brings up other aspects that may be collectively referred to as the Grimson-Guttag standards: Is the code well-written? Is the style safe/defensive? Is it properly commented? Can it be understood with clarity one year after it was written? In addition, what type of implementation strategy is used? I.e., is the model written in native code or using a toolkit? If a toolkit is used, which one, why, and how good is the application? Is the choice of code (native or toolkit) well-justified, given the research questions? In terms of "nuts and bolts", quality questions include such things as: What is the quality of the random number generator (RNG)? Is it Mersenne Twister, MT19937, or other PRNG? Which types of data structures are used, given the semantics? Are driven-threshold dynamics used? If so, how are the firing functions specified? In terms of algorithmic efficiency, what is the implementation difficulty of the problem(s) being addressed by the model? How efficient is the code in terms of implementing the main design ideas? In terms of computational efficiency, how efficient is the code in

terms of using computational resources? This aspect differs from algorithm efficiency. From the perspective of architectural design, is the code structured in a proper and elegant manner commensurate with the research question? In terms of object ontology, does the model instantiate the object-based ontology of the focal system for the chosen level of abstraction? Note that all these quality-related questions precede verification and validation.

3. **Verification**. Which passive and active tests were conducted to verify that the model is behaving in the way it is intended to behave? Social scientists also call this internal validity. Verification tests include but are not limited to the following: code walk-through, debugging, unit testing, profiling, and other common procedures used in software development, as we have already seen, and will examine more closely in the next chapters. What were the results of such verification tests? Quality assessment should cover investigation of which verification procedures were used, since results can range widely depending on the extent of verification methods employed. Unfortunately, most social simulations are reported without much (or any) information regarding verification procedures, as if it were true that "results speak for themselves"—quite often they do not.
4. **Validation**. Similarly, validation of a social simulation, what social scientists call external validation (or establishing a model's external validity), consists of a suite of tests, not a single procedure. Such tests are important for assessing quality in a social simulation. Which tests (histograms, RMSE for assessing goodness of fit, time series, spatial analysis, network structures, and other forms of real vs. artificial pattern matching tests) were conducted to validate the model? What were the results? Validation tests are often the focus of reporting results at the expense of all other phases in the life cycle of a social simulation model.
5. **Analysis**. The preceding aspects provide a basis for establishing overall confidence in a given model. What is the level of confidence in the model's results, given the combined set of verification and validation tests? If networks are present and significant in the focal system, does the model exploit theory and research in social network analysis (Chap. 4)? Does the model facilitate analysis of complexity as a system of non-linear interactions and emergent properties (Chap. 6)? Which features of complexity (emergence, phase transitions, power-laws or other heavy-tailed distributions, criticality, long-range dynamics, near-decomposability, serial-parallel systems, or other structural features) are relevant to the particular model? If spatial features are significant, does the simulation employ appropriate spatial metrics and statistical tools for spatial data? What is the overall analytical plan in terms of simulation runs and how is it justified? How does computational analysis advance fundamental or applied understanding of social systems? In terms of overall effectiveness, does the model render what is necessary for answering the initial research question(s) or class of research questions? This differs from efficiency. In terms of the simulation's computational facilities, does the model possess the necessary functionality for conducting extensive computational analysis to answer the research questions or even go beyond? How powerful is the model in terms of enabling critical or insightful

experiments, for example in terms of parameter exploration (evolutionary computation) and record-keeping? What is the quality of the physical infrastructure that renders the most effective simulation experience?

6. **Dissemination**. Finally, the quality of a social simulation should be assessed in terms of its "life-beyond-the-lab". For instance, in terms of pedagogical value: Does the model teach well; i.e., does it teach efficiently and effectively? In terms of communicative clarity and transparency, are useful flowcharts and diagrams of various kinds (e.g., UML class, sequence, state, and use case diagrams) provided for understanding the model? Are they drawn with graphic precision and proper style? In terms of replicability, what is the model's replication potential or feasibility? How is reproducibility facilitated? Aspects related to a model's graphics are also significant for assessing quality, not just "eye candy". In terms of GUI functionality, is the user interface of high quality according to its main users? Is the GUI foundational for answering the research questions? More specifically, in terms of visualization analytics, is visualization implemented according to high standards? This does not concern only visual quality, but analytics for drawing valid inferences as well. From a perspective of "long-term care", what is the quality of the model in terms of curatorial sustainability? How well is the model supported in terms of being easily available or accessible from a long-term perspective? In which venue (Google Code, Sourceforge, OpenABM, Harvard-MIT Data Center/Dataverse, or documentation archives such as the Social Science Research Network SSRN) is the model code and supplementary documentation made available? Finally, some social simulations are intended as policy analysis tools. Is the model properly accredited for use as a policy analysis tool, given the organizational mission and operational needs of the policy unit? Does the model add value to the overall quality of policy analysis? Does it provide new actionable information (new insights, plausible explanations, projections, margins of error, estimates, Bayesian updates) that may be useful to decision-makers?

The quality of a social simulation is proportional to the number of dimensions on which it is highly rated. Although these basic dimensions are not independent among themselves, their total contribution is what matters in terms of a comprehensive quality assessment.

8.9 Methodology of Complex Social Simulations

Some social simulations are called *toy models* because they represent a very simple referent system based on research questions that investigate a relatively narrow range of entities and dynamics. Some of the earliest social simulation models belong to this class, and they are still important today because they provide a unique way of understanding fundamental human and social dynamics. For example, toy models such as Heatbugs, Segregation, Hawks and Doves, or Boids—as well as many others provided by Netlogo—have significant pedagogical value for teaching the fundamentals of social simulation science.

Other models consist of **complex social simulations** and are characterized by numerous interacting entities, typically heterogenous in several respects, governed by multiple and typically nonlinear dynamics. Complex social simulations are normally built by interdisciplinary teams with distributed expertise among members. Typical cases in this group include coupled socio-techno-natural systems that require integrated application of knowledge across multiple domains. Such models also typically require years of development work, most often involving multiple research institutions.

The methodology of complex social simulation models requires special consideration in order to exploit the richness of such models while at the same time managing multiple challenges. A viable approach to complex social simulation modeling is to view model development as a spiraling, multi-stage process that proceeds from an initial, simple model and moves toward the much more complex final model. A famous example of this in the history of physical science was none other than Isaac Newton's research program on planetary dynamics (what prompted him to invent infinitesimal calculus), which has been studied in detail by the late Hungarian philosopher of science and mathematics, Imre Lakatos [1922–1974]. As described by Lakatos, Newton worked through a progressive sequence of models—not a single large model—before he arrived at his final, full model of the whole planetary system, complete with planets, moons, and the sun at its center. The initial simple model investigated by Newton bore no resemblance to the final model, except as a minuscule component. His first model consisted of a single perfect sphere rotating around its axis. Subsequent models in a cleverly chosen sequence of "progressive problemshifts" added moons, tilting axes of rotation, elliptical orbits, and numerous other carefully chosen empirical features as Newton approximated his final model of the planetary system. The entire movement from the initial, simple model to the final, complex model resembled the masterfully orchestrated music of Maurice Ravel's *Boléro*, which starts with a single, lonely drum and ends with a huge, full orchestra.

An example of a complex social stimulation, in many ways similar to Newton's final model of the planetary system, would be a coupled socio-techno-natural system. In order to develop such a simulation as a final model of a referent system representing some geographic region, the first initial model would represent a single territorial entity with minimal dynamics included in the simulation. Once such an initial model is well understood, additional features would be added. For example, the second model in the sequence would have heterogeneous agents, in order to understand more realistic cultural dynamics. A third model would add some simple weather dynamics, to further understand biophysical interactions between, say, precipitation and land cover used by agents. The fourth model could include multiple societies over a broader region. Subsequent models would add infrastructure systems and other technological artifacts.

The idea of a sequence of models for developing a complex simulation research program should not be misinterpreted as being a strictly linear process. Occasionally, it is necessary to make corrections and return to an earlier model that overlooked something important, or it may be necessary to develop deeper understanding of simpler dynamics. That being said, the methodology of complex social simulations should have a definite forward thrust, moving from simple (initial model) to complex (final model).

There are several distinctive features of the methodology of complex simulations.

1. It is necessary to identify an initial model that is simple enough to understand in full detail, while at the same time representing a core element of the envisioned final model of the referent system. Note that the very first model may not bear much resemblance to empirical entities, just as in Newton's case a perfect sphere did not represent any real planet.
2. The sequence of models leading up to the final simulation is not arbitrary; it must be carefully designed in order to provide cumulative insights as work proceeds toward the final model. The sequence of simulation models should follow a theoretically meaningful plan, not simply proceed by random accretion and incremental changes without theoretical justification.
3. Verification is an essential activity throughout the whole development process from one model to the next. However, validation should proceed in a very judicious way, lagging behind verification, because if the model is tested through validation procedures that are premature with respect to the final model, what happens is that theoretically significant models might be rejected because they lack sufficient empirical support. This was the case with Newton's initial models in the sequence, which is why he was not as concerned with empirical tests early on in the research program.
4. Defining a final simulation model for the referent system is essential, because a progressive sequence of models can go on indefinitely.

Again, a clear focus on core research questions is essential for governing the development of complex simulations, just as it is for simpler models.

8.10 Comparing Simulations: How Are Computational Models Compared?

Comparative research is a well-developed and fruitful endeavor with a rich history across the social sciences. In fact, the theory and practice of comparative methodology is viewed by many as a defining feature of social science. Systematic comparison of social simulations is insightful and instructive for multiple reasons:

1. *Research questions* investigated through social simulation are clearly highlighted when comparing simulation models because research questions define the simulations themselves.
2. Analyzing *similarities and differences* among social simulation models provides a deeper, more comprehensive way of understanding them.
3. Comparative analysis of two or more social simulations can help identify features such as overlaps, gaps, or questions in need of *further research*.
4. Insights from comparative analysis of social simulations can also be used to clarify and refine *fundamental dynamics*, such as key properties of emergent phenomena in social complexity.

Keeping in mind the three main types of models used across the social sciences—i.e., statistical, mathematical, and computational varieties—it is safe to say that social scientists have learned a great deal from comparing statistical and mathematical models. For example, social scientists often compare various types of statistical regression models, such as when deciding which type to use given a set of hypotheses being tested, or when analyzing results from alternative functional specifications. Another example is provided by comparing game theoretic models, such as the classic taxonomy of 2×2 games pioneered by the late Russian-American mathematical social scientist Anatol Rapoport. Comparing social simulation models is a newer endeavor when compared to statistical and mathematical models.

A first approach to comparing social simulations is based on generic characteristics such as their referent system, type of implementation, level abstraction, and basic science versus applied uses. Each of these features provides ample room for examining similarities and differences among models being compared. Moreover, depending on the purpose of comparison, these features can be investigated in various degrees of detail. For instance, comparing social simulations by type of implementation is something that can be done in coarse terms by simply identifying the programming languages or simulation systems, or it can be much more detailed, comparing architectural features and interaction networks captured by each implementation. Comparison by generic characteristics can also focus on behavioral dynamics, distributions and stochastic processes, forms of emergent complexity, and long-term asymptotic equilibria.

The more advanced comparison of social simulations should focus on detailed examination of ontologies (including details provided in technical diagrams), dynamic processes (for example, by comparing UML sequence diagrams and state diagrams, in the case of agent-based models), as well as numerous other software features.

Comparing social simulation models is also sometimes referred to as **model-to-model** comparison, or **M2M** for short. In the next two chapters we shall examine several examples.

Recommended Readings

H.R. Alker Jr., R.D. Brunner, Simulating international conflict: a comparison of three approaches. Int. Stud. Q. **13**(1), 70–110 (1969)

O. Balci, Verification, validation, and testing, in *Handbook of Simulation*, vol. 10, (1998), pp. 335–393

J.L. Casti, *Would-Be Worlds: How Simulation is Changing the Frontiers of Science* (Wiley, New York, 1997)

C. Cioffi-Revilla, On the methodology of complex social simulations. J. Artif. Soc. Soc. Simul. **13**(1), 7 (2010). Available online

C. Cioffi-Revilla, Comparing agent-based computational simulation models in cross-cultural research. Cross-Cult. Res. **45**(2), 1–23 (2011)

C. Cioffi-Revilla, Computer simulation, in *Leadership in Science and Technology: A Reference Handbook. Volume 1: General Principles*, ed. by W.S. Bainbridge (Sage, Thousand Oaks, 2012), pp. 345–354

B. Edmonds, R. Meyer (eds.), *Simulating Social Complexity* (Springer, Berlin, 2013)

W.G. Kennedy, C.R. Cotla, T. Gulden, M. Coletti, C. Cioffi-Revilla, Towards validating a model of households and societies of East Africa, in *Proceedings of the Fourth World Congress in Social Simulation (WCSS 2012)*, Taipei, Taiwan, Republic of China (2012)

C.A. Lave, J.G. March, *An Introduction to Models in the Social Sciences* (University Press of America, Lanham, 1993)

J. Rouchier, C. Cioffi-Revilla, J.G. Polhill, K. Takadama, Progress in model-to-model analysis. J. Artif. Soc. Soc. Simul. **11**(2–8) (2008)

R.G. Sargent, Verification and validation of simulation models, in *Proceedings of the 40th Winter Simulation Conference, WSC'09*, (2008), pp. 157–169. Available online

R.K. Sawyer, Social explanation and computational simulation. Philos. Explor. **7**(3), 219–231 (2004)

F. Squazzoni (ed.), *Epistemological Aspects of Computer Simulation in the Social Sciences* (Springer, Heidelberg, 2009)

9 Simulations II: Variable-Oriented Models

9.1 Introduction and Motivation

This chapter examines the superclass of *variable-oriented* social simulation models, also called *equation-based* social simulations. Historically, these were the first types of social simulations and they have formal roots in differential equation models of social dynamics. Today, these social simulation models consist primarily of **system dynamic (SD) models** and **queueing models**. Each class is examined using the MDIVVA social simulation methodology (Motivate-Design-Implement-Verify-Validate-Analyze) developed in Chap. 8.

Both of these social simulation models focus on complex social systems *over time*, which makes them applicable to theoretical application for basic science as well as policy analysis. Historically, however, applications to applied operational and management issues have prevailed. Hence, their use for advanced theoretical analysis awaits many fruitful applications, especially in light of experience acquired through practical uses in management, industrial, and operational settings.

9.2 History and First Pioneers

Social simulation models examined in this chapter have scientific roots in Isaac Newton's theory of dynamical systems and Girolamo Cardano's theory of events in probability—a prestigious pedigree. The following summary of major milestones includes developments in SD and queueing models as well as closely related advances in dynamic simulation models more broadly.

1909 Mathematician and engineer Agner Krarup Erlang pioneers scientific research on queuing systems by modeling the Copenhagen telephone exchange.

1953 Statistician and mathematician David G. Kendall proposes the standard formal notation still in use for queueing systems, published in *The Annals of Mathematical Statistics*.

C. Cioffi-Revilla, *Introduction to Computational Social Science*, Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_9,

1958 Richard Bennett at MIT creates SIMPLE (Simulation of Industrial Management Problems with Lots of Equations), the first system dynamics computer modeling language.

1959 DYNAMO (DYNAmic MOdels) v. 1.0, an improved version of SIMPLE, is invented by Phyllis Fox and Alexander Pugh. DYNAMO quickly becomes the formal *lingua franca* of management science and operations simulation models.

1960s SD models become widely adopted in operations research of complex social systems and management science, remaining prominent today.

1961 Engineering scientist Jay Forrester from MIT's Sloan School of Management publishes his pioneering book, *Industrial Dynamics*, the first in a series of SD classics.

1961 Applied mathematician Thomas L. Saaty publishes the queueing theory classic, *Elements of Queueing Theory with Applications*. In the same year J.D.C. Little publishes his famous law of queueing systems in the journal *Operations Research*, and J.F.C. Kingman publishes his equally famous law in *Mathematical Proceedings of the Cambridge Philosophical Society*.

1969 *Urban Dynamics* is published by Jay Forrester and John Collins (former mayor of Boston), expanding system dynamics simulation to social complexity and CSS in a proper sense.

1970 Forrester and his group at MIT create the first socio-environmental global models, WORLD1 and WORLD2, published as *World Dynamics*, of what eventually became the famous Club of Rome model.

1972 *The Limits to Growth*, the classic book that will make SD famous worldwide, is published by Donella Meadows under the sponsorship of engineer Aurelio Peccei's Club of Rome. It is immediately translated into many languages.

1972 Cultural anthropologist Linda S. Cordell pioneers the first social simulation of Puebloan (Anasazi) polities in the American Southwest with her Ph.D. dissertation on "The Whetherill Mesa Simulation" at the University of California at Santa Barbara. Cordell received the Lifetime Achievement Award from the Society for American Archaeology and the A.V. Kidder Medal from the American Anthropological Association, becoming a member of the US National Academy of Sciences in 2005.

1975 Political scientist Dieter Ruloff, disciple of CSS pioneer Daniel Frei from the University of Zürich, Switzerland, demonstrates the first application of SD to simulating insurgency and political stability. In the following years he publishes the first SD models of the collapse of Classic Maya polities and Soviet–Taliban insurgency dynamics in Afghanistan.

1975 Political scientists Nazli Choucri and Robert North publish *Nations in Conflict*, the first discrete-time simulation in international relations, modeling the onset of World War I .

1979 Political scientists Urs Luterbacher and Pierre Allan from the Graduate Institute of International Studies in Geneva, Switzerland, create SIMPEST, the first dynamic simulation model of USA-USSR-PRC strategic triad dynamics during the Cold War, correctly predicting the disintegration of the

Soviet Union. Their paper was presented at the World Congress of the International Political Science Association, Moscow, USSR.

1979 Archaeologists Colin Renfrew and K.L. Cooke co-edit the volume *Transformations: Mathematical Approaches to Culture Change*, another early pioneering collection.

1981 Archaeologist Jeremy Sabloff publishes *Simulation in Archaeology*, one of the first edited volumes of its kind. The same year Nazli Choucri publishes *International Energy Futures*, the first SD modeling book on the world energy market from an economic and politics perspective.

1984 The SD scientific journal, *System Dynamics Review*, is founded.

Mid-1980s Political scientist Michael Wallace publishes a paper demonstrating the implementation of Lewis F. Richardson's theory of arms races in SD models using DYNAMO.

1985 The Stella version 1.0 software for system dynamics modeling is released by the isee systems company.

1998 Nazli Choucri and her MIT students publish the first SD model of state stability in the *System Dynamics Review*.

2000 American management scientist John D. Sterman publishes *Business Dynamics: Systems Thinking and Modeling for a Complex World*, the first major, comprehensive textbook in SD.

9.3 System Dynamics Models

This section introduces the superclass of social simulations based on system dynamic (SD) models, used in significant social science applications, and examines their main features for understanding social complexity. SD models are introduced within the broader context of dynamical systems, which span an even larger class of formal models. The emphasis of SD is on discrete-time systems as the main formalism for characterizing social dynamics of various types observed in referent social systems. Mathematical aspects are important, especially for learning how qualitatively different dynamical processes—i.e., different forms of dynamic behavior—are modeled through different model specifications.

The following terms must be distinguished in the interest of clarity, since they are easily confused when not used with precision:

Definition 9.1 (System Dynamics Model) A system dynamics (SD) simulation is a variable-based *computational* model for analyzing complex systems containing feedback and feedforward dependencies among variables and rates of change, often with high-dimensionality.

Formally, an SD model consists of a system of discrete-time difference equations with forward or backward differencing. SD models can be purely deterministic or contain stochastic noise as defined by random variables. A complete SD social simulation model consists of causal diagrams explaining the network of dependencies and associated code implementation.

Fig. 9.1 Major pioneers of system dynamics models: Jay Forrester, founder of SD modeling (*upper left*); Dennis Meadows, director of the Club of Rome Project on the Predicament of Mankind, *The Limits to Growth* (*upper right*); Linda Cordell, pioneer in dynamical systems models in archaeology, elected to the National Academy of Sciences in 2005 (*lower left*); Nazli Choucri, MIT pioneer SD modeler of energy, conflict, and state stability dynamics (*lower right*)

Definition 9.2 (Dynamical System Model) A dynamical system (DS) is a variable-based *mathematical* model composed of a set of differential equations or differential and integral equations.

Dynamical system models in social science date to the first pioneering applications to the study of conflict, demographic, and economic dynamics almost a hundred years ago—i.e., they were used in mathematical social science much earlier than computational SD models. Formally, a DS model consists of a system of continuous-time equations. DS models can be purely deterministic or contain stochastic noise defined by random variables. Both SD *and* DS are formal models (computational and mathematical, respectively), and can be purely deterministic or contain stochastic components. The main difference lies in the discrete and continuous time domains, as well as the presence of forward and backward time delays in the former.

9.3.1 Motivation: Research Questions

SD models address research questions in numerous domains of CSS, especially those where the following features are present in a given referent system of interest:

1. Variables and their respective time trajectories are of immediate interest as stocks, sizes, or quantities of some kind. (State variables are later abstracted as levels, as detailed in the next stage of the modeling process.)
2. Causal relations among variables are responsible for observed changes in terms of temporal dependencies; they don't just occur for unknown reasons or through purely random mechanisms. (Change is later abstracted as caused by rates.)
3. Noise can affect resulting trajectories at various points in the causal network. (Noise is later abstracted as probability distributions.)
4. At the macroscopic system level trajectories of change can include stationarity, escalation, dampening, cycling, oscillations, asymptotic behaviors, and other temporal qualitative patterns.
5. Emergent properties of social complexity at the systemic level result from interactions at the level of variables at the lowest causal levels.

9.3.2 Design: Abstracting Conceptual and Formal Models

Given some referent system of interest S, a conceptual model C_S, consisting of a set of state variables and their respective rates of change, is abstracted by a two-stage process rendered through causal loop diagrams and stock and flow diagrams.

9.3.2.1 Causal Loop Diagrams

The first stage in SD abstraction to produce a conceptual model focuses on elementary causal relations called *loops*.

Definition 9.3 (Causal Loop) A causal loop is a feedback relation between a given variable x and its rate of change.

Causal loops are the basic elements of an SD model. In turn, feedback can be positive or negative, depending on whether it promotes or impedes a given variable.

Definition 9.4 (Positive Feedback) A positive feedback loop is a causal relation that increases the value of a variable.

Positive feedback is viewed as a **reinforcement dynamic** in SD terminology, producing an increasing effect: growth, expansion, gains, amplification, increases, improvements, enlargements, proliferation, or escalation, or other increasing patterns in the time trajectory of a variable, depending on the appropriate semantics of the referent system.

Definition 9.5 (Negative Feedback) A negative feedback loop is a causal relation that decreases the value of a variable.

Negative feedback is a said to be a **balancing dynamic** in SD terminology, producing a decreasing effect: fatigue, decline, reduction, loss, diminution, mitigation, depletion, contraction, restraint, decay, or other decreasing patterns in the time trajectory of a variable, again depending on appropriate semantics of the referent system.

Definition 9.6 (Causal Loop Diagram) A causal loop diagram is a graphic abstraction that describes positive and negative feedback in the behavior of a given variable.

Norm adoption by members of a community is an example of an emergent social phenomenon that can be represented by a causal loop diagram. This is useful for understanding how a new norm may be adopted as a social process from an SD perspective, as shown in Fig. 9.2. The figure shows two feedback loops operating simultaneously. The positive feedback loop R, on the right, denotes how social conformity tends to produce new norm adopters by peer pressure as the number of new adopters grows. This is a reinforcement dynamic. The more people conforming to the new norm, the greater the pressure to adopt it, which is abstracted as a positive feedback loop. The feedback loop B, on the left, represents negative reinforcement or "balancing" because the community has finite size, so the number of potential adopters decreases as more community members adopt a new norm. The higher the proportion of conformity with the new norm the lower the number of the non-conformists, so the loop on the left represents negative feedback. Related examples of social norms are fashions, opinions, technological innovations, attitudes, and behavioral patterns, so the norm adoption process has broad applicability across domains of social science.

A similar example is found in the domain of *inter-group conflict*, based on *Richardson's two-group rivalry model of arms race dynamics*, shown in Fig. 9.3. (Although this is sometimes referred to as a two-nation arms race model, Richardson intended it to be a general model for conflict between rival groups of any kind, nations and non-state actors alike, as reflected by his term "deadly quarrels.") In this case the rate of arms acquisition by each group is affected by two opposite dynamics produced by feedback loops. On one hand, there is an escalation dynamic because the rate of arms acquisition is driven by a rival's current (and threatening!) level of arms; the higher that is, the greater the need to catch up by increasing one's own rate. On the other hand, there is a mitigating dynamic driven by the cost of maintaining what one already has, so the higher the level of one's own armaments, the

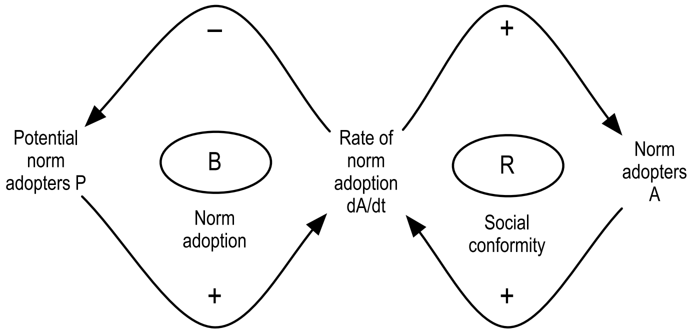


Fig. 9.2 Causal loop diagram for a system dynamics model of norm adoption

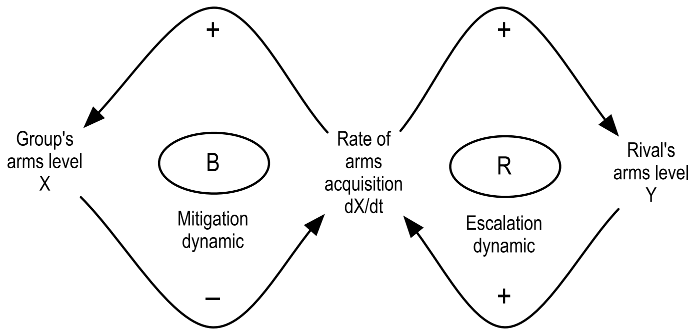


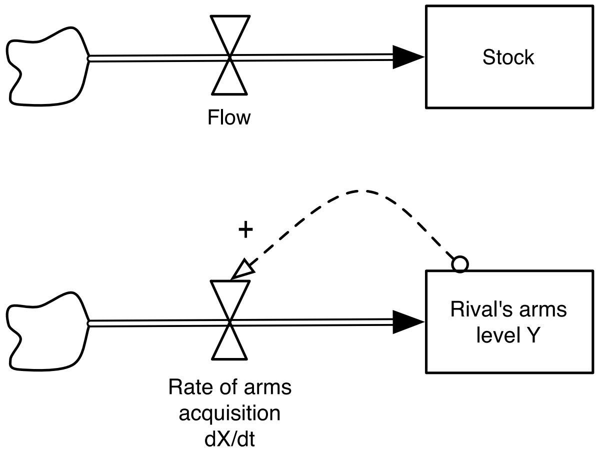
Fig. 9.3 Causal loop diagram for a system dynamics model of inter-group rivalry

greater the economic burden, so the more difficult it is to procure further increases. Today, organization complexity required to support advanced capabilities must be added to direct economic cost. Richardson called this restraining force "fatigue."

A system as a whole is represented by coupled causal loops representing how all elementary causal loops are related to one another. Note that in the last two examples overall system structure is the same, but the signs are not—the balancing signs of the mitigation dynamic are reversed. In the norm diffusion process in Fig. 9.2, the two feedback loops are assumed to be coupled, acting simultaneously. As shown in the diagram, the *rate* of norm adoption is a function of both the *number of potential norm adopters* and the *number of norm adopters*. Potential norm adopters and actual norm adopters are decreased and increased by the adoption rate, respectively. The result is that at different times the two coupled dynamics behave differently. During the early stages of the process, growth in the population of adopters will be greater than in latter stages when fewer non-conformists remain in the community.

In the rivalry process in Fig. 9.3, the two feedback loops are also assumed to be coupled, so they operate simultaneously. The *rate* of arms acquisition is a function of both the *rival's arms level* and the *group's (own) arms level*. The escalation dynamic on the right is a self-reinforcing drive (positive feedback). The mitigation dynamic

Fig. 9.4 SD stock and flow diagram for representing variables (stocks represented as *rectangles*) and rates of change (flow represented as *valves*)



on the left is a balancing drive (negative feedback). However, unlike the previous example, this case assumes *two different kinds couplings*, both acting on the rates of arms acquisitions:

1. Feedback couplings: positive and negative feedback processes are coupled, as in the norm emergence example.
2. Actor couplings: the two rivals are coupled through strategic interaction, in a game-theoretic sense, since the outcome for each (arms levels) is determined not only by what one decides, but also by what the rival decides.

These two coupled dynamics in the rivalry process in this case also behave differently at different times, depending on which dynamic drive prevails.

In sum, causal loop diagrams can contribute to building a conceptual SD model from a qualitative perspective by abstracting positive and negative feedback loops corresponding to reinforcing/escalating and dampening/mitigating drives, respectively. However, more is needed to build a sufficiently complete conceptual model of a referent system that can be computationally implemented in code.

9.3.2.2 Stock and Flow Diagrams

The second stage of abstraction in SD model development is to provide a more quantitative way of representing system structure and dynamics using stock and flow diagrams, as shown in Fig. 9.4. In this second kind of SD diagram, variables become stocks (rectangles) and rates become flow valves (bow ties). Unlike a feedback loop diagram, a stock and flow diagram can be directly translated into code.

The top of Fig. 9.4 shows a generic stock and flow diagram with its basic notation, where the source on the left represents a variable with realization determined by the flow valve that controls the stock or level on the right. The bottom of the figure uses the same notation applied to the case of the reinforcement loop or escalation dynamic of a conflict process (right portion of Fig. 9.3), where a group's rate of acquisition in military capabilities is determined by the level of its rival. The fully coupled conflict system is shown in Fig. 9.5.

Fig. 9.5 Stock and flow diagram for a system dynamics model of a two-group rivalry interaction

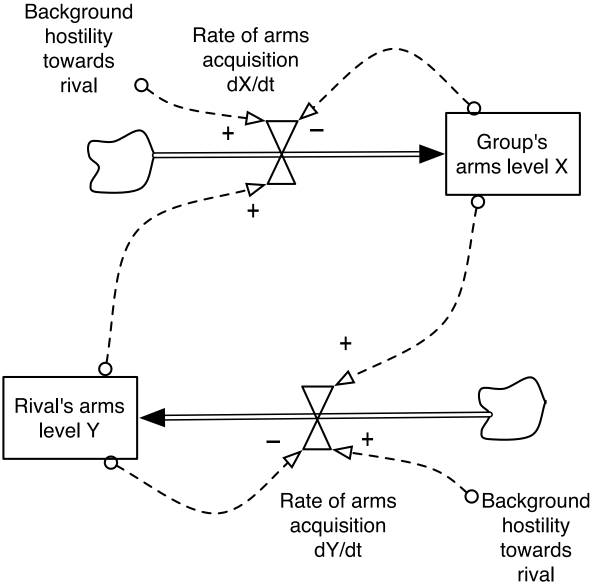


Figure 9.5 specifies how rival actors and feedback loops are mutually dependent on each other, to formalize the concept of strategic interaction. The figure uses the same basic stock and flow components as in Fig. 9.4, with the added element of background hostility acting as a parameter that also affects the rate of change, so now the dynamic process of each rival is driven by three factors:

1. The rival's current arms level (representing positive feedback, escalation force)
2. The group's own arms level (negative feedback, mitigation force) and
3. Background hostility acting as a constant background force, which captures the idea that a group would acquire some minimal military capabilities as insurance, regardless of a rival's arms level.

Diagrams such as these—usually involving many more stocks/variables, flows/rates, and parameters—are used in SD methodology for representing a conceptual model of a given referent system. Noise, stochastic shocks, and other elements are also added as necessary.

The main result of the design stage in system dynamics is a conceptual model of the referent social system specified by a set of equations. For example, in the conflict model, the following system of equations in continuous time specifies the rivalry dynamics:

$$\frac{dX}{dt} = aY - bX + g \tag{9.1}$$

$$\frac{dY}{dt} = \alpha X - \beta Y + h, \tag{9.2}$$

where a and α are reaction coefficients, b and β are mitigation coefficients, g and h are hostility coefficients, and X and Y are levels of armaments. The following system of equations is in discrete time:

$$X(t+1) = aY(t) - bX(t) + g \tag{9.3}$$

$$Y(t+1) = \alpha X(t) - \beta Y(t) + h. \tag{9.4}$$

In this case, the system of equations can be analyzed to obtain closed form solutions, since the system is simple. Solutions to these systems of equations yield time trajectories containing exponential terms, which can be easily verified. In most cases this is not possible, which is why simulation is required.

9.3.3 Implementation: System Dynamics Software

Given a sufficiently complete conceptual model of a referent system, the next stage in SD methodology consists of implementing the model in code using a simulation system. The key milestone activity in the implementation stage is marked by the transition from mathematical equations in the conceptual model to code in the simulation model.

The current, most utilized simulation system for implementing SD models is called VENSIM, which is the current successor to earlier DYNAMO and STELLA simulation systems software. Vensim PLE is an education version that is made available free of charge. The classic textbook by John D. Sterman, *Business Dynamics*, includes a CD (for PC and Macintosh) containing simulation software and models, including ithink, Powersim, and Vensim software. A major advantage of systems such as these is their close association with the SD community, specifically the System Dynamics Society. The Vensim website has numerous resources for beginning and advanced users, including tutorials and other helpful materials.

Figure 9.6 shows a screenshot of the Vensim system while implementing a conceptual stock and flow model of a simple customer base in a company. While Dynamo was a programing language that required writing code, Vensim can be used by selecting facilities for defining variables, equations, and other components by clicking options, using drop-down menus, and other features of the user interface.

Another option for developing SD social simulation is to implement the conceptual model in simulation systems such as Netlogo or Repast. Although these simulation systems were not originally designed to run SD models, they do have such facilities in addition to the agent-based models for which they were originally designed. For example, Netlogo has demonstration SD models for exponential growth,

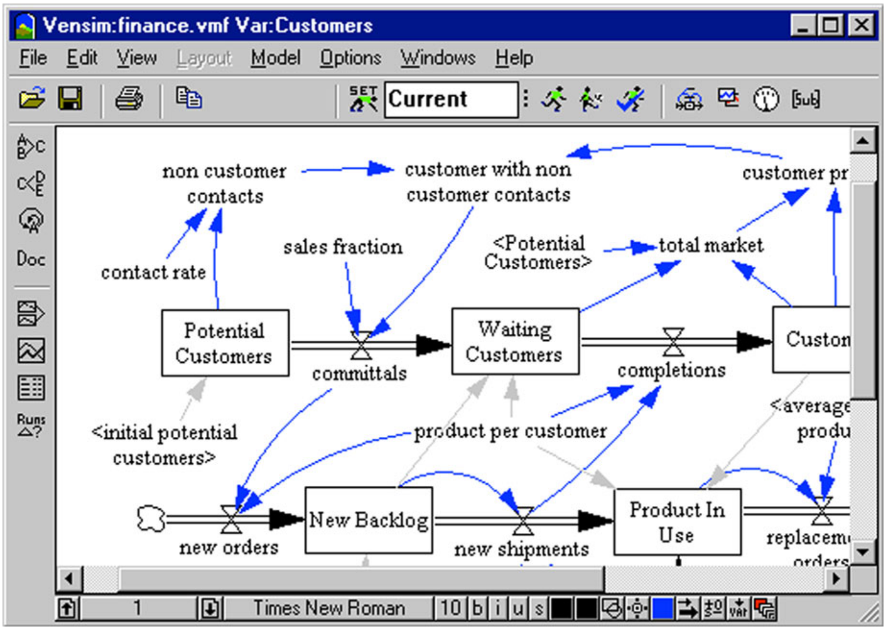


Fig. 9.6 Screenshot while implementing an SD social simulation using the Vensim system

logistic growth, prey-predator (wolf-sheep) dynamics (based on the classic Lotka-Volterra model), as well as other effective examples.

9.3.4 Verification

Recall the difference between verification and validation: the former is about ensuring a model is running the way it is supposed to, as guided by the conceptual model and any other simulation design specifications; the latter is about ensuring that the simulation model is a good representation of the referent system.

Once an SD social simulation model has been implemented, the next step involves verification procedures. Systems such as Vensim provide a number of facilities for verifying a model, such as checking that the right units are specified, rates are using the proper dependencies, and similar steps to ensure that the model is running the way it was intended by the conceptual model. Since an SD conceptual model, complete with stock and flow diagrams, uses the iconic metaphor of levels and flow valves, verifying an SD implementation essentially means checking that all "the plumbing" is working as it should according to the most minute details in the blueprints (stock and flow diagrams). Each element must be checked for accurate implementation, as well as every rate, feature, and connection. Facilities provided by whatever simulation system is chosen should be used in the context of the verification procedures examined in Chap. 8.

9.3.5 Validation

Validating an SD social simulation model that has been verified is accomplished from two main perspectives. **Structure validity** refers to internal features of the model, including all assumptions, relevant variables and their units, and the system of equations in all its component stocks and flows. The following are recommended tests of structure validity for SD models:

Empirical tests of validation: This is aimed at validating the specification of equations used in the model as well as parameter values being used. For example, in the case of the conflict model discussed earlier, this part of the validation process would focus on parameters such as the equation's coefficients being assumed, as well as constants, such as background hostility that affects armament rates. The equations themselves require validation, since different specifications will yield different results. The classic rivalry model assumes additive and symmetrical armament levels, which is an assumption that requires validation through using empirical tests. It is also assumed that reaction coefficients and hostility parameters are constant. These all add up to an overall assumption of structural stationarity, in the sense that all equations specified do not undergo significant change over time—i.e., the standard model assumes that the basic clockwork mechanism does not change as history evolves, which may or may not be a valid assumption.

Theoretical tests of validation: Model assumptions should also be confirmed by the extant theories being used, since even the simplest SD model assumes theoretical mechanisms that justify its causal structure. This is a broader perspective than empirical tests of structural validity, since it is based on fundamental causal arguments that are difficult if not impossible to quantify. For example, in the case of the conflict model, the overall structure is grounded on Richardson's theory of how rivalry between two groups is explained. The fundamental theory is based on three factors or dynamics driving the conflict process: escalation forces driven by positive feedback from a rival's stock of weapons; mitigation forces driven by fatigue and negative feedback from one's own stockpile of armaments; and some background constant force generated by hostility over disagreements and insecurity. Is this theory valid? Are there other factors as important or even more significant than these? The theory also assumes perfect symmetry between rivals; both make arms procurement decisions in the same way. Is it possible that the rivals in question decide with different goals, such as one trying to "catch up" with the other, so it reacts to the *gap* between its own level and the rival's [i.e., $dX/dt \propto (X - Y)$], not simply to the rival's level ($dX/dt \propto Y$ as in Eq. (9.1))?

As with any other kind of social simulation model, tests of structural validity for SD models are complex and require considerable attention. The empirical literature is of great value in navigating through these procedures.

By contrast, **behavior validity** concerns the actual results of simulation runs, primarily in terms of qualitative and quantitative features such as patterns of growth, decay, and oscillation, among others. Many of these procedures involve various forms of time-series analysis and extensions. Some of these were mentioned during the general methodological discussion in the previous chapter, including analyzing trends, comparing periodicities by means of autocorrelation functions, comparing distribution moments, and computing global statistics such as the discrepancy coefficient between simulated and observed time-series data (Barlas 1996: 207–208).

9.3.6 Analysis

The main goal of simulation research in CSS is to obtain qualitative and quantitative results to better understand the referent system. The previous forms of qualitative and quantitative analysis are primarily procedural, for purposes of gaining confidence in the veracity of a model by conducting verification and validation procedures. Obviously, the main goal of developing an SD social simulation—the reason for going through all the trouble—is to analyze it in substantive ways. Formal analysis, asking what-if questions, and scenario analysis constitute major forms of analyzing SD social simulations.

Formal analysis of an SD model yields results, such as *time trajectories* for all level variables (stocks), *phase portraits* in parameter spaces, *sensitivity analysis*, *comparative statics*, and similar sets of results in dynamical systems analysis. For example, the conflict model results from formal analysis would show the time trajectories of levels of armaments in the evolution of conflict between groups, phase portraits of trajectories as a function of parameter combinations, and similar qualitative and quantitative results. Results from formal analysis can reveal properties such as orbits, singularities, asymptotes, attractors, gradient fields, periodicities, chaos, bifurcations, ergodicities (equality between time averages and space averages), phase transitions, stability properties, and other significant dynamic features of social complexity that are typically not apparent from the model structure.

Asking **what-if questions** is another major approach to analyzing SD social simulations. For example, in the conflict model we may ask what happens when the hostility of one group versus its rival is some multiple of the other's hostility. Or, what happens when reaction coefficients differ significantly across the two groups? What-if questions can also extend to analysis of an SD model with alternative specifications of equations to explore what happens when rates of change are governed by different dynamics. For example, as was suggested earlier, in the conflict model we may wish to have a rival responding to the gap $(Y - X)$ in armament levels, as opposed to the original assumption of responding to just level Y.

A more comprehensive form of analysis used with SD social simulations is **scenario analysis**, which typically involves a suite of related questions defining a given scenario, rather than analyzing one question at a time. For example, in the conflict model we may wish to examine a scenario in which reaction coefficients are relatively small, mitigation coefficients are several times larger than reaction coefficients, and hostility coefficients are weak. Intuitively, such a scenario should lead toward lowering of the conflict by de-escalation and disarmament. The opposite scenario would have the set of coefficients changed in opposite ranges, leading to escalation and the system spiraling out of control (blowing up). Within these two extreme scenarios lie many others with interesting qualitative and quantitative properties, some of which are previously known through analytical methods that yield a closed-form solution—many more are not known and remain to be explored, especially in high-dimensionality systems with many actors and different structural specifications in terms of reaction dynamics.

These and other forms of analysis are used in SD simulation to investigate basic CSS questions as well as applied policy issues. SD can also be used in combination with other simulation models, such as agent-based models examined in the next chapter.

9.4 Queueing Models

This section examines the superclass of social simulations that use queuing models, covering their significant social applications and main features. The emphasis is on distributions as the main feature for characterizing queues of various types of processes observed in referent social systems. As always, mathematical aspects are foundational, especially for learning how qualitatively different process structures, representing different forms of randomness, are modeled by different probability distribution laws.

Definition 9.7 (Queue) A system consisting of one or more units or stations that service or process a stream of incoming demands or requests is called a queue. Formally, using Kendall's notation, a given queue Q is denoted by a triplet $A/S/C$, where A describes time between arrivals to the queue, S describes servicing or processing, and C is the number of processors, where $C = 1, 2, 3, \ldots$.

This initial definition is useful by itself, and provides the basis for more complex systems with multistage queues, as we will demonstrate with examples.

9.4.1 Motivation: Research Questions

Queue-like systems are ubiquitous and significant across domains of social science. Consider the following examples:

1. A *bank* (the classic example given in many queueing theory textbooks) is a queueing system where customers arrive with frequency A; they are served by tellers in time S; and there are C teller windows to service customers. If a teller cannot satisfy the customer, there would be another queue for speaking with a bank manager or supervisor. Supermarkets, fueling stations, hospitals (including emergency rooms embedded within), and registration desks are other common everyday examples.

Fig. 9.7 Major pioneers of queueing systems models: Agner Krarup Erlang, founding pioneer of queueing models (*upper left*); David George Kendall, inventor of the standard notation for queueing system models (*upper right*); Thomas L. Saaty (*lower left*), author of classic works in applied discrete mathematics, including queueing theory, and inventor of the Analytic Hierarchy Process; John Dutton Conant Little, discoverer of the law of mean arrival times and a founder of modern marketing research (*lower right*)

2. An *airport* check-in counter (and many other transportation nodes) is a queueing system where passengers arrive with a certain pattern A; the airline staff at the counter, or check-in machine, process passenger identification, flight information, and provide boarding passes in time S; and there are C counters with staff to assist passengers. Flight operations consist of other queuing systems, which are separate, albeit coupled, for processing arriving airplanes from entering the air space to the arrival gate. Modern airports are highly sophisticated, complex queuing systems.

3. A *polity* can be modeled as a queueing system where public issues arise with some temporal pattern A; each issue is addressed with policies S involving resources, processing time for decision-making, and implementation; and which uses a set C of agencies.
4. When a *disaster* occurs in a given society, demand for relief A increases significantly, which requires the immediate activation of emergency response services and humanitarian assistance supply chains S through multiple organizations C. The results are significant for societal welfare and even governmental stability, as seen in Haiti following the 2010 earthquake.
5. A *legislative body* is a queueing system where bills are introduced with frequency A and laws are passed in time S supported by C legislators and staff members.
6. *Human information processing* can be viewed as a queueing system where information arrives at rate A, is processed (decoded and interpreted) at rate S, and makes use of C cognitive elements (values, goals, belief systems, and heuristics, among other elements of bounded rational actors).

The key to recognizing queuing systems in human and social dynamics is identifying the $A/S/C$ pattern in a referent system of research interest. Courts, markets, organizations, and a vast array of institutions provide additional examples. Queuing systems are abundant in social systems and processes. Note that the entities processed or serviced by a queue can be human agents (customers, passengers, shoppers) as well as other socially relevant entities, such as laws and public issues, among many others, as suggested by the examples above.

The most obvious research questions that arise in queueing systems concern patterns of arrival A and servicing S, which are typically expressed in the form of distributions, as well as the organizational arrangement among the C processing components. Given some queuing system Q,

- What are the patterns of arrival and service times in terms of distributions and moments?
- Does the system have sufficient capacity for processing demands within reasonable time?
- Are patterns of arrival and service stationary with respect to epochal time τ?
- If non-stationary patterns exist, how can they be described?

Each of these questions in fact represents a whole set of research issues that are investigated through queuing models in CSS. For example, the question concerning the capacity of a given polity for dealing with a relentless stream of public issues that arise in the normal life of a society (example 3 in the list above) is anything but purely theoretical, although it may sound that way at first. A country that is overwhelmed by unresolved public issues and unattended policy demands will eventually experience state failure, *ceteris paribus* (all other variables held constant). In another example, people get killed when a stampede of panicked individuals seeks to exit a stadium, church, discotheque, or theater when some frightening event has occurred within. This happens because all of a sudden $A \ggg S$, whereas the system is normally designed for $A \leqslant S$ or, at best, $A \approx S$, from a queuing systems perspective.

From the preceding discussion it should be apparent that there are multiple theoretical and policy applications of queuing systems in CSS. That being said, applications of queuing systems to social simulation domains have prevailed in applied areas, such as management science and operations research, with fewer applications investigating basic theoretical questions in social theory (such as examples 3–6 in the list above). Such an imbalance is unjustified, as we will demonstrate by examining the process of social simulation model development from a queuing systems perspective.

There are also purely technological queueing systems, such as the physical Internet, with which we are not concerned. Queuing systems are also important in the context of coupled socio-techno-natural systems, especially in terms of social and technological components and all three coupling interfaces.

9.4.2 Design: Abstracting Conceptual and Formal Models

Given a referent system of interest, the next step toward developing a social queuing system simulation consists of identifying and abstracting relevant information for purposes of developing a conceptual model of the referent system. Based on Definition 9.7, the following three variables each require empirical identification and formal specification:

Definition 9.8 (Arrival Time A) Arrival time A is a continuous random variable defined by a probability density function $p(t)$, or p.d.f., with realizations $\{t_1, t_2, t_3, \dots, t_n\}$.

Definition 9.9 (Service Time S) Service time S is a continuous random variable defined by a p.d.f. $p(s)$ with realizations $\{s_1, s_2, s_3, \dots, s_m\}$.

Note that:

1. Both A and S are c.r.v.s (continuous random variables) measured in time units.
2. Accordingly, arrival and service are also defined by all probability functions formally derived from a p.d.f. $p(x)$, such as (1) the cumulative probability function (c.d.f.) $\Phi(x)$, (2) the intensity function $I(x)$, also known as the hazard rate function $H(x)$ or social force function $F(x)$, (3) the stress function $\Lambda(x)$ as the integral of $I(x)$, and (4) others, as defined earlier in Chap. 7. These other probability functions are important because each describes a different, specific facet of randomness that is important to understand.
3. All probability functions of A and S can be estimated from empirical data, using various methods, although some purposes require more data than others.
4. Density functions $p(t)$ and $p(s)$ provide precise descriptions of numerous forms of randomness, including the special case of deterministic arrival or service, as we shall examine below. Empirical data and social theory should be used for choosing distributions, not purely mathematical or algorithmic convenience.
5. Statistical moments m_i also characterize a given distribution, most importantly $m_1 = \overline{x}$ (mean), $m_2 = \sigma^2$ (variance), $m_3 =$ skewness, and $m_4 =$ kurtosis. The median and the mode are also useful, especially since many social distributions are not normal.

6. Empirically, A is often exponential while S is often normal, at least to a first approximation. The Weibull distribution is also significant for many social processes, as explained below.

The number of service or process components is the third element of a queueing system, based on Definition 9.7.

Definition 9.10 (Service Components C) The number of service components C in a queueing system is a discrete variable with finite integer values $1, 2, 3, \dots, k$.

Following Kendall's notation, the following are important *elementary types* of social queueing systems:

$$Q_1 = M/D/1 \tag{9.5}$$

$$Q_2 = M/M/1, \tag{9.6}$$

where M denotes a Markovian or memoryless process with simple negative exponential (Poissonian) arrivals, D denotes a deterministic processing time, and $C = 1$ component processing node. Equation (9.5) specifies a queueing system characterized by: Markovian (exponential) M arrivals given by $p(t) = \lambda e^{-kt}$, where k is the arrival rate measured in number of arrivals per unit of time; deterministic D (constant) time is required to process each arrival; and a single processing component. Equation (9.6) defines a similar but different queue with the same arrival and component features but processing is Markovian.

The Weibull distribution (Fig. 9.8) is also socially significant, because it includes the simple exponential distribution, an approximation of the normal distribution, as well as a variety of qualitatively different intensity functions that are applicable to many social systems and processes. The Weibull distribution is defined by the following probability functions:

$$p(x) = \kappa x^{\alpha} \exp\left(-\frac{\kappa}{\alpha} x^{\alpha+1}\right) \tag{9.7}$$

$$\Phi(x) = 1 - \exp\left(-\frac{\kappa}{\alpha} x^{\alpha+1}\right) \tag{9.8}$$

$$I(x) = \kappa x^{\alpha} \tag{9.9}$$

$$\Lambda = \frac{\kappa}{\alpha} x^{\alpha+1}, \tag{9.10}$$

where κ and α are scale and shape parameters, respectively.

Numerous mathematical results exist for queueing systems, although the theory of an $M/G/k$ queue, where G is a generic probability distribution, remains incomplete—so simulation methods are appropriate for obtaining computational solutions. The following eponymous laws are among the better known for queues with one or more generic probability distributions G:

Little's law Average number of units being processed in a $G/G/1$ queue in steady-state.

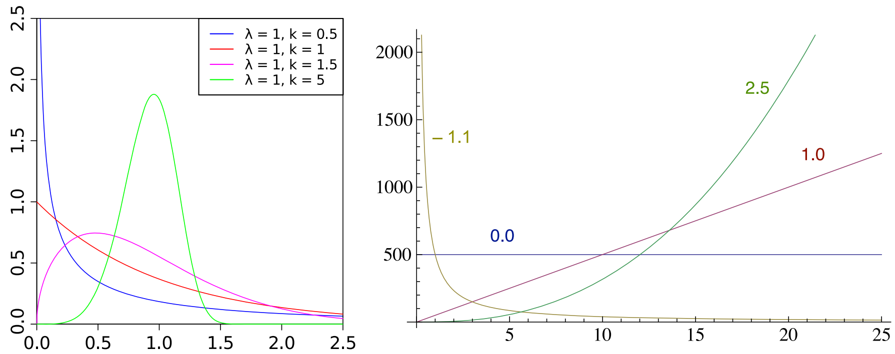


Fig. 9.8 The Weibull distribution. Probability density functions (*left*) and associated event intensity functions (*right*) shown for different values of the shape parameter. The Weibull distribution reduces to the exponential distribution when the shape parameter is 1.0 and approximates the normal distribution when the shape parameter is around 5

Pollacsek-Khinchine's equation Expected waiting time for a $M/G/1$ queue.
Kingman's formula Expected waiting time for a $G/G/1$ queue.

The behavior of processing components in a queue matters greatly and can be based on various specific **scheduling policies**. These are usually described in terms of agents, but they can refer to any entity being processed by a queue.

First-in-first out (FIFO) The agent with longest waiting time is served first.
First-in-last-out (FILO) The agent who arrived first is served last.
Last-in-last-out (LILO) The agent who arrived last is served last.
Last-in-first-out (LIFO) The agent with shortest waiting time is served first, or stack.
Sharing Processing capacity C is shared equally among agents.
Priority Agents are processed according to some ranking.
Fastest job first The agent with the shortest processing time is served first.
Preemptive Processing is interrupted to permit servicing a priority agent.

Some queuing policies, such as FIFO and LIFO, are also used in accounting systems. The LIFO stack was discussed in Chap. 2 as a data structure.

Simple, unitary queues are important for understanding how queue-based processes operate under various probability distributions. However, it is often the case that a real-world referent system will contain a **network of queues**, as well as internal queues embedded within larger queues as in a **system-of-systems**. A common example would be a bookstore where one would enter, browse, select some books, and then proceed to the cashier for payment. The bookstore as a whole is a queueing system where one enters, shops, and departs. But within the store, the time spent browsing, as well as the time spent paying, constitute queues within the "macro" store-level queue. The same is true in the example of a polity that processes public issues, and within governmental institutions, laws and policies have their own, internal processing dynamics. Assuming an exponential onset of issues as well as policymaking and implementation—which is a reasonable approximation in many cases—an abstract queue-based model of a polity can be presented as an $M/M/k$ system, where k denotes the number of government agencies involved. Hence, a network of this kind has links provided by probability distributions and nodes by distributions and service stations.

9.4.3 Implementation: Queuing Systems Software

Social simulations based on queuing models can be implemented in native code or by using a specialized simulation system. There are scores of simulation software packages for queueing systems—not counting some ingenious spreadsheet implementations (not really recommended). Two simulation systems that are frequently used are the Queuing Package for GNU Octave and a suite of queueing modeling software in the Java Modeling Tools, both available at Sourceforge.

9.4.4 Verification

Verifying a social simulation using a queuing model involves several aspects. First, it is good practice to verify that proper ranges are being used for arrival and service random variables, as these need to be positive real values. Second, results need to be consistent with assumed parameter values and at least the qualitative form of probability distributions being used. For example, a queue that has Weibull arrival time A with shape parameter value 2.0 should show an intensity function that is approximately linear, corresponding to the Rayleigh distribution.

More specifically, verifying a social queuing model usually begins with verifying that the entities being processed or serviced (whether they are human agents or other entities such as public issues, vehicles, or other) are being generated in a proper way. This means verifying that the relevant probability distribution function is operating properly. Features to check include low or high arrival frequencies, as well as any temporal clustering that is deemed significant in the conceptual model. Arrival volume might also be variable, which should also be verified.

Common sense is one valid way to verify a queue-based model. For instance, changes in arrival and service times should have direct and measurable effects on the length of queues, otherwise something is wrong with the implementation. Another way is to have the implementation checked by someone other than the coder, which is a general verification procedure along with others discussed in the previous chapter.

Another feature of the model to verify concerns the possible presence of bottlenecks, saturation effects, and issues regarding processing capacity. For example, bottlenecks tend to produce departures at an approximately constant rate. Models of pedestrian traffic as well as vehicular traffic use many of these considerations in terms of verification standards. When agents have a choice in terms of which station or service node to use, decision-making rules must be properly verified. When multiple components are used, such as in a system of standby backup service stations, switching mechanisms for engaging backup units must also be carefully verified.

In sum, verification of a social simulation based on a queuing model always depends on the structure and details of the queuing network system. Each component and the overall organizational structure must be verified at a level of detail required by the research questions being asked.

9.4.5 Validation

The principal way of validating a social simulation based on a queuing model is to match simulated distributions with real-world empirical distributions. Face validity, as always, is a fundamental way of assessing a queueing model, and should be tried first, just as in all other social simulations. Direct familiarity with the referent system is fundamental for establishing face validity. Common technical ways of validating a queueing model include assessing goodness of fit using statistics, regression analysis, distribution moments, time series analysis, and Monte Carlo (MC) methods.

9.4.6 Analysis

Analysis of queue-based social systems and processes from a purely theoretical perspective is vastly undeveloped in social research. This is because there is a paucity of social theory that has been implemented in queueing models as opposed to other areas of social simulation. There are multiple reasons for this. One is that social scientists have favored other forms of formalization rather than making extensive use of probability distributions to model randomness. Another is simply lack of familiarity with the scientific potential offered by queueing systems. Finally, there is a misconception that this class of models is mostly intended for managers and systems operators. This is a fertile area for novel forms of analysis in CSS.

By contrast, analysis of queue-based social systems and processes from an applied, operational perspective is highly developed in management science, operations research, and related disciplines. Traffic flows, customer servicing systems, hospitals and healthcare facilities, supply chains, industrial production systems, and numerous other domains have benefitted from decades of practical applications that have improved many real-world systems through optimization, increasing resiliency, and numerous other improvements ranging from trivial to vitally important.

Recommended Readings

On System Dynamics (and Dynamical Systems)

Y. Barlas, Formal aspects of model validity and validation in system dynamics. System Dynamics Review **12**(3), 183–210 (1996)

N. Choucri, R.C. North, *Nations in Conflict: National Growth and International Violence* (Freeman, San Francisco, 1975)

N. Choucri, *International Energy Futures: Petroleum Prices, Power, and Payments* (MIT Press, Cambridge, 1981)

N. Choucri, D. Goldsmith, S. Madnick, J.B. Morrison, M. Siegel, Using System Dynamics to Model and Better Understand State Stability. Paper presented at the 25th International Conference of the System Dynamics Society, Boston, MA. MIT Sloan School working paper 4661–07, 7/1/2007

J.W. Forrester, *Industrial Dynamics* (MIT Press, Cambridge, 1961)

J.W. Forrester, *Principles of Systems* (Wright-Allen Press, Cambridge, 1968)

J.W. Forrester, *Urban Dynamics* (MIT Press, Cambridge, 1969)

J.W. Forrester, *World Dynamics* (Wright-Allen Press, Cambridge, 1973)

S. Gavrilets, D. Anderson, P. Turchin, Cycling in the complexity of early societies. Cliodynamics: Journal of Theoretical and Mathematical History **1**(1), 58–80 (2010)

R.A. Hanneman, *Computer-Assisted Theory Building: Modeling Dynamic Social Systems* (Sage, Newbury Park, 1988)

B.B. Hughes, E.E. Hillebrand, *Exploring and Shaping International Futures* (Paradigm Publishers, Boulder, 2006)

C.L. Lofdahl, *Environmental Impacts of Globalization and Trade: A Systems Study* (MIT Press, Cambridge, 2002)

U. Luterbacher, Simulation models, global environmental change, and policy, in *International Relations and Global Climate Change*, ed. by U. Luterbacher, D.F. Sprinz (MIT Press, Cambridge, 2001), pp. 183–197

D.H. Meadows, D.L. Meadows, J. Randers, W.B. William III., *The Limits to Growth: A Report to the Club of Rome's Project on the Predicament of Mankind* (New American Library, New York, 1974)

J.D. Sterman, *Business Dynamics: System Thinking and Modeling for a Complex World* (McGraw-Hill, Boston, 2000)

P. Turchin, *Historical Dynamics: Why States Rise and Fall* (Princeton University Press, Princeton, 2003)

A. Wils, M. Kamiya, N. Choucri, Threats to sustainability: simulating conflict within and between nations. System Dynamics Review **14**(2–3), 129–162 (1998)

On Queueing Systems

P. Bratley, B.L. Fox, L.E. Schrage, *A Guide to Simulation*, 2nd edn. (Springer, New York, 1987)

L. Kleinrock, R. Gail, *Queueing Systems: Problems and Solutions* (Wiley-Interscience, New York, 1996)

W. Kreutzer, *System Simulation: Programming Styles and Languages* (Addison-Wesley, Sidney, 1986)

T.L. Saaty, *Elements of Queueing Theory with Applications* (Dover, New York, 1961)

J.A. Sokolowski, C.M. Banks (eds.), *Handbook of Real-World Applications in Modeling and Simulation* (Wiley, New York, 2012)

B.P. Zeigler, H. Praehofer, T.G. Kim, *Theory of Modeling and Simulation* (Academic Press, San Diego, 2000)

Simulations III: Object-Oriented Models 10

10.1 Introduction and Motivation

This chapter examines the superclass of *object-oriented* social simulation models, also called *object-based* social simulations. The main families of simulation models in this area of CSS consist primarily of **cellular automata models** and **agent-based models**. As in the previous chapter, each will be examined using the MDIVVA social simulation methodology (Motivate-Design-Implement-Verify-Validate-Analyze) developed in Chap. 8.

Both families of object-oriented social simulation models use the *simplest social entities* (cells or agents, respectively) as elementary units to understand emergent complexity, rather than variables (as in system dynamics and queueing models). Both families are applicable to theoretical research for developing basic science, as well as practical application for policy analysis, as was the case before for variable-oriented models. Historically, agent-based models have enabled theoretical as well as policy applications, whereas cellular automata models have been more confined to theoretical analysis. However, this is a broad generalization regarding the majority of research. Policy applications of cellular automata models also exist, as we will examine in this chapter.

10.2 History and First Pioneers

Object-oriented social simulation models presented in this chapter have scientific roots in John von Neumann's theory of automata and Thomas Schelling's social segregation model. The following summary of major milestones includes developments in cellular automata (CA) and agent-based models (ABM) and some closely related advances in areas such as organizational and spatial models, including geographic information systems (GIS). The chronology is unavoidably incomplete after the late 1990s, when the field exploded (exponentially) with a doubling time of just a few years.

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1_10,

1940s John von Neumann [1903–1957] and mathematician Stanislaw Ulam [1909–1984] pioneer the theory of automata, publicly presented for the first time in 1948 and published in 1951 as *The General and Logical Theory of Automata*.

1949 Sociologist James M. Sakoda pioneers CA modeling in the social sciences in his doctoral dissertation on "Minidoka: An Analysis of Changing Patterns of Social Interaction" at the University of California at Berkeley, published in 1971 in the *Journal of Mathematical Sociology*, calling it a "checkboard model."

1960s Computer scientist Edward Forrest Moore [1925–2003] invents the concept of 8 neighbors surrounding a given cell in a CA landscape, providing an alternative to the 4-neighbor von Neumann neighborhood.

1966 The University of Illinois Press publishes *The Theory of Self-reproducing Automata* by von Neumann.

1969 Mathematician Gustav A. Hedlund publishes his influential CA paper on symbolic dynamics in the journal *Mathematical Systems Theory*.

1969 Economist Thomas C. Schelling publishes his first CA segregation modeling work in the *American Economic Review*, among the leading journals in economics.

1970 Mathematician John Horton Conway invents his famous CA model, Game of Life, popularized by Martin Gardner in *Scientific American*.

1970s–1980s Psychologist Bibb Latané formulates his theory of social impact, a milestone in social CA modeling.

1971 Schelling publishes his seminal paper on a CA of racial segregation by migration in the *Journal of Mathematical Sociology*.

1975 Economist Peter S. Albin [1934–2008] approaches checkerboard models as CA in his seminal book *Analysis of Complex Socioeconomic Systems*.

1977 Political scientist Stuart A. Bremer [1943–2002] pioneers CA modeling in political science with a hexagon-based simulation of war and peace in the international system, "Machiavelli in Machina," published in Karl W. Deutsch's seminal *Problems in World Modeling*.

1978 Mathematicians J.M. Greenberg and S.P. Hastings develop a true cellular automaton model of excitable media as a 3-state 2-dimensional CA, published in the *SIAM Journal of Applied Mathematics*.

ca. 1981 Physicist Stephen Wolfram begins work on elementary CA theory and modeling, publishing his first paper two years later in *Reviews of Modern Physics*, and later proposing a general classification of CA models in four major classes.

1987 Computer scientist James (Jim) E. Doran publishes his seminal agent-based modeling paper "Distributed Artificial Intelligence and the Modelling of Socio-Cultural Systems."

1987 Mathematician and theologian Edwin A. Abbott publishes his famous mathematical fiction book, *Flatland*, inspiring German computational social scientists Rainer Hegselmann and Andreas Flache to write their 1998 seminal paper, "Understanding Complex Social Dynamics: A Plea For Cellular Automata Based Modelling," in the first volume of the *Journal of Artificial Societies and Social Simulation*.

1990 Political scientists Thomas R. Cusack and Richard J. Stoll publish the realpolitik CA hex-based model of inter- and intra-national conflict, building on S. A. Bremer's earlier work.

1994 Computational social scientist Nigel Gilbert and computer scientist James Doran publish one of the earliest collections of papers on computational applications in social science, *Simulating Societies*, including chapters by other pioneers such as Rosaria Conte, Klaus Troitzsch, Francois Bousquet, Robert Reynolds, Helder Coelho, and Cristiano Castelfranchi.

1995 Computational social scientists Rosaria Conte and Cristiano Castelfranchi publish their seminal work on *Cognitive and Social Action*.

1996 Computational social scientists Joshua Epstein and Robert Axtell publish their influential book on the Sugarscape model, *Growing Artificial Societies*.

1996 Rainer Hegselmann publishes his two influential papers, "Cellular Automata in the Social Sciences" and "Understanding Social Dynamics," still considered among the best introductions to CA simulation models in the social sciences.

1997 Computational social geographer Lena Sanders and her team in Paris publish a seminal paper on SIMPOP, one of the earliest ABM systems for modeling historical urban growth, in the journal *Environment and Planning B: Planning and Design*.

1997 Computational social scientist Robert Axelrod publishes his seminal book on social agent-based modeling, *The Complexity of Cooperation*, as well as his influential paper, "Advancing the Art of Simulation in the Social Sciences," in the journal *Complexity* published by the Santa Fe Institute.

1997 Leigh Tesfatsion at Iowa State University publishes the first newsletter of ACE, Agent-based Computational Economics, which rapidly becomes a major resource for the CSS community.

1998 The *Journal of Artificial Societies and Social Simulation* is founded by computational social scientist Nigel Gilbert, quickly becoming one of the most influential CSS journals. Rainer Hegselmann and Andreas Flache publish their influential paper on CA, and the same year computational social scientist Domenico Parisi publishes the first CA model of ancient Mesopotamian empires, collaborating with historian Mario Liverani.

1999 Computational sociologist Kathleen M. Carley of Carnegie Mellon University and computer scientist Les Gasser of the University of Illinois at Urbana-Champaign publish their seminal paper on "Computational Organization Theory" in G. Weiss's influential *Multiagent Systems* textbook reader.

1999 Nigel Gilbert and Klaus Troitzsch publish the first edition of the classic textbook, *Simulation for the Social Scientist*.

1999 Chris Langton of the Santa Fe Institute establishes the Swarm Development Group for developing the eponymous ABM simulation system that later inspired NetLogo (designed by Uri Wilensky of Northwestern University the same year), Repast (since 2002), and MASON (2002).

1999 Computational archaeologists Timothy Kohler and George Gummerman from the Santa Fe Institute co-edit the influential volume *Dynamics in Human and Primate Societies*, including the so-called Anasazi model.

2002 Stephen Wolfram publishes *A New Kind of Science*, his *magnum opus* in 1280 pages.

2002 The US National Academy of Sciences holds its first Sackler Colloquium and publishes its first *Proceedings* dedicated to social ABM, co-edited by renowned geographer and NAS member Brian L. Berry, L. Douglas Kiel, and Euel Elliott.

2002 The North American Association for Computational Social and Organizational Sciences (NAACSOS) is founded at its first annual meeting and Kathleen Carley becomes its first President. Co-founders include Claudio Cioffi-Revilla (4th president), Charles Macal, Michael North, and David Sallach (2nd president).

2002 The first semester-long courses in CA and ABM are taught in George Mason University's Program in Computational Social Science by an initial faculty consisting of Claudio Cioffi-Revilla (founding chairman, CSS Department), Dawn C. Parker, Robert Axtell, Jacquie Barker, and Timothy Gulden.

2003 Computer scientist Sean Luke and Claudio Cioffi-Revilla release the first version of the MASON (Multi-Agent Simulator of Networks or Neighborhoods) system at the Agent 2003 annual conference in Chicago, demonstrating the new system with the Wetlands ABM and a suite of other classic models (HeatBugs, Conway's Life, Flockers, and Boids).

2004 Andrew Ilachinski of the Center for Naval Analysis publishes *Artificial War*, the largest multi-agent analysis of conflict thus far.

2005 Thomas Schelling of the University of Maryland and former president of the International Studies Association is awarded the Nobel Memorial Prize in Economic Sciences, with Robert Aumann, for his work on conflict theory and social simulations. He is the first computational social scientist to win such an honor.

2005 The first US National Science Foundation grant for a large-scale ABM-GIS simulation model of coupled socio-natural systems using remote sensing and ethnographic methods from field research is awarded to the Mason-Smithsonian Joint Project on Inner Asia, led by Claudio Cioffi-Revilla (principal investigator), Sean Luke, and J. Daniel Rogers.

2006 The first issue of the *Journal of Cellular Automata* is published, with the goal of disseminating "high-quality papers where cellular automata are studied theoretically or used as computational models of mathematical, physical, chemical, biological, social and engineering systems."

2010 Computer scientist Andrew I. Adamatzky from the University of the West of England in Bristol publishes the edited volume *Game of Life Cellular Automata*. The same year Alfons G. Hoekstra, Jiri Kroc and Peter M.A. Stout publish the edited volume entitled *Simulating Complex Systems by Cellular Automata*. Both books demonstrate the scientific maturation of Conway's seminal model.

2010 Claudio Cioffi-Revilla is elected first president of the Computational Social Science Society of the Americas (CSSSA), founded as the successor to NAACSOS.

2010 Princeton University Press publishes Michael Laver and Ernest Sergenti's *Party Competition: An Agent-Based Model*, the first major significant advance in the computational political science of multi-party systems for modeling democratic regimes.

10.3 Cellular Automata Models

This section introduces the superclass of social simulations based on cellular automata (CA) models, used in social science spatial applications, and examines their unique characteristics for understanding emergent social complexity. CA models are presented within the broader context of object-oriented models, which includes an even larger class of computational spatial and organizational models. The emphasis of CA is on neighboring cell-like sites interacting in discrete time steps that resemble a broad variety of social phenomena. Formal aspects involving interaction topologies and behavioral rules are important.

We begin with the following definition:

Definition 10.1 (Cellular Automaton Model) A cellular automaton (CA) simulation is an object-oriented *computational* model for analyzing complex systems consisting of neighboring entities (x, y), called cells, that change their state s_{xy} as they interact in a (typically two-dimensional) grid-like landscape L using some rule set $\mathbb{R}$.

The following are examples of CA social simulation models:

- Sakoda's Group Attitudinal Model
- Schelling's Urban Racial Segregation Model
- Conway's Game of Life
- Hegselman's Opinion Dynamics Model
- Bremer-Mihalka's and Cusack-Stoll's Realpolitik Models
- Axelrod's Tribute Model
- Parisi's Model of the Neo-Assyrian Empire

While we cannot examine all of them in detail, we use these examples to explain basic features of CA social simulations.

Formally, a CA model consists of an array of cells, each of which is in one of a finite number of states. Neighboring cells are defined with respect to a given cell. The dynamic behavior of a CA begins at $t = 0$ when each cell is initialized in a given

Fig. 10.1 Major pioneers of cellular automata models: John von Neumann, inventor of cellular automata (*upper left*); John Horton Conway, inventor of the CA-based Game of Life (*upper right*); Stuart A. Bremer, pioneer computational political scientist in the use of CA models of international conflict (*lower left*); Nobel prize winner Thomas C. Schelling, famous for his model of racial segregation (*lower right*)

state. Given a cell in an initial state s_0, the state at the next step $t + 1$ is determined by rules specified by some mathematical function(s) that determines s_{t+1} based on information concerning one or more neighboring cells. Rules are *local*, in the sense that they affect cells, not the *global* landscape where emergent behavior may occur.

In the simplest CA models all cells are the same and rule sets are homogenous and constant for all cells. **Stochastic cellular automata** and **asynchronous cellular automata** are different from simple CA models and use non-deterministic and other rule sets. As suggested by this distinction, CA models can be purely deterministic or contain stochastic elements defined by probability distributions.

A complete CA social simulation model consists of all elements in Definition 10.1. Accordingly, these models are appropriate for rendering the following formal features of a referent social system:

Discreteness: Spatio-temporal discreteness means that a landscape is divided into cells and time passes in integer units.

Locality: Cells interact only with contiguous neighbors, not with other cells far away.

Interaction topology: Square cells may interact with their north-south-east-west neighbors (called a 4-cell **von Neumann neighborhood**) or with corner neighbors (8-cell **Moore neighborhood**).

Scheduled updating: All cells update their state after each time step according to simple rules, resulting in emergent patterns at the macroscopic, global level of the entire landscape.

CA models in social science date to the first pioneering applications to the study of racial segregation and opinion dynamics, followed by models of territorial growth. These models were initially called "checkerboard" and "chicken wire" models, in reference to square and hexagonal cells, respectively. They are also widely used in fields closely related to CSS, such as ecology. Figure 10.2 illustrates racial segregation and territorial growth models, running from initialization at $t = 0$ to long-run conditions at some t_N.

10.3.1 Motivation: Research Questions

CA models address research questions in many domains of CSS. They are most appropriate for modeling referent systems with the following features, assuming unit cells are simple in terms of attributes and rules, as explained earlier:

1. A *landscape*, physical or conceptual, well describes the referent system. Examples include urban areas, belief systems, and networks of actors ranging from small groups of individuals to the international system of nations.
2. *Actors* located on the landscape have *information* about neighboring actors and use it to update their own state.
3. The *state* of each actor is determined by *rules* that govern behavior conditional on information concerning self and relevant neighbors.
4. At the macroscopic system level the landscape of cells might evolve toward some *stationary state*, *oscillate* between different patterns, or show *chaotic* behavior.
5. *Emergent* properties of social complexity at the systemic level result from interactions at the level of individual cells—the phenomenon known as *emergence*.

Research questions commonly addressed by CA social simulations typically include one or more of the following:

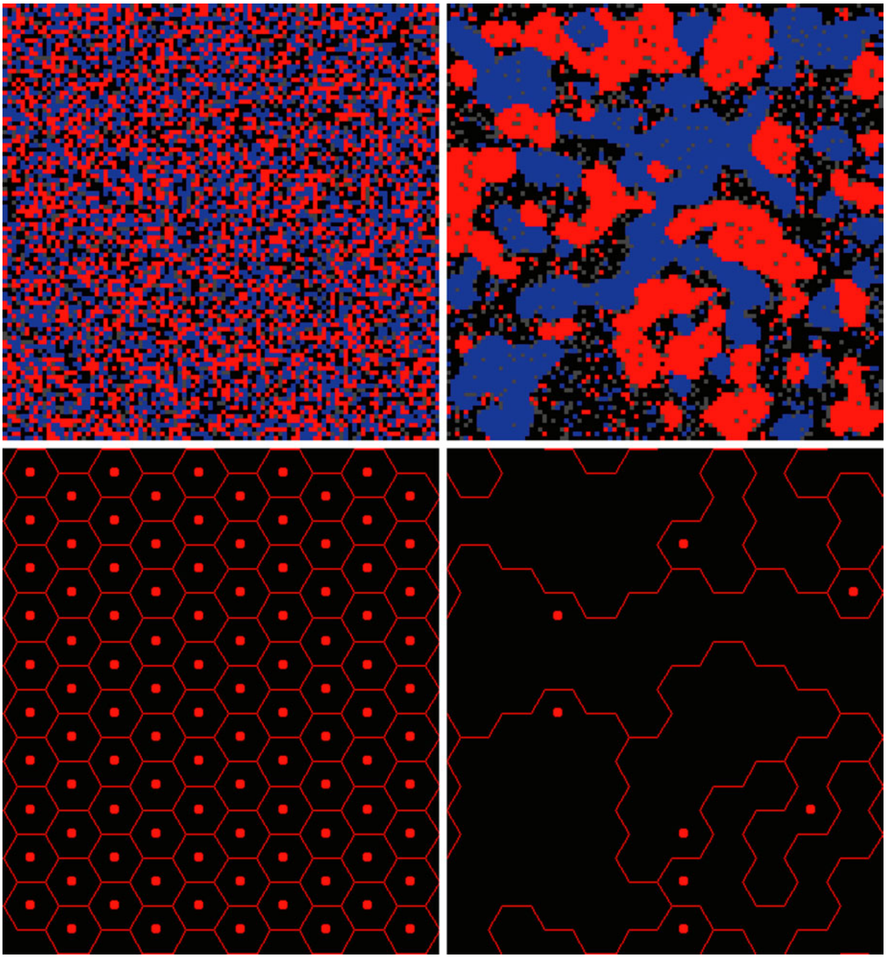


Fig. 10.2 Examples of cellular automata models: The Schelling model with square cells and Moore neighborhood is initialized with ethnically mixed population (*upper left*). Racial segregation emerges as neighbors become cognizant of their surroundings and decide to move away from where they started (*upper right*). The Interhex model with hexagonal cells representing small, simple polities begins with uniformly distributed capabilities (*lower left*). As neighboring polities interact through normal balance of power dynamics, mild stochasticity is sufficient to grow a system of countries. Both models shown in this figure were implemented in MASON, discussed in Sect. 10.3.3

- What is the effect of local cell-level rules on emergent social phenomena?
- Do different interaction topologies (e.g., von Neumann or Moore neighborhoods) matter significantly?
- Are emergent patterns stationary, fluctuating, or chaotic?
- If stationary or fluctuating, what determines the time period for convergence or periodicity of fluctuations?

- Are there patterns of diffusion across the landscape and, if so, how are they characterized?

CA models provide answers to questions such as these through simulation, as long as cell attributes and rules are kept relatively simple, as in the examples provided below.

10.3.2 Design: Abstracting Conceptual and Formal Models

Given some referent system of interest S, a conceptual model C_S, consisting of a cellular automaton and its respective cells, topology, and rule set, is abstracted by a three-stage process consisting of landscape tessellation, interaction topology, and behavioral rules.

Thinking one step ahead, in the case of CA models there are no major design or abstraction considerations that have significant consequences for implementation. All CA models discussed in this chapter and most others in the extant literature run fast on basic laptops. (By contrast, implementation in agent-based models can be highly affected by design/abstraction decisions.) Hence, virtually all CA models are considered "lightweight," computationally speaking. Even when they are large, CA models are easy to distribute due to the total absence of global or long-range interactions.

10.3.2.1 Cellular Tessellation

The first stage in CA abstraction to produce a conceptual model will focus on the referent system's landscape, which should consist of actors represented by *cells*.

Definition 10.2 (Cell) A cell is a tile-like object defined by attributes and located adjacent to other, similar objects. The state of a cell is given by its attribute values, where one or more attribute is a function of the state of neighbors.

The procedure of abstracting cells is called **tessellation**. Cells are the basic elements of a CA model. They can be *square* (most common form), *triangular*, *hexagonal*, or *irregular*, depending on a landscape's tessellation and features of the referent system. Square cells make sense for urban models, whereas hexagonal cells are sometimes preferable for large territories or open terrain. From a computational perspective each has advantages and disadvantages, depending on multiple factors such as number of cells, movement, and scheduling.

For example, in **Conway's Game of Life** cells are square in the classic version, defining a rectangular landscape. In other versions cells can also be hexagonal. Regardless of form, each cell can be in one of two states, alive or dead. What happens to each cell and the whole population in the simulation depends on the condition of neighboring cells in the landscape.

As another example, in **Schelling's Segregation Model** (Fig. 10.2, upper frames) each cell represents a person with a given level of racial tolerance (attribute). Each person is happy or unhappy (the cell's two states) depending on the race of neighbors, which, in turn, will determine whether the person moves away from his/her present neighborhood.

Urban sprawl is a more complex example of a CA-like social phenomenon. Each area surrounding a city may become suburbanized or not, depending on factors (attributes) such as population growth, cost of land, proximity to work, and other variables considered by actors who may decide to move away from a downtown urban center to a suburban neighborhood.

Before the advent of airplanes, when military conquest was mostly land-driven, *territorial polities* grew and contracted based on the ability of a population center to expand its territory into increasingly large swaths of neighboring territories. Hexagonal cells—such as those in the Interhex model, Fig. 10.2—are good tessellations for open territory, as demonstrated by tabletop games played by the military since the German army (Prussian General Staff) pioneered war games in the early 19th century. However, square cells are also used for modeling polity expansion, as demonstrated by Domenico Parisi in his study of the growth of the Neo-Assyrian Empire during the 9th–7th centuries BC using a CA model.

A distinctive feature of cells in CA models is that the *number of attributes* they contain is *relatively small*. (By contrast, agent-based models examined in the next section commonly encapsulate numerous attributes, sometimes in the hundreds, as well as complex methods for updating attribute values.) In the previous examples each cell has just one or a few attributes, such as being alive or dead in the Game of Life, or happy or unhappy in Schelling's segregation model.

The *size* of a CA landscape in terms of number of cells also matters, since larger numbers can often generate emergent phenomena not possible with smaller worlds. Size is determined by tessellation.

10.3.2.2 Interaction Topology

The second stage of abstraction in developing a CA model consists of specifying the **interaction topology**—how cells are "wired" to neighboring cells, so to speak. Interaction topology defines an array of local, short-range interactions. This step comes second, because it depends in part on the form of cells. Square cells can have either von Neumann or Moore neighborhoods, as already mentioned. Hexagonal cells commonly have six neighbors, although they can also have three by alternating neighbors. Triangular cells can have the equivalent of von Neumann and Moore neighborhoods, depending on whether they have three side neighbors or all six, including apical neighbors (sometimes referred to somewhat imprecisely as "corner neighbors").

Another defining feature of interaction topology is **neighborhood radius**, defined as distance from a cell to its farthest neighbor, normally not more than two or three cells away. Most CA models operate with an interaction topology of radius 1 to ensure only local, short-range interactions.

In the Game of Life, interaction topology is defined by a Moore neighborhood of radius 1, thus including all eight surrounding cells, as is also the case for the Schelling segregation model. CA models of other referent systems can assume different interaction topologies, such as when triangular or hexagonal cells are used to represent a landscape. (Compare square cells to hexagonal cells in Fig. 10.2.) In the interaction topology of the Bremer-Mihalka and Cusack-Stoll inter-state CA systems of hexagons, all six neighbors affect a cell (country or province). This is also typically the case in wargaming (tabletop or computational) simulations.

For some global emergent phenomena in a CA model, details of the interaction topology (cell shapes, neighborhood radius, as examples) may or may not matter. In fact, an interesting research question to analyze is the *sensitivity of results with respect to interaction topology*, a topic to which we shall return later.

10.3.2.3 Rules of Cell Behavior

The third and final stage of abstraction in a CA model development effort is to specify **rules** followed by cells. Rules are translated into code when a CA model is implemented. Simple rules are what make a CA interesting in terms of generating unexpected emergent patterns.

In the Game of Life, a cell maintains its current state if it has two dead neighbors. When a cell has three dead neighbors, it too becomes dead. This simple rule generates many different patterns that are unexpected, including "gliders"—collectives of cells that move across the landscape.

In Schelling's segregation model the basic rule is that an agent moves to a different neighborhood when it becomes unhappy. The surprising result is that even when agents have a high level of tolerance for neighbors of different race (i.e., $> 50\,\%$ of different ethnicity among surrounding neighbors), segregated neighborhoods still emerge. In the Interhex model the core rule regards the result of neighboring conflicts and what happens to the territory of the vanquished.

In models of opinion formation, rules specify when an agent changes opinion. Numerous CA models of opinion dynamics show surprising results when seemingly simple rules give rise to divided, uniform, or fluctuating opinion groups.

Other CA spatial models, such as those simulating territorial polities, have simple rules capable of generating complex patterns of land borders.

The main result of the design stage of a CA model is a conceptual and formal model of the referent social system specified by a landscape of cells (specifying their total number and individual geometry), their interaction topology (specifying how cells are wired together in an array), and behavioral rules (specifying what each cell does).

10.3.3 Implementation: Cellular Automata Software

Given a sufficiently complete conceptual or formal model of a referent system as a CA, the next methodological stage consists of implementing the model in code using a simulation system. (As always, the model can also be implemented in native code using an OOP language, such as Python, Java, or C++.) The main milestone in implementation is the transition from CA diagrams and mathematical equations in the conceptual model to code in the simulation model.

Swarm, NetLogo, Repast, and MASON are among the most widely utilized CSS simulation systems that offer CA implementation facilities. Conway's Game of Life and Schelling's Social Segregation have also served as demonstration models for CA social simulations. NetLogo offers several already-built CA models that are easy to use and learn with. In the early 2000's, Repast and MASON used the segregation model among the earliest demos to showcase the new simulations systems. They are still in use today. The choice among these alternative simulation systems for learning purposes largely depends on access and familiarity. NetLogo is often the toolkit of choice for learning a new class of models. For research purposes, the others, especially MASON, assume familiarity with Java.

Figure 10.3 shows a screenshot of a 2-dimensional stochastic CA model running in NetLogo. Simulation systems such as these offer new users several pre-set analytical options. In this case NetLogo makes available several neighborhood topology options, shown by "switches" on the left side of the screen. Screenshots and movies are easy to produce with appropriate software running on a computer's operating system.

In addition to "The Big Four" (Swarm, NetLogo, Repast, and MASON), other software systems are also available for implementing CA social simulation models. Mathematica has powerful CA modeling facilities, and many other systems are included in the Nikolai-Maddey 2009 survey of simulation *Tools of the Trade*.

10.3.4 Verification

Verifying a CA social simulation model involves ascertaining that cells, interaction topology, and behavioral rules are all working in the way they are intended according to the conceptual model. In the case of square cells, verification is simplest and relatively straightforward, including checking to see whether landscape borders are behaving properly (edged or toroidal).[1] Behavioral rules are best verified by detailed tracing of each discrete interaction event within a single simulation step. As always, all general verification procedures examined earlier in Sect. 8.7.4 also apply to CA models, including code walkthrough, profiling, and parameter sweeps.

10.3.5 Validation

Validating a CA social simulation model that has been verified involves two main perspectives. **Structure validity** refers to internal features of the model, including main assumptions concerning relevant cell attributes, interaction topology, and behavioral rules. The following should be considered when testing structure validity in a CA model:

[1] A **toroidal** landscape is one where the borders wrap around, such that the landscape is continuous, without an edge.

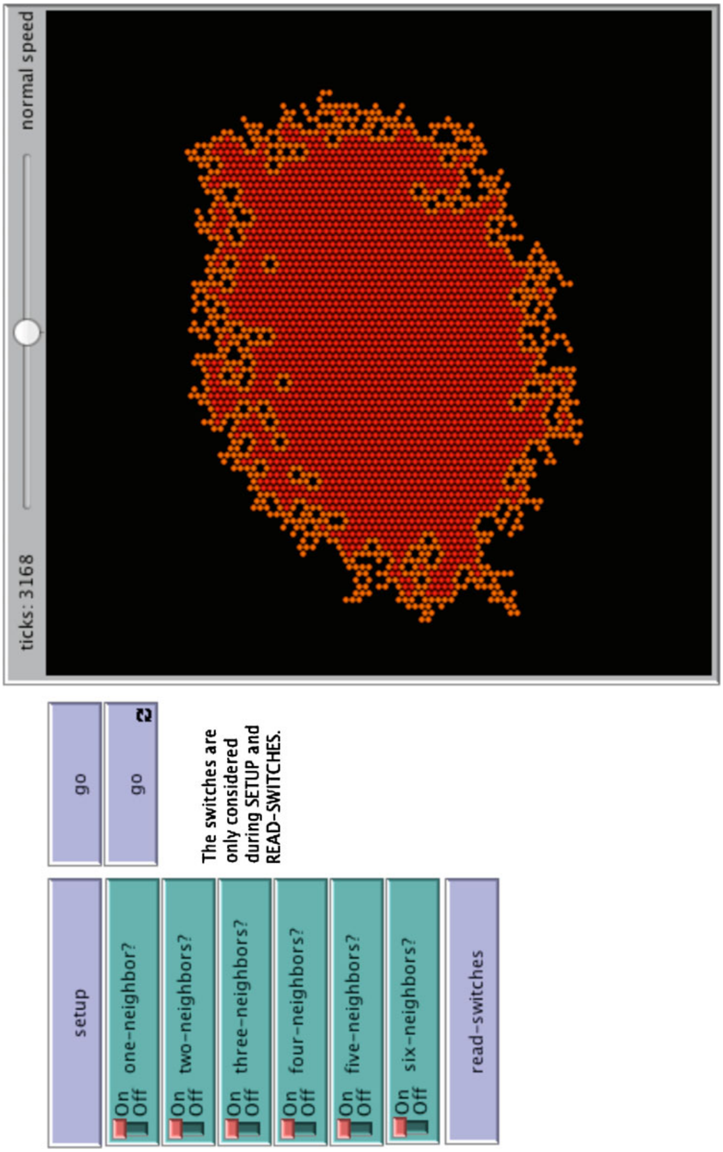


Fig. 10.3 Screenshot of a 2-dimensional cellular automata model of growth with varying number of neighbors running in NetLogo

Empirical tests of validation The specification of equations used in the model, as well as parameter values, are features requiring validation. For example, in the case of Schelling's segregation model discussed earlier, this part of the validation procedure would focus on parameters such as an individual's racial tolerance being assumed, as well as the number of neighbors taken into consideration. The classic model assumes a Moore neighborhood, which is an assumption that requires validation using empirical tests. It is also often assumed that coefficients are constant throughout a given simulated run. These are assumptions of structural stationarity, in the sense that cell rules specified do not change over time; i.e., classical CA models assume that the basic clockwork among cells in a landscape does not change throughout history, which may or may not be a valid assumption about the referent system. For example, education may prevent segregation, or household attention may focus more on neighbors next door rather than across the street or around the block.

Theoretical tests of validation CA model assumptions should also be checked in terms of theories being used, because the simplicity of these models should not distract attention from theoretical underpinnings. Again, this is a broader perspective than empirical tests of structural validity, because it is based on fundamental, causal arguments that are difficult if not impossible to quantify. For example, in the case of the segregation model, the overall structure is based on Schelling's theory of how interaction between two groups is explained. The fundamental theory is based on three factors or dynamics driving the cells' happiness and its decision to stay in the neighborhood or move away: one's own identity; the identity of neighbors; and distance from neighbors. Is this theory valid? Are there other factors as important or even more significant than these? The theory also assumes perfect symmetry among neighbors; i.e., both make residential decisions in the same way. Is it possible that different neighbors decide based on different criteria, such as, one on racial factors and another by education levels?

Tests of structural validity for CA social simulation models can be quite complex and require considerable attention, as seen for other kinds of models. Again, the empirical social science literature is of great value in navigating through these procedures.

Behavior validity is about actual results from simulation runs, especially in terms of qualitative and quantitative features such as cellular landscape patterns of growth, decay, and oscillation, among others. What matters most in the context of ascertaining behavioral validity in CA models is checking whether simulated spatial patterns correspond to empirical patterns.

10.3.6 Analysis

Cellular automata social simulations are analyzed in a variety of ways, including formal analysis, asking what-if questions, and scenario analysis.

Formal analysis of cellular automata, a tradition begun by von Neumann and Ulam, is a field that extends far beyond CSS, but one that provides insights for better understanding social dynamics. For example, Wolfram's classification of CA into a small number of types (stable, oscillating, chaotic, complex) highlights similarities and differences that can be socially meaningful. Formal analysis of rules can also yield theoretical expectations for testing through simulation.

Asking **what-if questions** is another way of analyzing CA social simulations. For example, in a racial segregation model we may ask what happens when tolerance coefficients differ significantly across the two groups. Or, what if tolerance deteriorates as a function of time, as can happen when conflict breaks out in a previously integrated community when previously peaceful but heterogenous neighbors no longer trust each other, as happens in many civil wars. What-if questions can also be used to analyze a CA model using different rule sets. For example, in a racial-migration model we may wish to have one group responding to a Moore neighborhood while another uses a von Neumann neighborhood, based on different attitudes toward physical distance.

Scenario analysis provides a more comprehensive analytical approach to CA simulations by using a set of related questions defining a given scenario, rather than analyzing one question at a time. For example, in a racial-migration model interest may lie in examining a scenario in which tolerance coefficients are relatively large, neighborhood radii are short, and the number of cells is large. Intuitively, such a scenario should not generate segregated neighborhoods. By contrast, an opposite scenario would analyze what happens when tolerance is low, radii are long, and the landscape is smaller. Exploring scenarios between these two extremes can uncover interesting qualitative and quantitative properties, some of which may not be as well-known.

CA models are primarily intended for basic CSS research and theoretical analysis, not for developing actionable policy analysis, given their emphasis on simple interaction rules and overall homogeneity of cells, neighborhoods, and rules. Practical policy analysis can only be obtained through social simulations that allow sufficient empirical specificity and high-fidelity calibration, which is generally not viable with CA—but eminently feasible, if not always easy, with agent-based models.

10.4 Agent-Based Models

This section introduces **agent-based models** (ABM) in CSS, also called social multi-agent systems in computer science. Social ABM simulations are one of the largest and most rapidly growing varieties of computational models. Informally, an ABM can be thought of as a CA with a more sophisticated landscape and actors that come closer to emulating humans through various aspects of reasoning, decision-making, and behaviors.

We begin with the following working definition, which we will later use to examine its main components:

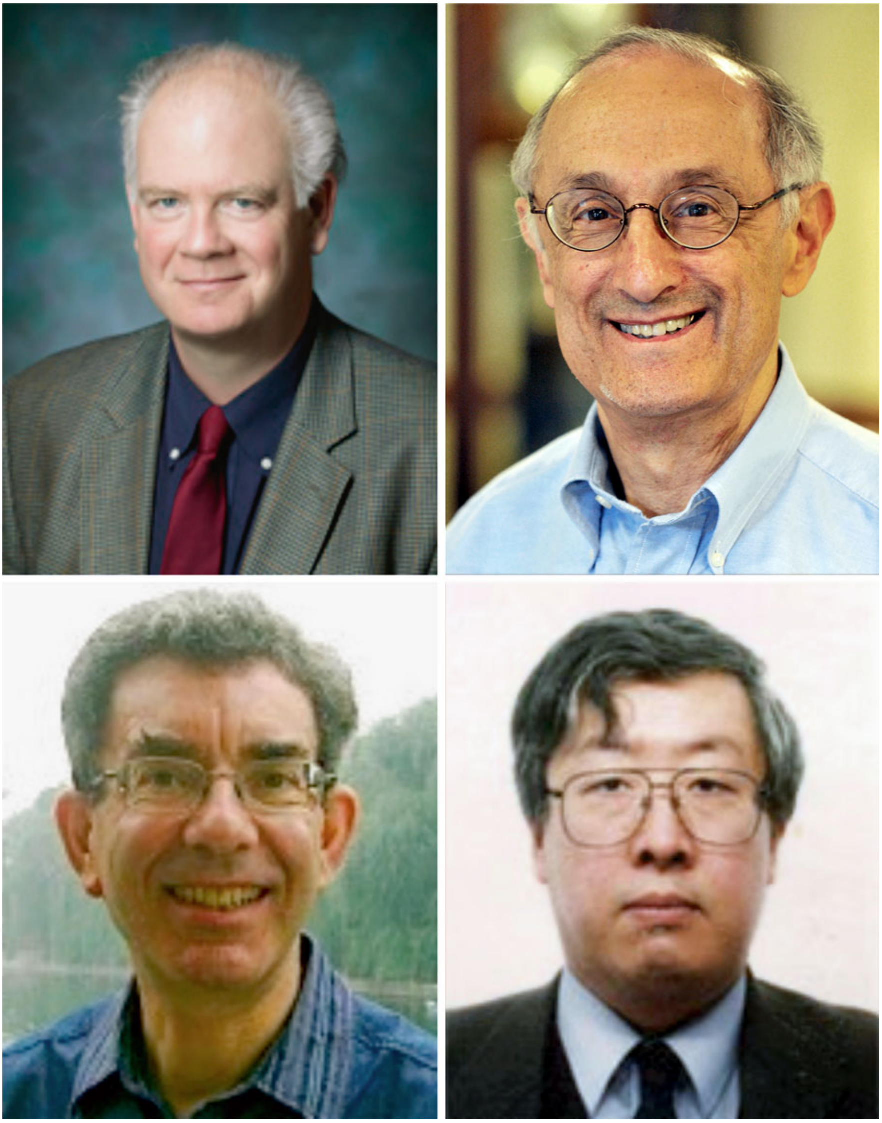


Fig. 10.4 Pioneers of agent-based models. Joshua Epstein, creator of Sugarscape (with R. Axtell) (*upper left*); Robert Axelrod, author of *The Complexity of Cooperation* and other CSS classics (*upper right*); Nigel Gilbert, editor of *Journal of Artificial Societies and Social Simulation* (*lower left*); Hiroshi Deguchi, president of the Pacific-Asian Association for Agent-based Social Science (*lower right*)

Definition 10.3 (Agent-Based Model) A social agent-based model (ABM) is an object-oriented computational model for analyzing a social system consisting of autonomous, interacting, goal-oriented, bounded-rational *set of actors* $\mathbb{A}$ that use a given *rule set* $\mathbb{R}$ and are situated in an *environment* E.

Table 10.1 Examples of agent-based models in CSS by empirical calibration

Model name	Referent system and research questions	Empirical calibration	Source code	Bibliographic reference
RiftLand model	East African coupled socio-techno-natural system; hazards and disaster scenarios	High	MASON	Cioffi-Revilla et al. (2012)
Anasazi	Long House Valley, Arizona; population dynamics and carrying capacity	High	Ascape, NetLogo	Dean et al. (1999), Axtell et al. (2002)
Sugarscape	Theoretical system of agents; social consequences of agent rules	Medium	Ascape, NetLogo	Epstein and Axtell (1996)
RebeLand	Political stability in a country; insurgency and state-failure dynamics	Medium	MASON	Cioffi and Rouleau (2010)
GeoSim	Balance of power system; territorial change	Medium	Repast	Cederman (2003)
FEARLUS	Land-use and cover change; farming dynamics	Medium	Swarm	Gotts and Polhill (2010)
SIMPOP	Urban systems; growth dynamics	Medium	C++	Sanders et al. (1997)
Heatbugs	Abstract social system; agent happiness and social proximity	Low	Swarm	C.G. Langton, Swarm Development Group
Wetlands	Hunter-gatherers affected by weather; social effects of memory	Low	MASON	Cioffi et al. (2004)

Formally, therefore, an ABM consists of the three main components in Definition 10.3: agents, rules, and environments where agents are situated, as we will examine more closely below.

Table 10.1 provides some examples of social ABM models in various domains of CSS. They address a variety of research questions using models calibrated at different empirical levels and built with various simulation toolkits or programming languages (Java and C++). We will draw on some of these examples to explain features of ABM social simulations. Paraphrasing an earlier distinction between a chiefdom and a state, an agent-based model is *not* simply a cellular automaton on hormones—no more so than a jet airliner is a flying bus. The addition of autonomy, goal-directed behavior, and environmental complexity adds entirely new qualitative and quantitative features to a social ABM, compared to the relatively simpler class of cellular automata models.

The *dynamic behavior of an ABM* begins at $t = 0$ when each agent is initialized in a given state. Given an agent in an initial state s_0, the state at the next step $t + 1$ is determined by rules applied to each agent's situation. The next state s_{t+1} will then be based on information processed by rules. Such dynamic behavior is similar but more complex than that of a CA model because now agents have (a) *autonomy* (whereas cells were strongly dependent on their neighborhood), (b) *freedom of movement* (whereas cells had fixed locations), and (c) *reason-based behavior*, among other salient differences. None of these were CA features.

Clearly, agents have more human-like features than cellular automata, making ABMs methodologically appealing and powerful *formalisms for social and behavioral science*. This is especially so in the case of social theories that are expressed primarily in terms of actors, including their cognitive and decision-making processes, and patterns of social behaviors, including collective behavior and organizational and spatial dynamics.

In the simplest ABM models (e.g., Heatbugs, Sugarscape, Boids) all agents are usually the same and rule sets are homogenous and constant for all agents. **Stochastic ABM** and **asynchronous ABM** are different from simple models and use non-deterministic and other rule sets. As suggested by this distinction, ABM models can be purely deterministic or contain stochastic elements defined by probability distributions.

The earliest ABM simulations in social science were Heatbugs (late 1980s), Sugarscape (1996), SIMPOP (1997), and similar spatial "landscape" models that were the first to demonstrate the emergence of social complexity in ways never before seen by social scientists. These pioneer models were followed by many others built during the past decade. ABM simulations are also widely used in ecology and population biology, where they are called *individual-based models*. Figures 10.5 and 10.6 illustrate behavioral patterns and wealth distribution of agents in Sugarscape, running from initialization at $t = 0$ to long-run conditions at some t_N.

10.4.1 Motivation: Research Questions

Agent-based simulation models address research questions in many domains of CSS—whether from basic research or applied policy perspectives. They are most appropriate for modeling referent systems with the following features, where agents can range from "light" cognition and decision-making capacity to "heavy" agents with more detailed cognitive architecture:

Bounded rationality: Agents make decisions under conditions of bounded rationality, as examined earlier in Sect. 7.5.2.

Decision-based behavior: Agents behave based on choices determined by some form of reasoning. This is in contrast to the unreasoned, purely rule-based behavior of cellular automata examined earlier.

Artifacts and artificial systems: When built artifacts such as institutions or infrastructure matter in a referent system, those entities can be represented in an ABM in a number of ways.

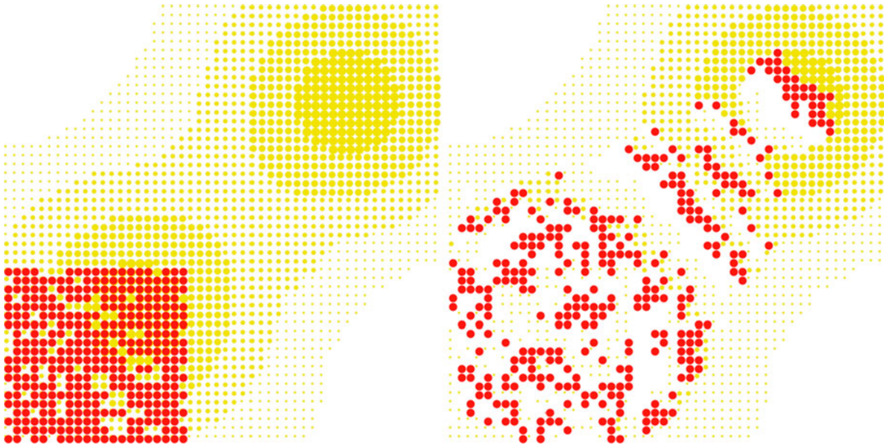


Fig. 10.5 *The Sugarscape agent-based model: agent behavior.* The Sugarscape model consists of a society of agents (*red dots*) situated on a landscape consisting of a grid of square sites where agents with von Neumann neighborhood-vision feed on sugar (*yellow dots*). *Left*: At initialization agents are assigned a uniform distribution of wealth and they reside in the southwestern region. *Right*: After a number of time steps, most agents have migrated away from their original homeland as they move around feeding on the landscape. This MASON implementation by Tony Bigbee also replicates the "wave" phenomenon generated by the original (and now lost) implementation in Ascape, observed here by the northwest-southeast formations of diagonally grouped agents in the northeast region

Social or physical spaces: Referent systems may contain organizational (e.g., social networks), territorial (physical spaces), or other spatial aspects (policy spaces) that are important to model.

Besides these features, ABMs can also have characteristics shared with CA, including various kinds of discreteness, interaction topologies, vision or range, and scheduled updating. All these are ubiquitous and significant features of social complexity that are difficult or impossible to formalize using other modeling approaches (e.g., dynamical systems or game-theoretic models).

Some typical research questions commonly addressed by ABM social simulations may include the following:

- What is the effect of local agent-level rules and micro behaviors on emergent social phenomena at the macro level?
- How do alternative assumptions about human cognition and individual decision-making affect emergent collective behavior?
- Do different interaction topologies (e.g., von Neumann or Moore neighborhoods) or the radius of agents' vision matter significantly?
- Are emergent societal patterns globally stationary, fluctuating, periodic, or chaotic?
- If stationary or fluctuating, what determines the time period for convergence or periodicity of fluctuations?
- Are there patterns of diffusion across the landscape and, if so, how are they characterized?

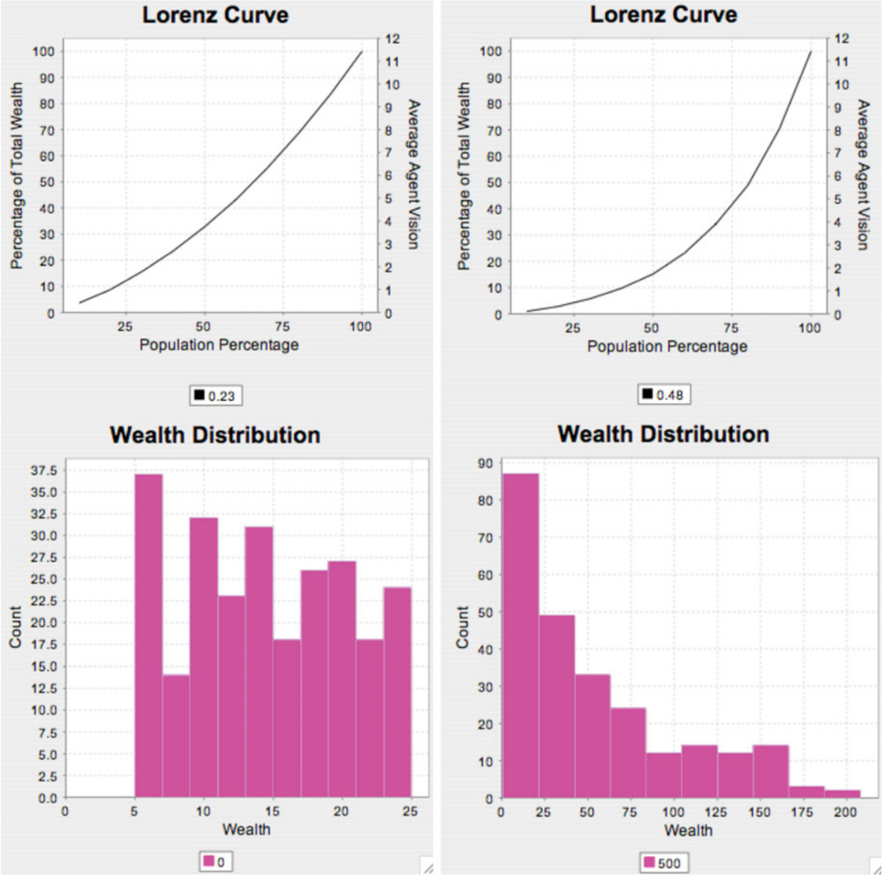


Fig. 10.6 *The Sugarscape agent-based model: emergence of inequality.* Lorenz curves (*top*) and histograms (*bottom*) portray the distribution of agents' wealth. *Left*: Agents are assigned some wealth at initialization $t = 0$, following an approximately uniform distribution, as shown by the nearly straight Lorenz curve and wealth histogram. *Right*: After some time, inequality emerges as a social pattern, as shown by the more pronounced Lorenz curve and much more skewed histogram, similar to Pareto's Law and diagnostic of social complexity

- What is the effect of different distance-dependent functions in human and social dynamics?

Comparing these questions with comparable sets of questions for system dynamics models (Sect. 9.3), queueing models (Sect. 9.4), and cellular automata models (Sect. 10.3), it is clear that these have significantly broader scientific scope as well as analytical depth. Questions addressed by social ABMs also have the feature of being inter-, multi-, or cross-disciplinary, or scientifically integrative, because ABM methodology lends itself to leveraging knowledge across the social, natural, and engineering sciences—which is required for understanding complexity in coupled socio-techno-natural systems. Of all the social simulation methodologies seen thus far, ABMs are arguably among the most versatile in terms of the range of feasible research questions that can be addressed. Research questions in the context of scenario analysis are a major application of ABM social simulations. Asking what-if questions of social complexity is an excellent way to motivate an agent-based simulation.

10.4.2 Design: Abstracting Conceptual and Formal Models

Given some referent system of interest S, a conceptual agent-based model $\mathsf{C_S}$ is abstracted by identifying relevant agents, environments, and rules, as suggested by Definition 10.3.

10.4.2.1 Agents

Human actors in an ABM—whether individuals or collectives (e.g., households, groups, other social aggregates)—are represented as agent-objects that encapsulate attributes and dynamics (computational methods or operations). The **state of an agent** is determined by its attributes, just as in any object.

The following are standard features of agents:

- Each agent is *aware* of its own state, including its environmental situation.
- An agent is said to be *autonomous*, in the sense that it can decide what to do based on endogenous goals and information, much like a social actor, without necessarily requiring exogenous guidance.
- Besides making decisions based on its own internal state, an agent can also decide to act in *reaction* to some perceived environmental situation.
- Moreover, agents can also behave *proactively*, based on goals.
- Agents can *communicate*, sometimes generating emergent patterns of sociality (e.g., collective behavior), by making their attributes visible or actually passing information.

Accordingly, we can use these features to define an agent.

Definition 10.4 (Agent) An agent is an environmentally situated object with encapsulated attributes and methods that enable self-awareness, autonomy, reactivity, proactivity, and communication with other agents and environments. The state of an agent is given by its attribute values.

For example, the agents in Sugarscape satisfy each of these properties: they are aware of being hungry or satisfied; they decide where to move with complete autonomy; they can decide to seek a better patch of sugar, doing so proactively since they seek to survive; and, based on some additional rules, they can communicate and exchange sugar for spice, thereby generating a simple market. Similarly, in the Wetlands model (Table 10.1) agents know their own state: they decide to migrate with autonomy and use memory about various locations; they react to the distribution of other agents and food sites; they communicate among members of their own group, avoiding communication with foreigners. Agents in all models in Table 10.1 share comparable characteristics.

10.4.2.2 Environments

Agents are situated in an environment, which can consist of any number of components related through loose or tight *coupling*. From a complexity-theoretic perspective, natural and artificial systems are assumed to be disjoint components of agents' environment.

- **Natural environments** generally consist of biophysical landscape, sometimes including weather. In turn, landscape can consist of topography, land cover, hydrology, and other biophysical features, depending on what parts of the referent system the model needs to render. Natural environments are governed by biophysical laws, including thermodynamic laws.
- **Artificial environments**—what we may call Simon's environment of artifacts—can include any number of human-built or engineered systems, such as buildings, streets, markets, and parks in urban areas, or roads, bridges, and transportation nodes linking urban areas. *Critical infrastructure systems*, specifically, are comprised of several major components, such as roads, energy, telecommunications, water supply, public health, and sanitation, among others, depending on a country's statutory taxonomy. Artificial environments are also governed by physical laws, except thermodynamics. This is because artificial systems generate more order (decreasing entropy) by using resources, which is the reverse of thermodynamic disorder (increasing entropy).

For example, in terms of ABMs in Table 10.1, Anasazi and Wetlands comprise natural environments, whereas RiftLand, RebeLand, SIMPOP, and FEARLUS also include artificial environments.

10.4.2.3 Rules

Agents and environmental components interact among themselves as well as with each other, generating emergent behavior through the following inter-agent, agent-environment, and intra-environment interactions. Rules are generally *local*, in the sense that they affect agents but not the *global* landscape where emergent behavior may occur—similar to micro-motives generating macro-behavior (paraphrasing T.S. Schelling's famous 1978 book). In turn, however, agents can also be affected by global conditions.

- **Inter-agent rules** govern interactions among agents through communication, exchange, cooperation, conflict, migration, and other patterns of social behavior, including particularly significant patterns such as *collective action* and *social choice*. Generally these rules are grounded in social theory and research. For example, in Wetlands, agents communicate among members of the same group; in RebeLand, government agents and insurgent agents fight each other while general population agents express support for or against government or insurgents.
- **Agent-environment rules** govern effects of environmental conditions on agents and, vice versa, environmental impacts on agents' decisions and behaviors (simulating anthropogenic effects on the environment). These rules are also grounded in social theory, as well as environmental science and related disciplines. For example, in RiftLand farmers are affected by rainfall and land cover, whereas in GeoSim and similar war-games countries are affected by balance of power processes with neighboring rivals.

- **Intra-environmental rules** pertain to cause and effect mechanisms within biophysical components of the environment, such as effects of rainfall on vegetation, or effects of natural hazards on infrastructure. This third type of rule is grounded in the physical, biological, and engineering sciences. For example, in the Wetlands model and others like it, rainfall affects vegetation. In Riftland, herds of animals are also affected. In turn, herd grazing affects ground cover, which can affect infrastructure by causing erosion and making severe precipitation more hazardous during rainy seasons.

In the case of abstracting a referent system as being agent-based (unlike the earlier case of cellular automata), there are significant design or abstraction implications that must be considered in terms of subsequent implementation. Most ABM models discussed in this chapter and most others in the extant literature run fast on basic laptops. But some models cannot, requiring distributed computational resources, either through multiple processors or an actual cluster. An effective balance between high-fidelity and viable computational speed can be difficult to accomplish in the case of models having more than just local interactions.

The landscape of an ABM can also be tessellated, where sites can be *square* (most common form), *triangular*, *hexagonal*, or *irregular* (vector shapes), depending on a landscape's features in the referent system. As mentioned for CA, square cells normally are used for urban landscapes, whereas hexagonal cells are often preferable for large territories or open terrain. Each geometry has computational advantages and disadvantages, depending on factors such as total number of agents, sites, decision-making, behaviors, and scheduling. Needed data structures are also a consideration, such as preferring square sites over hexes when remote sensing imagery (using square pixels) is used in a model.

For square grids, agents may have von Neumann, Moore, or other neighborhood topology. For example, the original Sugarscape used von Neumann neighborhoods, whereas hexagonal neighborhoods in Wetlands and GeoSim use all six neighbors. Interaction or visual radii can also vary, depending on what is being abstracted from the referent system.

The main result of the design stage of an ABM is a conceptual and formal model of the referent social system specified by agents (social actors), their behavioral rules (what each agent does), and an environment (where agents are situated). Class, sequential, and state diagrams in UML are useful for specifying a conceptual model, along with traditional flowcharts. Mathematical models are also helpful in specifying a formal model of the referent system of interest.

10.4.3 Implementation: Agent-Based Simulation Systems

Having developed a sufficiently complete conceptual or formal model of a referent system as an ABM, the next methodological stage consists of implementing the model in code using a simulation system. As always, the model can also be implemented in native code using an OOP language, such as Python, Java, or C++. Currently available simulation systems are mostly Java-based. The main milestone

Fig. 10.7 Pioneers of ABM toolkits. Swarm's Chris Langton (*upper left*); NetLogo's Uri Wilensky (*upper right*); Repast's David Sallach (*lower left*); MASON's Sean Luke (*lower right*). All of them collaborated with others in creating today's leading simulation systems for building social ABMs

in implementation is the transition from UML diagrams and mathematical equations in the conceptual model to code in the simulation model.

The number of agent-based simulation systems (toolkits) today ranges somewhere between fifty and a hundred, with more being created to provide new facilities. Swarm, NetLogo, Repast, and MASON are among the most widely utilized

ABM simulation systems. The choice among these alternative simulation systems for learning purposes largely depends on access and familiarity. As was the case earlier for cellular automata, NetLogo is often the toolkit of choice for learning agent-based modeling, although Python software is becoming increasingly available. For advanced research purposes, Repast and, in particular, MASON assume familiarity with Java. Both Repast and GeoMASON can also implement true GIS for developing spatial ABMs with high-fidelity calibration to represent realistic empirical features of terrain and other features of a referent system.

Figure 10.8 shows a screenshot of the Sugarscape model implemented in NetLogo.

In addition to "The Big Four" (Swarm, NetLogo, Repast, and MASON), other software systems are also available for implementing ABM simulation models. Mathematica has demonstrated several simple ABMs, such as Sugarscape and Boids. Other ABM simulation systems are included in the Nikolai-Maddey 2009 survey.

10.4.4 Verification

Verifying an ABM social simulation model requires making sure that agents, rules, and environments are all working the way they are supposed to according to the conceptual model. In the case of relatively few agents and square cells, verification is simplest and relatively straightforward. Part of verification must include close examination of landscape borders (edged or toroidal). Behavioral rules are best verified by detailed tracing of each discrete interaction event within a single simulation step. As always, all general verification procedures examined earlier in Sect. 8.7.4 also apply to ABM social simulations, including code walk-through, unit testing, profiling, and parameter sweeps.

10.4.5 Validation

Validating an ABM social simulation model that has passed its verification tests involves the same two main perspectives mentioned earlier for other models: structural and behavioral validity.

Structural validity refers to internal features of the model, including main assumptions concerning relevant agent attributes, interaction rules, and environments. The following should be considered when testing structural validity in an ABM:

Empirical tests of validation The specification of equations used by object methods, as well as attribute and parameter values, are features requiring validation. For example, in the case of the Anasazi and Riftland models, this part of the validation procedure focused on parameters such as vegetation grow-back rates, as well as features of weather and land use. The radius of vision or communication used is another assumption requiring validation using empirical tests. It is also often assumed that coefficients are constant throughout

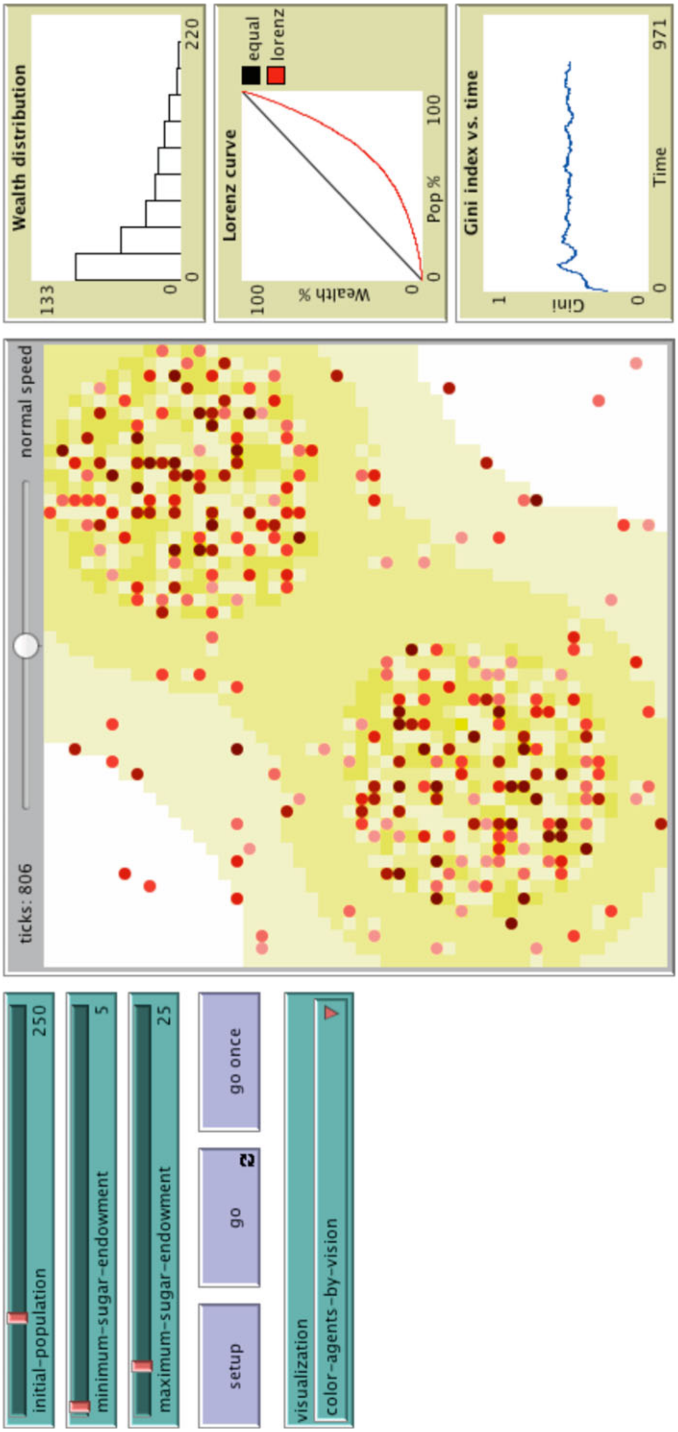


Fig. 10.8 Screenshot of a Sugarscape model implemented in NetLogo

a given simulated run. These are assumptions of structural stationarity, in the sense that agent rules do not change over time; i.e., classical object models assume that the basic clockwork of agents, rules, and environment does not change throughout history, which may or may not be a valid assumption in regards to a given referent system. For example, poverty may impair decision-making, or conflict may reduce cognitive bandwidth and complicate reasoning caused by unresolved dissonance (Sect. 4.8.1).

Theoretical tests of validation ABM simulation assumptions must also be checked in terms of theories being used, especially concerning knowledge taken from various disciplines. This is a broader perspective than empirical tests of structural validity, as already noted, because it is based on fundamental causal arguments that are sometimes difficult—if not impossible—to quantify. For example, in the case of GeoSim and similar models, the overall structure is based on balance of power and deterrence theory concerning how nations are supposed to interact in an international system. In this case, the fundamental theory is based on factors such as objective capabilities untransformed by perceptions, calendar time undistorted by tension and stress, and other simplifying features. Is such a theory valid? Are there other factors as important or even more significant than these? The underlying theory used in an ABM may also assume perfect symmetry among agents, even when they are heterogeneous in some respects. Even bounded rationality is often implemented in simplistic ways. Is it possible that actors decide with time-dependent or other forms of heterogeneity?

Tests of structural validity for ABM social simulation models can be laborious, but are always necessary to develop confidence in a model. Again, the empirical literature is of critical value in conducting these tests.

Behavioral validity is about actual results from ABM simulation runs, especially in terms of qualitative and quantitative features such as patterns of growth, decay, or oscillation. What matters most for ascertaining behavioral validity is whether simulated spatial patterns generated by an ABM correspond to known empirical patterns in its referent system. Time series, histograms, specialized metrics, and similar results are among the most commonly used. For example, Figs. 10.6 and 10.8 showed the Lorenz curves and wealth distribution histograms generated by the Sugarscape model. The long-run patterns of these (shown on the right side of the figure) are a close match to known empirical patterns in many societies (Pareto's Law). The RiftLand model is capable of generating ground cover patterns that are almost indistinguishable from empirical imagery satellite data obtained through remote sensing. The Anasazi model was among the first empirically referenced ABMs to demonstrate a close fit between simulated results and empirically measured patterns.

10.4.6 Analysis

ABM social simulations are susceptible to many forms of analysis, including formal analysis, asking what-if questions, and scenario analysis.

Formal analysis of ABM, a tradition exemplified by urban dynamics and human geography, is a major field extending far beyond the confines of CSS. For example, various *gravity models* of agent interactions, as well as *driven-threshold systems* of agents display significant properties that can be investigated through formal analysis. For the most part, CSS researches have paid relatively little attention to formal analysis of spatio-temporal interactions of agent communities. For example, different distance or temporal interaction structural specifications, and different types of driven-threshold mechanisms remain largely unexplored, in spite of their fundamental theoretical interest. Formal analysis of agent rules can also yield theoretical expectations for testing through simulation.

Another way of analyzing ABM social simulations is by asking **what-if questions**. For example, in a model such as Sugarscape we may ask what may happen when a Moore neighborhood is used, as opposed to the standard von Neumann neighborhood. Or, what if agent vision deteriorates as a function of time, as can happen also in times of conflict ("fog of war" effect). What-if questions can also be used to analyze an ABM simulation using different rule sets. For example, in an agent migration model we may wish to have one group responding to a Moore neighborhood while another uses a von Neumann neighborhood, perhaps based on different attitudes toward physical distance. Or, one group may be endowed with vision having longer range.

Scenario analysis provides a more comprehensive and versatile methodological approach to analyzing ABM social simulations. A scenario uses a set of related research questions, rather than analyzing one question at a time. For example, in a model such as RiftLand, it is possible to investigate a scenario such as prolonged drought in a given country: Given a three-year drought that has been going on in, say, Kenya, what may happen to crops and herds should the drought continue for another year or two? How might social relations be affected? Will governmental institutions of the polity have sufficient capacity to mitigate the societal effects caused by drought? Will there be displaced persons? Will large-scale refugee flows be generated by the drought? Will refugee flows remain internal or cross boundaries into neighboring countries? Can such analyses provide novel insights that may be valuable to relief planners and responders? Sets of scenarios can also be used for investigating natural, engineering, and anthropogenic (human-caused) disasters.

ABM social simulations are still primarily intended for basic CSS and theoretical analysis, but increasingly they are being called upon to address policy analysis to provide actionable results. Significant methodological and theoretical advances are still necessary to satisfy demand, but sustained progress will enable future generations of CSS researchers to build upon and surpass these recent achievements.

Recommended Readings

On Cellular Automata

P.S. Albin, *The Analysis of Complex Socioeconomic Systems* (Lexington Books, Lexington, 1975)

S.A. Bremer, M. Mihalka, Machiavelli in Machina: or politics among hexagons, in *Problems in World Modeling*, ed. by K.W. Deutsch (Ballinger, Boston, 1977)

T.R. Cusack, R.J. Stoll, Adaptation, state survival and system endurance: a simulation study. Int. Polit. Sci. Rev. **11**(2), 261–278 (1990)

R.J. Gaylord, L.J. D'Andria, *Simulating society: a Mathematica toolkit for modelling socioeconomic behavior* (Springer, Berlin, 1998)

T. Hägerstand, A Monte Carlo approach to diffusion. Eur. J. Sociol. **63**, 43–67 (1965)

R. Hegselmann, Cellular automata in the social sciences: perspectives, restrictions and artifacts, in *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, ed. by R. Hegselmann et al.(Kluwer, Dordrecht, 1996a), pp. 209–234

R. Hegselmann, Understanding social dynamics: the cellular automata approach, in *Social Science Microsimulation*, ed. by K.G. Troitzsch et al.(Springer, Berlin, 1996b), pp. 282–306

B. Latane, The psychology of social impact. Am. Psychol. **36**(4), 343–356 (1981)

D. Parisi, A cellular automata model of the expansion of the Assyrian empire, in *Cellular Automata: Research Towards Industry*, ed. by S. Bandini, R. Serra, F.S. Liverani (Springer, London, 1998)

J.D. Rogers, T. Nichols, M. Latek, C. Cioffi-Revilla, Modeling scale and variability in human-environmental interactions in Inner Asia. Ecol. Model. **241**, 5–14 (2012)

T.C. Schelling, Dynamic models of segregation. J. Math. Sociol. **1**(2), 143–186 (1971)

U. Wilensky, *NetLogo* (1999). http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL

U. Wilensky, *NetLogo Hex Cell Aggregation Model* (2007). http://ccl.northwestern.edu/netlogo/models/HexCellAggregation. Center for Connected Learning and Computer-Based Modeling, Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL

S. Wolfram, *A New Kind of Science* (Wolfram Media, Champaign, 2002)

On Agent-Based Models

R. Axelrod, *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration* (Princeton University Press, Princeton, 1997)

M. Batty, *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals* (MIT Press, Cambridge, 2006)

A. Bigbee, C. Cioffi-Revilla, S. Luke, Replication of sugarscape using MASON, in *Agent-Based Approaches in Economic and Social Complex Systems IV: Post-Proceedings of the AESCS International Workshop 2005 (vol. 3)*, ed. by T. Terano, H. Kita, H. Deguchi, K. Kijima (Springer, Tokyo, 2007), pp. 183–190

R.B. Boone, K.A. Galvin, S.B. BurnSilver, P.K. Thornton, D.S. Ojima, J.R. Jawson, Using coupled simulation models to link pastoral decision making and ecosystem services. Ecol. Soc. **16**(2), 6 (2011). online

C. Cioffi-Revilla, Invariance and universality in social agent-based simulations. Proc. Natl. Acad. Sci. USA **99**(14), 7314–7316 (2002)

C. Cioffi-Revilla, S. Luke, D.C. Parker, J.D. Rogers, W.W. Fitzhugh, W. Honeychurch, C. Amartuvshin, Agent-based modeling simulation of social adaptation and long-term change in Inner Asia, in *Advancing Social Simulation: The First World Congress in Social Simulation*, ed. by S. Takahashi, D. Sallach, J. Rouchier (Springer, Tokyo, 2007), pp. 189–200

C. Cioffi-Revilla, J.D. Rogers, M. Latek, The MASON HouseholdsWorld of pastoral nomad societies, in *Simulating Interacting Agents and Social Phenomena: The Second World Congress in Social Simulation*, ed. by K. Takadama, C. Cioffi-Revilla, G. Deffaunt (Springer, Berlin, 2010), pp. 193–204

C. Cioffi-Revilla, M. Rouleau, MASON RebeLand: an agent-based model of politics, environment, and insurgency. Intl. Stud. Rev. **12**(1), 31–46 (2010)

C. Cioffi-Revilla, J.D. Rogers, S. Wilcox, J. Alterman, Computing the steppes: data analysis for agent-based models of polities in Inner Asia, in *Xiongnu Archaeology: Multidisciplinary Perspectives of the First Steppe Empire in Inner Asia*, ed. by U. Brosseder, B. Miller (Bonn University Press, Bonn, 2011a), pp. 97–110

C. Cioffi-Revilla, J.D. Rogers, A.B. Hailegiorgis, Geographic information systems and spatial agent-based model simulations for sustainable development. ACM Trans. Intell. Syst. Technol. **3**(1), 10 (2011b)

C.J.E. Castle, A.T. Crooks, M.J. de Smith, M.F. Goodchild, P.A. Longley, Geocomputational methods and modelling, in *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, ed. by M.J. de Smith, M.F. Goodchild, P.A. Longley (Winchelsea Press, Winchelsea, 2007), pp. 383–450

J.E. Doran, M. Palmer, N. Gilbert, P. Mellars, The EOS project: modelling upper paleolithic social change, in *Simulating Societies*, ed. by N. Gilbert, J.E. Doran (UCL Press, London, 1994), pp. 195–221

J.M. Epstein, R. Axtell, *Growing Artificial Societies: Social Science from the Bottom Up* (MIT Press, Cambridge, 1996)

J. Ferber, *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence* (Addison-Wesley, Reading, 1998)

N. Gilbert (ed.), *Computational Social Science* (Sage, Los Angeles, 2010)

N. Gilbert, R. Conte (eds.), *Artificial Societies: The Computer Simulation of Social Life* (University College Press, London, 1995)

A.J. Heppenstall, A.T. Crooks, L.M. See, M. Batty (eds.), *Agent-Based Models of Geographical Systems* (Springer, Berlin, 2011)

D. Helbing (ed.), *Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior* (Springer, Berlin, 2012)

W.G. Kennedy, C.R. Cotla, T. Gulden, M. Coletti, C. Cioffi-Revilla, Validation of a household agent-based model of the societies of East Africa, in *Proceedings of the 2012 Human, Social, Cultural, and Behavioral Conference*, San Francisco, CA, Proceedings of the Fourth World Congress on Social Simulation, Taipei, Republic of China (2012)

T.A. Kohler (ed.), *Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes* (Addison-Wesley, Reading, 2000)

T.A. Kohler, D. Cockburn, P.L. Hooper, R.K. Bocinsky, Z. Kobti, The coevolution of group size and leadership: an agent-based public goods model for prehispanic Pueblo societies. Adv. Complex Syst. **15**(1–2), 115007 (2012) 115029 pp.

L.A. Kuznar, High-fidelity computational social science in anthropology. Soc. Sci. Comput. Rev. **24**(1), 15–29 (2006)

M. Laver, E. Sergenti, *Party competition: an agent-based model* (Princeton University Press, Princeton, 2012)

S. Luke, Multiagent simulation and the MASON Library (2011). Retrieved from http://cs.gmu.edu/~eclab/projects/mason/

S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, MASON: a Java multi-agent simulation environment. Simul. Trans. Soc. Model. Simul. Int. **81**(7), 517–527 (2005)

M.J. North, C.M. Macal, *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation* (Oxford University Press, London, 2007)

D. Phan, F. Amblard (eds.), *Agent-Based Modelling and Simulation in the Social and Human Sciences* (Bardwell Press, Oxford, 2007)

D.L. Poole, A.K. Macworth, *Artificial Intelligence: Foundations of Computational Agents* (Cambridge University Press, Cambridge, 2010)

S.F. Railsback, V. Grimm, *Agent-Based and Individual-Based Modeling: A Practical Introduction* (Princeton, Princeton University Press, 2012)

J.D. Rogers, T. Nichols, T. Emmerich, M. Latek, C. Cioffi-Revilla, Modeling scale and variability in human-environmental interactions in Inner Asia. Ecol. Model. **241**, 5–14 (2012)

T. Salamon, *Design of Agent-Based Models: Developing Computer Simulations for a Better Understanding of Social Processes* (Eva & Tomas Bruckner Publishing, Repin, 2011)

R. Sun (ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation* (Cambridge University Press, Cambridge, 2006)

K. Takadama, C. Cioffi-Revilla, G. Deffaunt (eds.), *Simulating Interacting Agents and Social Phenomena: The Second World Congress in Social Simulation (vol. 7)* (Springer, Tokyo, 2010)

W. Tang, D.A. Bennett, The explicit representation of context in agent-based models of complex adaptive spatial systems. Ann. Assoc. Am. Geogr. **100**(5), 1128–1155 (2010)

Y. Shohan, K. Leyton-Brown, *Multiagent systems: algorithmic, game-theoretic, and logical foundations* (Cambridge University Press, Cambridge, 2008)

M. Wooldridge, *An Introduction to MultiAgent Systems* (Wiley, New York, 2009)

References

A.V. Aho, J. Larus, Programming languages and computer science, in *Computer Science*, ed. by M. Shaw (National Academies Press, Washington, 2004), pp. 74–78

R.P. Abelson, Modes of resolution of belief dilemmas. J. Confl. Resolut. **3**(4), 343–352 (1959)

R. Albert, H. Jeong, A.L. Barabasi, Diameter of the world-wide web. Nature **401**, 130–131 (1999)

W. Albig, The content of radio programs, 1925–1935. Soc. Forces **16**, 338–349 (1938)

G. Algaze, *Ancient Mesopotamia at the Dawn of Civilization: The Evolution of an Urban Landscape* (University of Chicago Press, Chicago, 2008)

M. Allais, Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'ecole Americaine. Econometrica **21**, 503–546 (1953)

G.W. Allport, *The Use of Personal Documents in Psychological Science* (Social Science Research Council, New York, 1942)

T.W. Anderson, D.A. Darling, A test of goodness of fit. J. Am. Stat. Assoc. **49**, 765–769 (1954)

R.L. Axtell et al., Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. Proc. Natl. Acad. Sci. USA **3**(99), 7275–7279 (2002)

P. Bak, *How Nature Works: The Science of Self-Organized Criticality* (Springer, New York, 1996)

P. Bak, K. Chen, Self-organized criticality. Sci. Am. **264**, 46–53 (1991)

A.L. Baldwin, Personal structure analysis: a statistical method for investigating the single personality. J. Abnorm. Soc. Psychol. **37**, 163–183 (1942)

R.F. Bales, *Interaction Process Analysis* (Addison-Wesley, Reading, 1950)

A.L. Barabasi, *Linked: The New Science of Networks* (Perseus Publishing, Cambridge, 2002)

A.L. Barabasi, R. Albert, Emergence of scaling in random networks. Science **286**, 509–512 (1999)

A. Barrat, M. Weight, On the properties of small-world networks. Eur. Phys. J. B **13**, 547 (2000)

M. Batty, *Cities and Complexity* (MIT Press, Cambridge, 2006)

M. Batty, *The New Science of Cities* (MIT Press, Cambridge, 2013)

A. Bavelas, A mathematical model for small group structures. Human Organ. **7**, 16–30 (1948)

A. Bavelas, Communication in task-oriented groups. J. Acoust. Soc. Am. **22**, 725–730 (1950)

O. Benson, A simple diplomatic game, in *International Politics and Foreign Policy*, ed. by J.N. Rosenau (Free Press, New York, 1961), pp. 504–511

O. Benson, Simulation in international relations and diplomacy, in *Computer Applications in Political Science*, ed. by H. Borko (Prentice-Hall, Englewood Cliffs, 1963), pp. 574–595

B. Berelson, *Content Analysis in Communication Research* (Free Press, New York, 1952)

B. Berelson, P.F. Lazarsfeld, *The Analysis of Communication Content* (University of Chicago Press, Chicago, 1948)

H.R. Bernard, The science in social science. Proc. Natl. Acad. Sci. USA **109**, 20796 (2012)

P.E. Black, Beta structures, in *Dictionary of Algorithms and Data Structures*, ed. by P.E. Black (US National Institute of Standards and Technology, Washington, 2004)

C. Cioffi-Revilla, *Introduction to Computational Social Science*,
Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1,

L.P. Bloomfield, W. Beattie, Computers and policy-making: the CASCON experiment. J. Confl. Resolut. **15**(1), 33–46 (1971)
S.P. Borgatti, M.G. Everett, L.C. Freeman, *UCINET for Windows: Software for Social Network Analysis* (Analytic Technologies, Harvard, 2002)
S.P. Borgatti, M.G. Everett, J.C. Johnson, *Analyzing Social Networks* (Sage, Los Angeles, 2013)
H. Borko (ed.), *Computer Applications in the Behavioral Sciences* (Prentice-Hall, Englewood Cliffs, 1962)
J.P. Boyd, The algebra of group kinship. J. Math. Psychol. **6**, 139 (1969)
R. Breiger, K. Carley, P. Pattison (eds.), *Dynamic Social Network Modeling and Analysis* (National Academies Press, Washington, 2003)
S.A. Bremer, *Simulated Worlds* (Princeton University Press, Princeton, 1977)
S.A. Bremer, M. Mihalka, Machiavelli in Machina, in *Problems in World Modeling*, ed. by K.W. Deutsch (Ballinger, Boston, 1977)
S.A. Bremer (ed.), *The GLOBUS Model: Computer Simulation of Worldwide Political and Economic Developments*, (Westview, Boulder, 1987)
R.G. Busacker, T.L. Saaty, *Finite Graphs and Networks* (MacGraw-Hill, New York, 1965)
K. Carley, J. Pfeffer, J. Reminga, J. Storrick, D. Columbus, ORA User's Guide. Technical Report, CMU-ISR-13-108, School of Computer Science, Institute for Software Research, Carnegie Mellon University (2013)
P.J. Carrington, J. Scott, Introduction, in *The SAGE Handbook of Social Network Analysis*, ed. by J. Scott, P.J. Carrington, (Sage, Los Angeles, 2011), pp. 1–8
D. Cartwright, F. Harary, Structural balance: a generalization of Heider's theory. Psychol. Rev. **63**, 277–292 (1956)
T.W. Casstevens, Flows in politics and government, in *Applied Systems and Cybernetics*, vol. 6, ed. by G.E. Lasker (Pergamon, New York, 1981), pp. 3092–3096
T.W. Casstevens, H.T. Casstevens II, The circulation of elites: a review and critique of a class of models. Am. J. Polit. Sci. **33**, 296–297 (1989)
L.E. Cederman, Modeling the size of wars: from billiard balls to sandpiles. Am. Polit. Sci. Rev. **97**(1), 135–150 (2003)
I.M. Chakravarti, R.G. Laha, J. Roy, *Handbook of Methods of Applied Statistics* (Wiley, New York, 1967)
W.S. Churchill, *The Gathering Storm* (Houghton Mifflin, New York, 1948)
C. Cioffi-Revilla, A canonical theory of origins and development of social complexity. J. Math. Sociol. **29**(2), 133–153 (2005)
C. Cioffi-Revilla, Simplicity and reality in computational modeling of politics. Comput. Math. Organ. Theory **15**(1), 26–46 (2008)
C. Cioffi-Revilla, Advances in social simulation, in *Proceedings of the 9th Conference of the European Social Simulation Association*, ed. by B. Kaminski, G.. Koloch, (Springer, Heidelberg, 2013), pp. 13–23
C. Cioffi-Revilla et al., MASON RiftLand: an agent-based model for analyzing conflict, disasters, and humanitarian crises in East Africa, in *XXII World Congress of the International Political Science Association*, Madrid (2012)
C. Cioffi-Revilla, M.I. Midlarsky, Power laws, scaling, and fractals in the most lethal international and civil wars, in *The Scourge of War: New Extensions on an Old Problem*, ed. by P.F. Diehl (University of Michigan Press, Ann Arbor, 2004), p. 327
C. Cioffi-Revilla, S. Paus, S. Luke, J.L. Olds, J. Thomas, Mnemonic structure and sociality: a computational agent-based simulation model, in *Proceedings of the Agent 2004 Conference on Social Dynamics: Interaction, Reflexivity and Emergence*, ed. by D. Sallach, C. Macal (Argonne National Laboratory and University of Chicago, Chicago, 2004)
C. Cioffi-Revilla, M. Rouleau, MASON RebeLand: an agent-based model of politics, environment, and insurgency. Int. Stud. Rev. **12**, 31–46 (2010)
D. Collier, J. Gerring, *Concepts and Method in Social Science* (Routledge, New York, 2009)
R. Conte, B. Edmonds, M. Scott, R.K. Sawyer, Sociology and social theory in agent based social simulation: a symposium. Comput. Math. Organ. Theory **7**, 183–205 (2001)

J.H. Conway, Game of life, review by James Gardner. Sci. Am. **223**, 120–123 (1970)
L.S. Cordell, The Wetherhill Mesa simulation: a retrospective, in *Simulations in Archaeology*, ed. by J.A. Sabloff (University of New Mexico Press, Albuquerque, 1981)
R.A. Dahl, *Polyarchy: Participation and Opposition* (Yale University Press, New Haven, 1971)
I. de Sola Pool (ed.), *Trends in Content Analysis* (University of Illinois Press, Urbana, 1959)
I. de Sola Pool, M. Kochen, Contacts and influence. Soc. Netw. **1**, 5 (1978)
I. de Solla Price, Networks of scientific papers. Science **149**, 510 (1965)
J.S. Dean et al., Understanding Anasazi culture through agent-based modeling, in *Dynamics in Human and Primate Societies*, ed. by T. Kohler, G.J. Gummerman (Oxford University Press, Oxford, 1999), pp. 179–205
K.W. Deutsch, *The Nerves of Government* (Free Press, New York, 1963)
J.E. Doran, Systems theory, computer simulations and archaeology. World Archaeol. **1**, 289–298 (1970)
T. Earle (ed.), *Chiefdoms: Power, Economy, and Ideology* (Cambridge University Press, Cambridge, 1991)
T. Earle, *How Chiefs Come to Power: The Political Economy in Prehistory* (Stanford University Press, Stanford, 1997)
D. Easton, *The Political System: An Inquiry into the Study of Political Science* (Knopf, New York, 1953)
D. Easton, *A Framework for Political Analysis* (Prentice-Hall, Englewood Cliffs, 1965)
J. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling* (Princeton University Press, Princeton, 2006)
P. Erdos, A. Renyi, On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. **5**, 17–61 (1960)
M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the internet topology. Comput. Commun. Rev. **29**, 251–262 (1999)
D.P. Fan, Symposium, possibilities in computer content analysis. Soc. Sci. Comput. Rev. **15**(4), 349–350 (1997)
R. Florida, T. Gulden, C. Mellander, The rise of the mega-region. Camb. J. Reg. Econ. Soc. **1**, 459–476 (2008)
E. Forsyth, L. Katz, A matrix approach to the analysis of sociometric data: preliminary report. Sociometry **9**, 340–347 (1946)
L.C. Freeman, *The Development of Social Network Analysis: A Study in the Sociology of Science* (Empirical Press, Vancouver, 2004)
L.C. Freeman, The development of social network analysis with an emphasis on recent events, in *The SAGE Handbook of Social Network Analysis*, ed. by P.J. Carrington, J. Scott, (Sage, Los Angeles, 2011), pp. 26–40
J. Gao, K.H. Leetaru, J. Hu, C. Cioffi-Revilla, P. Schrodt, Massive media event data analysis to assess world-wide political conflict and instability, in *Advances in Social Computing: Proceedings of the 2013 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP13)*, ed. by A. Greenberg, W. Kennedy, (Springer, Berlin, 2013), pp. 284–292
N. Gilbert, R. Conte (eds.), *Artificial Societies: The Computer Simulation of Social Life* (University College Press, London, 1995)
N. Gilbert, K. Troitzsch, *Simulation for the social scientist*, 2nd edn. (Open University Press, Buckingham, 2005)
N.M. Gotts, J.G. Polhill, When and how to imitate your neighbours: lessons from and for FEARLUSS. *J. Artif. Soc. Soc. Simul.* **12** (2009). Online
N.M. Gotts, J.G. Polhill, Size matters: large-scale replications of experiments with FEARLUS. Adv. Complex Syst. **13**, 453–467 (2010)
A.F. Griffin, C. Stanish, An agent-based model of settlement patterns and political consolidation in the Lake Titicaca Basin of Peru and Bolivia. *Structure and Dynamics: eJournal of Anthropological and Related Sciences* **2**, 2 (2007)

E. Grimson, J. Guttag, MIT 6.00 Introduction to Computer Science and Programming, Fall 2008. Available online at: http://videolectures.net/mit600f08_intro_compuper_science_programming/

H. Guetzkow, C.F. Alger, R.A. Brody, R.C. Noel, R.C. Snyder (eds.), *Simulation in International Relations: Developments for Research and Teaching* (Prentice-Hall, Englewood Cliffs, 1963)

T.R. Gulden, Spatial and temporal patterns in civil violence. Polit. Life Sci. **21**, 26–37 (2002)

G. Hardin, The tragedy of the commons. Science **162**, 1243–1248 (1968)

D.R. Heise, Measuring attitudes with PET, BYTE. Small Syst. J. **7**(7), 208–246 (1982)

D.R. Heise, Affect control theory: concepts and model. J. Math. Sociol. **13**(1–2), 1–33 (1987)

T. Hermann, A. Hunt, J.G. Neuhoff, *The Sonification Handbook* (Logos Publishing House, Berlin, 2011)

B.M. Hill, A simple general approach to inference about the tail of a distribution. Ann. Stat. **13**, 331–341 (1975)

A.O. Hirschman, *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States* (Harvard University Press, Cambridge, 1970)

J.H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975)

J.H. Holland, Using classifier systems to study adaptive nonlinear networks, in *Lectures in the Sciences of Complexity*, ed. by D.L. Stein (Addison-Wesley, Reading, 1989)

B.B. Hughes, *International Futures: Choices in the Search for a New World Order* (Westview Press, Boulder, 1999)

B.B. Hughes, E.E. Hillebrand, *Exploring and Shaping International Futures* (Paradigm Publishers, Boulder, 2006)

K. Janda, Keyword indexes for the behavioral sciences. Am. Behav. Sci. **7**, 55–58 (1964)

K. Janda, W.H. Tetzlaff, Trial: a computer technique for retrieving information from abstracts of literature. Behav. Sci. **11**, 480–486 (1966)

H.J. Jensen, Self-organized criticality: emergent complex behavior, in *Physical and Biological Systems* (Cambridge University Press, Cambridge, 1998)

G. King, How not to lie with statistics: avoiding common mistakes in quantitative political science. Am. J. Polit. Sci. **30**, 666–687 (1986)

J.W. Kingdon, *Agendas, Alternatives and Public Policies*, 2nd edn. (Harper Collins, New York, 1995)

T.A. Kohler, D. Cockburn, P.L. Hooper, R.K. Bocinsky, Z. Kobti, The coevolution of group size and leadership: an agent-based public goods model for Prehispanic Pueblo societies. Adv. Complex Syst. **15**, 115007 (2012), 29 pp.

T. Kohler, G.J. Gumerman, R.G. Reynolds, Simulating ancient societies. Sci. Am. **293**(1), 77–84 (2005)

T.A. Kohler, S.E. van der Leeuw (eds.), *The Model-Based Archaeology of Socionatural Systems* (School for Advanced Research Press, Santa Fe, 2007)

G. Kramer, *Auditory Display: Sonification, Audification, and Auditory Interfaces* (Addison-Wesley, Reading, 1994)

K. Krippendorf, *Content Analysis*, 2nd edn. (Sage, Los Angeles, 2013)

C.G. Langton (ed.), *Artificial Life III* (Addison-Wesley, Reading, 1994)

H.D. Lasswell, Why be quantitative? in *Language of Politics: Studies in Quantitative Semantics*, ed. by H.D. Lasswell et al., (George W. Stewart, New York, 1949), pp. 40–52

C.A. Lave, J.G. March, *An Introduction to Models in the Social Sciences* (University Press of America, Lanham, 1993)

K. Leetaru, P.A. Schrodt, GDELT: global data on events, location and tone, 1979–2012, in *Proceedings of the Annual Convention of the International Studies Association*, San Francisco (2013)

K. Lewin, *Field Theory in Social Science: Selected Theoretical Papers* (University of Chicago Press, Chicago, 1952)

M.P. Lewis (ed.), *Ethnologue: Languages of the World* (SIL International, Dallas, 2009)

D. Lewis-Williams, D. Pearce, *Inside the Neolithic Mind: Consciousness, Cosmos and the Realm of the Gods* (Thames and Hudson, London, 2009)

M.I. Lichbach, *The Cooperator's Dilemma* (University of Michigan Press, Ann Arbor, 1996)
J.D.C. Little, Little's law as viewed on its 50th anniversary. Oper. Res. **59**(3), 536–549 (2011)
J.C. Loehlin, *Computer Models of Personality* (Random House, New York, 1968)
S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, MASON: a Java multi-agent simulation environment. Trans. Soc. Model. Simul. Int. **81**, 517–527 (2005)
U. Luterbacher, *Dimension Historiques de Modele Dinamique de Conflit: Application aux Processus de Course aux Armements, 1900–1965* (Sijthoff, Leiden, 1974)
U. Luterbacher, Simulation models, global environmental change, and policy, in *International Relations and Global Climate Change*, ed. by U. Luterbacher, D.F. Sprinz (MIT Press, Cambridge, 2001), pp. 183–197
U. Luterbacher, P. Allan, A. Imhoff, SIMPEST: a simulation model of political, economic, and strategic interaction among major powers, in *World Congress of the International Political Science Association*, Moscow, USSR (1979)
U. Luterbacher, M.D. Ward (eds.), *Dynamic Models of International Conflict* (Lynne Rienner Publishers, Boulder, 1985)
J. Marcus, Zapotec chiefdoms and the nature of formative religions, in *Regional Perspectives on the Olmec*, ed. by R.J. Sharer, D.C. Grove (Cambridge University Press, Cambridge, 1989), pp. 148–197
J. Marcus, K.V. Flannery, Zapotec civilization: how urban society evolved, in *Mexico's Oaxaca Valley*, (Thames and Hudson, London, 1996)
A. Marin, B. Wellman, Social network analysis: an introduction, in *The SAGE Handbook of Social Network Analysis*, ed. by P.J. Carrington, J. Scott, (Sage, Los Angeles, 2011), pp. 11–25
A.A. Markov, An example of statistical investigation of the text Eugene Onegin concerning the connection of samples in chains. Sci. Context **19**, 591 (2006 [1913])
M.I. Midlarsky, Power, uncertainty, and the onset of international violence. J. Confl. Resolut. **18**(2), 395–431 (1974)
M.I. Midlarsky, Stochastic modeling in political science research, in *Statistical Distributions in Scientific Work*, vol. 4, ed. by C. Taillie, G.P. Patil, B. Baldessari (Reidel, Dordrecht, 1981), pp. 139–140
M.I. Midlarsky, Absence of memory in the nineteenth-century alliance system: perspectives from queuing theory and bivariate probability distributions. Am. J. Polit. Sci. **4**, 762–784 (1983)
M.I. Midlarsky, A hierarchical equilibrium theory of systemic war. Int. Stud. Q. **30**, 77–105 (1986)
M.I. Midlarsky, *The Evolution of Inequality* (Stanford University Press, Stanford, 1998)
E.J. Milner-Gulland, C. Kerven, R. Behnke, I.A. Wright, A. Smailov, A multi-agent system model of pastoralist behaviour in Kazakhstan. Ecol. Complex. **3**, 23–36 (2006)
P. Morasso, V. Tagliasco (eds.), *Human Movement Understanding: From Computational Geometry to Artificial Intelligence* (Elsevier, Amsterdam, 1986)
J.L. Moreno, *Who Shall Survive?* (Nervous and Mental Disease Publishing Company, Washington, 1934)
C. Nikolai, G. Madey, Tools of the trade: a survey of various agent based modeling platforms. J. Artif. Soc. Soc. Simul. **12**, 2 (2009)
M. Olson Jr., *The Logic of Collective Action: Public Goods and the Theory of Goods* (Harvard University Press, Cambridge, 1965)
J. Ortega y Gasset, *Meditaciones del Quijote* (Taurus, Madrid, 1914)
C.E. Osgood, The representation model and relevant research methods, in *Trends in Content Analysis*, ed. by I. de Sola Pool (University of Illinois Press, Urbana, 1959)
C.E. Osgood, G.J. Suci, P.H. Tannenbaum, *The Measurement of Meaning* (University of Illinois Press, Urbana, 1957)
C.E. Osgood, W.H. May, M.S. Miron, *Cross-Cultural Universals of Affective Meaning* (University of Illinois Press, Urbana, 1975)
V. Ostrom, C. Tiebout, R. Warner, The organization of government in metropolitan areas. Am. Polit. Sci. Rev. **55**, 831–842 (1961)
D. Pumain (ed.), *Hierarchy in Natural and Social* (Springer, Berlin, 2006)

E.H.L. Quarantelli, in *What Is a Disaster? New Answers to Old Questions*, ed. by R.W. Perry, E.H.L. Quarantelli, (International Research Committee on Disasters, 2005), pp. 325–395

A.R. Radcliffe-Browne, *Structure and Function in Primitive Society* (Free Press, New York, 1952)

O. Ramsbotham, The analysis of protracted social conflict: a tribute to Edward Azar. Rev. Int. Stud. **31**(1), 109–126 (2005)

A. Rapoport, Contribution to the theory of random and biased nets. Bull. Math. Biophys. **19**, 257–277 (1957)

A. Rapoport, Uses and limitations of mathematical models in social science, in *Symposium on Sociological Theory*, ed. by L. Gross (Harper Row, New York, 1959), pp. 348–372

A. Rapoport, *Mathematical Models in the Social and Behavioral Sciences* (Wiley, New York, 1983)

A. Rapoport, M. Guyer, A taxonomy of 2×2 games. Gen. Syst. **11**, 203–214 (1966)

A. Rapoport, W. Horvath, A study of a large sociogram. Behav. Sci. **6**, 279–291 (1961)

C.W. Reynolds, Flocks, herds, and schools: a distributed behavioral model. Comput. Graph. **21**(4), 25–34 (1987)

L.F. Richardson, Is it possible to prove any general statements about historical fact? Br. J. Sociol. **3**, 77–84 (1952a)

L.F. Richardson, Contiguity and deadly quarrels: the local pacifying influence. J. R. Stat. Soc. A **115**, 219–231 (1952b)

L.F. Richardson, *Statistics of Deadly Quarrels* (Boxwood Press, Pacific Grove, 1960a)

L.F. Richardson, *Arms and Insecurity* (Quadrangle Press, Philadelphia, 1960b)

L.F. Richardson, The problem of contiguity: an appendix to statistics of deadly quarrels. Gen. Syst. Yearb. Soc. Gen. Syst. Res. **6**, 139–187 (1961)

W.H. Riker, *The Theory of Political Coalitions* (Yale University Press, New Haven, 1962)

D.C. Roberts, D.L. Turcotte, Fractality and self-organized criticality of wars. Fractals **6**(4), 351–357 (1998)

G. Robins, P. Pattison, Y. Kalish, D. Lusher, An introduction to exponential random graph (p^*) models for social networks. Soc. Netw. **29**, 173–191 (2007)

G. Robins, P. Pattison, J. Woolcock, Small and other worlds: global network structures from local processes. Am. J. Sociol. **110**(4), 894–936 (2005)

M.S. Rothman (ed.), *Mesopotamia in the Era of State Formation* (School of American Research Press, Santa Fe, 2001)

J.B. Rundle, K.F. Tiampo, W. Klein, J.S.S. Martins, Self-organization in leaky threshold systems: the influence of near-mean field dynamics and its implications for earthquakes, neurobiology, and forecasting. Proc. Natl. Acad. Sci. USA **99**, 2514–2521 (2002)

J.B. Rundle, D.L. Turcotte, W. Kline (eds.), *Reduction and Predictability of Natural Disasters* (Addison-Wesley, Reading, 1996)

P.A. Samuelson, The pure theory of public expenditure. Rev. Econ. Stat. **36**, 387–389 (1954)

L. Sanders, D. Pumain, H. Mathian, SIMPOP: a multiagent system for the study of urbanism. Environ. Plan. B, Plan. Des. **24**, 287–305 (1997)

T. Sandler, *Collective Action: Theory and Applications* (University of Michigan Press, Ann Arbor, 1992)

G. Sartori, Concept misformation in comparative politics. Am. Polit. Sci. Rev. **64**, 1033–1053 (1970)

T.C. Schelling, *Micromotives and Macrobehavior* (Norton, New York, 1978)

P.A. Schrodt, *Precedents, Progress and Prospects in Political Event Data* (Department of Political Science, Pennsylvania State University, University Park, 2012)

P.A. Schrodt, D.J. Gerner, O. Yilmaz, D. Hermreck, The CAMEO (Conflict and Mediation Event Observations) actor coding framework. Paper presented at the Annual Meeting of the American Political Science Association (2005)

C.E. Shannon, W. Weaver, *Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949)

M. Shaw, in *Computer Science: Reflections on the Field, Reflections from the Field*, ed. by M. Shaw (National Academies Press, Washington, 2004), pp. 66–73

H.A. Simon, A formal theory of interaction in social groups. Am. Sociol. Rev. **17**(2), 202–212 (1952)
H.A. Simon, The architecture of complexity. Proc. Am. Philos. Soc. **106**, 467–482 (1965)
H.A. Simon, *The Sciences of the Artificial*, 1st edn. (MIT Press, Cambridge, 1969)
H.A. Simon, *The Sciences of the Artificial*, 2nd edn. (MIT Press, Cambridge, 1981)
H.A. Simon, *The Sciences of the Artificial*, 3rd edn. (MIT Press, Cambridge, 1996)
A.G. Smith, From words to action: exploring the relationship between a group's value references and its likelihood of engaging in terrorism. Stud. Confl. Terror. **27**, 409 (2004)
R. Solomonoff, A. Rapoport, Connectivity of random nets. Bull. Math. Biophys. **13**, 107–117 (1951)
D. Sornette, *Critical Phenomena in Natural Sciences* (Springer, Berlin, 2003)
R. Suleiman, K.G. Troitzsch, G.N. Gilbert (eds.), *Tools and Techniques for Social Science Simulation* (Physica-Verlag, Berlin, 2000)
S.J. Messick, *Computer Simulation of Personality: Frontier of Psychological Theory* (Wiley, New York, 1963)
S. Toulmin, *The Discovery of Time* (Penguin, London, 1967)
United Nations Development Program, Human Development Report (UNDP, New York, 2012)
J. von Neumann, The general and logical theory of automata, in *Cerebral Mechanisms in Behavior*, ed. by A.L. Jeffress (Wiley, New York, 1951), p. 141
J. von Neumann, *The Computer and the Brain* (Yale University Press, New Haven, 1958)
W.A. Wallis, The Poisson distribution and the Supreme Court. J. Am. Stat. Assoc. **31**, 376–380 (1936)
W.D. Wallis, *A Beginner's Guide to Graph Theory* (Birkhauser, Boston, 2000)
D. Waples, B. Berelson, What Voters Were Told (Unpublished ms., University of Chicago, Graduate Library School, Chicago, 1941)
M.D. Ward (ed.), *Theories, Models, and Simulations in International Relations: Essays in Honor of Harold Guetzkow* (Westview Press, Boulder, 1985)
D.J. Watts, A simple model of global cascades on random networks. Proc. Natl. Acad. Sci. USA **99**(9), 5766–5771 (2000)
D.J. Watts, S.H. Strogatz, Collective dynamics of "small world" networks. Nature **393**, 440–442 (1998)
W. Weibull, A statistical distribution function of wide applicability. J. Appl. Mech. **18**, 293–297 (1951)
R.K. White, Black boy: a value analysis. J. Abnorm. Soc. Psychol. **42**, 440–461 (1947)
H.C. White, *An Anatomy of Kinship* (Prentice-Hall, Englewood Cliffs, 1963)
T.J. Wilkinson, J.H. Christiansen, J. Ur, M. Widell, M. Altaweel, Urbanization within a dynamic environment: modeling Bronze Age communities in Upper Mesopotamia. Am. Anthropol. **109**(1), 52–68 (2007)
E.O. Wilson, *Consilience: The Unity of Knowledge* (Knopf, New York, 1998)
R.J. Wilson, *Introduction to Graph Theory*, 3rd edn. (Longman, New York, 1985)
A. Wohlstetter, The delicate balance of terror. Foreign Aff. **37**(1), 211–234 (1959)
A. Wohlstetter, Theory and opposed-systems design. J. Confl. Resolut. **12**(3), 302–331 (1968)
J.L. Woodward, Quantitative newspaper analysis as a technique of opinion research. Soc. Forces **12**, 526 (1934)
H.T. Wright, Recent research on the origin of the state. Annu. Rev. Anthropol. **6**, 379–397 (1977)
Q. Wright, *A Study of War* (University of Chicago Press, Chicago, 1942)
L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. Syst. Man Cybern. **SMC-3**, 28–44 (1973)
W.C. Zinck, *Dynamic Work Simplification* (Krieger, Melbourne, 1971)

Index

A

B

C. Cioffi-Revilla, *Introduction to Computational Social Science*, Texts in Computer Science, DOI 10.1007/978-1-4471-5661-1,

C

D