Chapter 11
Exploring Social and Asocial Agency in Agent-Based Systems

Sabine Thürmel

Abstract Agent-based systems focus on the simulation of complex interactions and relationships of human and/or non-human agents. Social and asocial agency in agent-based systems depends on the conceptual engineering performed by computer scientists and application engineers. The perspective of multidimensional, gradual agency allows both agency (potentiality) and action (actuality) in socio-technical systems to be examined. A conceptual framework is presented which permits the phenomena of complex regulation of behavior and execution control in computer-mediated environments to be characterized. On this basis scenarios in which humans and non-human agents interact can be analyzed. Emergence in such systems may be described. Distributed action in the material reality can be compared to test-bed simulations. It is shown how the exploration of social and asocial agency in virtual environments may profit from work done in machine ethics.

Keywords Distributed agency • Multi-agent systems • Social computing systems

11.1 Introduction

Since the early 1990s computational science and engineering approaches have profited from agent-based models (ABM) and multi-agent systems (MAS) in general. ABM and MAS focus on the simulation of complex interactions and relationships of human and/or non-human agents. Natural scientists apply ABM to the study of complex adaptive systems in biology or many-particle physics. Engineers use MAS to realize distributed problem solving based on bionic or societal metaphors. Swarm intelligence systems (Dorigo, Maniezzo, & Colorni, 1996) or electronic auctioning systems (Woolridge, 2009) are a case in point. In silico experiments have been performed in the humanities too. Academics have applied ABM to study the evolution of norms (Muldoon et al., 2014), languages (Cowley, 2014) and to explore the impact of different social organization models on

S. Thürmel (✉)
Technische Universität München, Munich, Germany
e-mail: sabine@thuermel.de

D. Secchi, M. Neumann (eds.), *Agent-Based Simulation of Organizational Behavior*, DOI 10.1007/978-3-319-18153-0_11

settlement locations in ancient civilizations (Chliaoutakis & Chalkiadakis, 2014). Thus, agent-based simulators are well-established tools for gaining insights into the dynamics of complex systems, experimenting with behavioral variants and designing virtual and hybrid environments. Following Ferber (1999) ABM are regarded as a special variant of MAS employed for the simulation of complex distributed systems. MAS in general may not only be used for simulation but also for distributed problem solving. MAS systems are found in virtual worlds, in scenarios where robots collaborate, and as a novel approach to govern and control distributed processes in the real world.

This paper provides a short overview of the conceptual engineering currently performed when agent-based approaches are used in computational science and sociology. Following Floridi the task of the "philosophy of information" is "conceptual engineering, that is, the art of identifying conceptual problems and of designing, proposing, and evaluating explanatory solutions" (Floridi, 2011, p. 11). In our case, the conceptual problem is how to characterize agency and interagency between humans, robots, and software agents in such a way that all current forms of interplay can be analyzed. Moreover the approach should be flexible in order to allow future technical developments to be included. However, it should be as straightforward as possible. The proposed solution is a multidimensional gradual classification scheme which is presented in this paper.

The multidimensional gradual framework for agency and action enables the engineer, the socio-technograph, and the philosopher to evaluate agent-based behavior in socio-technical systems. Following Ropohl (1999) socio-technical systems are "action systems" where technical agents and humans interact in order to achieve pre-set goals. It is a rationalistic approach which intends to do justice to the broad capabilities of technical actors today. It does not pretend to capture the complexities of human organizations.

Agent-based systems are inspired by nature as well as by human coordination and collaboration. In the natural sciences they allow complex adaptive systems to be modeled as demonstrated in (Scheutz, Madey, & Boyd, 2005). In the humanities they are used to model and experiment with certain aspects of human agency, e.g. behavioral variants in joining in on standing ovations (Muldoon et al., 2014). In engineering they are deployed as a means to an end to provide embedded governance in smart energy grids (Wedde, Lehnhoff, Rehtanz, & Krause, 2008) or distributed health monitoring systems (Nealon & Moreno, 2003). Varieties in agent-based systems are outlined in Sect. 11.2.

The computer is employed both as a multipurpose machine and a unique medium. Since the computer is indifferent to the applications being run on it is an ideal multipurpose machine where formally specified programs may be executed. Section 11.3 is dedicated to describing the specific mediality of computing systems that is the relation between their specifications and runtime instantiations. Their potentiality and actuality is described based on Hubig's two-tiered presentation of technology in general as a medium (2006).

Floridi's "method of levels of abstraction" (2008) lets us focus on agential perspectives. A multidimensional gradual framework for agency and action is

presented in Sect. 11.4. It may be applied to the role-based modeling of socio-technical systems and to the observation and interpretation of scenarios where humans and non-humans interact. Variants of agent-based systems may be studied and the benefits and limits of current approaches may be characterized.

In order to demonstrate that agent-based approaches are a complex tool in itself emergence of novel behavior and the potential for social and asocial behavior are briefly introduced in Sects. 11.5 and 11.6, respectively. This paper does not intend to present them in depth but relies on them to formulate a caveat when in trying to transfer the insights gained in the laboratory to real world scenarios: due to the nondeterministic nature of agent-based systems one must proceed with great care.

11.2 Varieties of Agent-Based Systems

The goal of the agent-oriented programming paradigm is the adequate and intuitive modeling and implementation of complex interactions and relationships. Software agents were introduced by Hewitt's Actor Model (Hewitt, Bishop, & Steiger, 1973). They were presented as encapsulated objects possessing an internal model and capabilities for communication which may be executed in parallel to others. Today a whole variety of definitions for software agents exist (Woolridge, 2009) but all of them include mechanisms to support persistence, autonomy, interactivity, and flexibility. Persistence refers to the fact that agents are permanent objects outliving the processes which created them. They can decide in an autonomous way when and how to pursue their (currently predefined) goals. They may interact both with other software agents and human users. Flexibility may be supported by specific learning strategies thus allowing the agents to adapt to their environment and especially to other agents.

MAS are suited to role-based modeling and simulation in such diverse fields as biology, economics, and sociology, (1) if the information and expertise is distributed in space and time, (2) if the relationships among the entities may be dynamically changed, and (3) if new organizational structures may arise and change over time. From a computer scientist's perspective using MAS may be a way to realize heuristics for NP-complete problems pursuing a distributed problem solving approach.

In the natural sciences agent-based systems are employed in diverse ways: While the classical biosciences intend to better understand life-as-we-know-it, engineering focuses on life-as-it-could-be. In 1987 the notion of artificial life (AL) was created at the first AL conference organized by Chris Langton. Today the field encompasses the in silico simulation of synthetic life (soft AL), multi-robot systems (hard AL), and biochemical systems (wet AL). The origin of life, evolutionary and ecological dynamics, and the social organization of virtual agents and robots are studied. In AL the focus is not so much on the single organism but on the population as a whole. Both digital evolution and self-organization are important research areas. The goal is to simulate life from the bottom up.

In engineering, bionic approaches such as swarm intelligence as well as societal models are adapted in order to implement collaborative approaches to distributed problem solving. Cooperation strategies provide new heuristics for decentralized planning and optimization (Eymann, 2003, p. 100). Both purely reactive and proactive approaches exist. MAS toolboxes provide a basis for crowd simulation as well as electronic market mechanisms. The latter may be deployed to coordinate emergency response services in disaster recovery systems (Jennings, 2010). MAS provide a basis to cyber-physical systems (CPS). While classical computer systems separate physical and virtual worlds, CPS observe their physical environment by sensors, process their information, and influence their environment with actuators while being connected by a communication layer. Agent-based CPS may be found in distributed rescue systems (Jennings, 2010), smart energy grids (Wedde et al., 2008) or distributed health monitoring systems (Nealon & Moreno, 2003). These systems are first simulated and intended to be deployed to control processes in the material world. In the latter case humans may be integrated for clarifying and/or deciding non-formalized conflicts in an ad hoc manner.

The humanities, social and political science, behavioral economics, and studies in law have discovered agent-based modeling too. Social and asocial agency has been studied by performing in silico experiments and comparing the results with real world behavior. ABM may provide a better fit than conventional economic models to model the "herding" among investors. Early-warning systems for the next financial crisis could be built based on ABM (Agents of Change, 2010). These early results may extend to other domains of behavioral economics. The emergence of social norms can be simulated (Muldoon et al., 2014). Even criminal behavior, deliberate misinterpretations of norms or negligence can be studied. Therefore it is hardly surprising that the Leibniz Center for Law at the University of Amsterdam had been looking—although in vain—for a specific Ph.D. candidate in legal engineering: He or she should be capable of developing new policies in tax evasion scenarios. These scenarios were planned to be based on ABM (Leibnizcenter for Law, 2011). Such a "social computing" approach does not only offer to model social behavior, it could also suggest ways to change it. Policies can be inscribed in semiotic or virtual devices. Constellations of inter-agency and distributed agency materialize.

While technographs such as Latour (2005) strive to observe and analyze the interactions without prejudices by "opening up black boxes" this paper advocates making use of computational science and engineering knowledge in order to enhance the understanding of socio-technical environments. If one understands the capabilities of technical agents due to being familiar with their design and their implementation the analysis is grounded in knowledge about their inner workings and not in observations alone. Such a twofold approach takes both the potential and the actuality of computer-mediated artifacts into account.

11.3 Potentiality and Actuality of Agent-Based Approaches

Virtuality in technologically induced contexts is best explained if Hubig's two-tiered presentation of technology in general as a medium (2006) is adopted. He distinguishes between the "potential sphere of the realization of potential ends" and the "actual sphere of realizing possible ends" (Hubig, 2010, p. 4). Applied to agent-based systems—or IT systems in general—it can be stated that their specification corresponds to the "potential sphere of the realization of potential ends" (Hubig, 2010, p. 4) and any run-time instantiation to a corresponding actual sphere. In other words: Due to their nature as computational artifacts, the potential of social computing systems becomes actual in a concrete instantiation. Their inherent potentiality is actualized during run-time. "A technical system constitutes a potentiality that only becomes a reality if and when the system is identified as relevant for agency and is embedded into concrete contexts of action" (Hubig, 2010, p. 3).

Since purely computational artifacts are intangible, i.e. existing in time but not in space, the situation becomes even more challenging: One and the same social computing program can be executed in experimental environments and in real-world interaction spaces. The demonstrator for the coordination of emergency response services may go live and coordinate human and non-human actors in genuine disaster recovery scenarios. With regard to its impact on the physical environment, it possesses a virtual actuality in the test-bed environment and a real actuality when it is employed in real time in order to control processes in the natural world.

In test-bed environments and real-time deployments, the potential of agent-based systems becomes actual. Thus, the "actual sphere of realizing possible ends" (Hubig, 2010, p. 4) can either be an experimental environment composed exclusively of software agents or a system running in real time. When MAS are used for distributed problem solving the overall objective is to automate processes as far as possible. Thus humans are integrated only if need arises, e.g. for solving potential conflicts in an ad hoc manner. Automatic collaborative routines or new practices for ad hoc collaboration are established. Novel, purely virtual or hybrid contexts realizing collective and distributed agency materialize.

In the following, the agency of technology is not considered a "pragmatic fiction" as it was by Rammert (2011). It is perceived as a (functional) abstraction corresponding to a level of abstraction (LoA) as defined by Floridi: A LoA "is a specific set of typed variables, intuitively representable as an interface, which establishes the scope and type of data that will be available as a resource for the generation of information" (Floridi, 2008, p. 320). For a detailed definition, see (Floridi, 2011, p. 46). A LoA presents an interface where the observed behavior—either in virtual actuality or real actuality—may be interpreted. Under a LoA, different observations may result due to the fact that social computing software can be executed in different run-time environments, e.g. in a test-bed in contrast to a real-time environment. Different LoAs correspond to different abstractions of one and the same behavior of computing systems in a certain run-time environment.

Different observations under one and the same LoA are possible if different versions of a program are run. This is the case when software agents are replaced by humans. Conceptual entities may also be interpreted at a chosen LoA. Note that different levels of abstraction may co-exist. Since levels of abstractions correspond to different perspectives, the system designer's LoA may be different from the sociologist's LoA or the legal engineer's LoA of one and the same social computing system. These LoAs are related but not necessarily identical. By choosing a certain LoA a theory commits itself to a certain interpretation of the object types (Floridi, 2008, p. 327) and their instantiations, e.g. the software agent types and their realizations.

In the design phase ideas guiding the modeling phase are often quite vague at first. In due course their concretization results in a conceptual model (Ruß, Müller, & Hesse, 2010) which is then specified as a software system. From the user's or observer's point of view during run-time the more that is known about the conceptual model the better its potential for (distributed) agency can be predicted, and the better the hybrid constellations of (collective) action, emerging at run-time, may be analyzed. Snapshots of technical agents in action may be complemented by a perspective on the system model. The philosophical benefit of this approach does not only lie in a reconstructive approach as intended by Latour (2005) and Rammert (2011) but also in the conceptual engineering of the activity space. Under a LoA for agency and action, activities may be observed as they unfold.

Using a multidimensional gradual agency concept such a LoA may be characterized in more detail. The classification scheme may already be used in the design phase, when a system based on different agent types is to be modeled. Moreover, the system may be analyzed and educated guesses about its future behavior can be made. Both the specifics of distinct systems and their commonalities may be compiled.

11.4 A Multidimensional Gradual Framework for Evaluating Socio-Technical Systems

A multidimensional gradual framework for agency and action is introduced in Thürmel's paper (2012) and expanded in her dissertation (2013). It is a classification scheme developed so that the potential for individual and joint action of technical agents may be characterized in an efficient way. It is based on as few dimensions as possible doing justice to their capabilities but at the same time demonstrating the width of the gap between humans and current technical agents. Moreover, suggestions are made how this gap could be closed in the future.

The multidimensional classification scheme may used to analyze in detail the role-based modeling of socio-technical systems and to the observation and interpretation of scenarios where humans and non-humans interact. In the following a short overview of the classification scheme is given: In order to demonstrate the

potential for agency not only the activity levels of any entities but also their potential for adaptivity, interaction, personification of others, individual action, and conjoint action have to be taken into account.

The potential for individual action in technical agents such as individual software agents or individual robots depends on their activity level that is their potential for self-induced action and their potential for adapting to their environment. A hammer is just a passive tool unable to act. In contrast a (software) bid agent in high frequency bidding system proactively makes its bid without any human intervention and may even learn to adapt its strategy based on trading patterns and information available in the market.

The potential for conjoint action in MAS or multi-robot systems requires the capability for interaction. If plans and strategies are shared and labor is (re)distributed during execution the technical agents and the humans must attribute capabilities to others, treat them as some kind of person. Suitable distinctions must be made: a filter-agent must be treated in totally different way than an avatar if collaboration is to be successful.

In the following the different dimensions are presented in more detail: The activity level permits the characterization of individual behavior depending on the degree of the self-inducible activity potential. It starts with passive entities such as road bumpers or hammers. Reactivity, realized as simple feedback loops or other situated reactions, is the next level. Active entities permit individual selection between alternatives resulting in changes in the behavior. Pro-active ones allow self-reflective individual selection. The next level corresponds to the capability of setting one's own goals and pursuing them. These capabilities depend on an entity-internal system for information processing linking input to output. In the case of humans it equals a cognitive system connecting perception and action. For material artifacts or software agents an artificial "cognitive" system couples (sensor) input with (actuator) output. Based on such a system for (agent-internal) information processing the level of adaptivity may be defined. It characterizes the plasticity of the phenotype, i.e. the ability to change one's observable characteristics including any traits, which may be made visible by a technical procedure, in correspondence to changes in the environment. Models of adaptivity and their corresponding realizations range from totally rigid to simple conditioning up to impressive cognitive agency, i.e. the capability to learn from past experiences and to plan and act accordingly. A wide range of models co-exist allowing one to study and experiment with artificial "cognition in action." This dimension is important to all who define agency as situation-appropriate behavior and who deem the plasticity of the phenotype as an essential assumption of the conception of man. Based on activity levels and on being able to adapt in a "smart" way acting may be discerned from just behaving and the potential for individual action may be defined. A hammer is just a passive tool, a sensor-controlled power drill demonstrates reactive behavior. A robopet may display active behavior. An automatic bid agent or a car on auto-pilot may perform proactive actions.

Conjoint actions depend on interaction and the potential for personification. The potential for interaction, i.e. the coordination by means of communications, is the

basis of most if not all social computing systems and approaches to distributed problem solving. It may range from uncommunicative, to hard-wired cooperation mechanisms, up to ad hoc cooperation.

The potential for the personification of others enables agents to integrate predicted effects of own and other actions. "Personification of non-humans is best understood as a strategy of dealing with the uncertainty about the identity of the other... Personifying other non-humans is a social reality today and a political necessity for the future" (Teubner, 2006, p. 497). Alluding to Dennett's intentional stance (Dennett, 1987) Rammert talks about "as-if intentionality" which humans must attribute to technical agents for goal-oriented interaction (Rammert, 2011, p. 19). I deem it more appropriate to focus on the capability for personification of others. Behavioral patterns may be explained based on the respective ability to perceive another agent as such. Thus a level of abstraction is found which focuses on agency and interaction and not on the ontological statuses of the involved agents. The personification of others lays the foundation for interactive planning, sharing strategies and for adapting actions. The personification of others in technical agents may lead to an interdisciplinary sort of conceptual engineering, Floridi had in mind, when named it "the art of identifying conceptual problems and of designing, proposing, and evaluating explanatory solutions" (Floridi, 2011, p. 11). The designers' and philosopher's task would be to define the concrete form of collaboration, the concrete level of abstraction one is interested in and the computer scientists' and engineers' task would be to realize such collaborative technical agents and their capabilities of personalization. One approach could be to focus on "shared cooperative activities" and "shared agency" in the sense of Bratman (1992, 2014): Mutual responsiveness, commitment to joint activity and commitment to mutual support form the basis of "shared cooperative activity" (Bratman, 1992). Examples include rational, self-governing groups formed in order to realize a joint project. Human groups of that kind display a so-called "modest sociality" (Bratman, 2014) which may be explained based on Bratman's notion of "shared agency" (2009, 2014). Such agency emerges from structures of interconnected planning agency. Practical rationality forms its core. Examples of such shared agency can be found in distributed health monitoring systems (Nealon & Moreno, 2003). These endeavors and similar research intend to support the vision of "smart health," i.e. patients' competent treatment to be offered by collaborating humans, robots, and software agents at any location while constantly monitoring the patients' health. Another example would be the collaboration of humans, robots, and software agents in "smart manufacturing" offering customized products in highly flexible production environments. The ad hoc coordination of activities in these environments would benefit from even a basic understanding of others and their capabilities for social interaction.

Today, this capability for personification is non-existent in most material and software agents. Some agents have more or less crude models of others, e.g. realized as so-called minimal models of the mind. A further qualitative level may be found in great apes which also have the potential for joint intentionality (Call & Tomasello, 2008). This provides the basis for topic-focused group decision

making based on egoistical behavior. Understanding the other as an intentional agent allows even infants to participate in so-called shared actions (Tomasello, 2008). Understanding others as mental actors lays the basis for interacting intentionally and acting collectively (Tomasello, 2008). Currently there is quite a gap between non-human actors and human ones concerning their ability to interact intentionally. This strongly limits the scope of social computing systems when they are used to predict human behavior or if they are intended to engineer and simulate future environments.

The capabilities for individual action and conjoint action may be defined based on activity levels, the potential for adaptivity, interaction, and personification of others possessed by the involved actor(s). One option is the following: In order to stress the communalities between human and non-human agents, an agent counts as capable of acting (instead of just behaving), if the following conditions concerning its ontogenesis hold: "the individual actor [evolves] as a complex, adaptive system (CAS), which is capable of rule based information processing and based on that able to solve problems by way of adaptive behaviour in a dynamic process of constitution and emergence" (Kappelhoff, 2011, p. 320). Based on the actor's capability for joint intentionality respective of understanding the other as an intentional agent or even as a mental actor, the actor may be capable of joint action, shared or collective action. These levels show how the gap between non-human and human actors could eventually be closed.

Constellations of inter-agency and distributed agency in social computing systems or hybrid constellations, where humans, machines, and programs interact, may be described, examined, and analyzed using the classification scheme for agency and action introduced above. These constellations start with purely virtual systems like swarm intelligence systems and fixed instrumental relationships between humans and assistive software agents where certain tasks are delegated to artificial agents. They continue with flexible partnerships between humans and software agents. They range up to loosely coupled complex adaptive systems. The latter may model such diverse problem spaces as predator–prey relationships of natural ecologies, legal engineering scenarios, or disaster recovery systems. Their common ground and their differences may be discovered when the above outlined multidimensional, gradual conceptual framework for agency and action is applied. A subset of these social computing systems, namely those which may form part of the infrastructure of our world, provide a new form of "embedded governance." Their potential and limits may also be analyzed using the multidimensional agency concept.

Since MAS and most multi-robot systems are not centrally controlled but rely on some sort of distributed control based on self-organization where behavior on the meso- or macro-level emerges from the interaction of the individual agents the next section gives a short introduction to emergence in agent-based systems.

11.5 Emergence in Agent-Based Systems

Starting with Anderson's seminal paper "More is Different" (1972) a revival of the discussion on emergence has taken place: "Emergence, largely ignored just thirty years ago, has become one of the liveliest areas of research in both philosophy and science" (Bedau & Humphreys, 2008). In the current literature a wide variety of emergence concepts is discussed. Important distinctions are to be found between diachronic and synchronic emergence and between weak and strong emergence.

"Diachronic emergence is "horizontal" emergence evolved through time in which the structure from which the novel property emerges exists prior to the emergent." (Vintiadis, 2014, 2.ii). Concerning the novelty of a property, a pattern or a phenomenon in agent-based systems one may follow Darley (1994) and define that "true emergent phenomenon is one for which the optimal means of prediction is simulation." Thus emergence as "the arising of novel and coherent structures, patterns and properties during the process of self-organization in complex systems" (Goldstein, 1999, p. 49) seems appropriate when focusing on agent-based behavior in ABM and MAS. Diachronic emergence due to adaptive behavior in agent-based systems may occur on different levels: Adaptivity on system, i.e. macro-level, e.g. of whole organizations, on the meso-level, e.g. of groups of agents, and on the individual agent level. On all these levels interrelated dynamic processes of constitution and emergence may take place.

In simulation based on certain multi-layered models one may find synchronic emergence, too. In contrast to diachronic emergence "in synchronic emergence [...] the higher-level, emergent phenomena are simultaneously present with the lower-level phenomena from which they emerge" (Vintiadis, 2014, 2.ii). One example is:

> a multi-scale agent-based framework to model phenomena at different levels of organization even if the exact dependence and determination relations are not known. Such models provide insights into the inter-level dynamics of complex systems and might help scientists to discover and formulate equation-based models for multi-scale phenomena, which would otherwise be difficult (if not impossible) to detect (Scheutz et al., 2005).

This agent-based approach was employed in a four-level biological model used for the study of the effects of low-level synaptic and neuro-chemical processes on social interactions in bull frogs (Scheutz et al., 2005, p. 3). Such an approach displays not only diachronic emergence but it may also offer snapshots of synchronic emergence.

One may speak of weak emergence in a system if one focuses on the unpredictability or unexpectedness of a systemic property, a pattern or a phenomenon given its components (Vintiadis, 2014, 2.ii). It may be found in swarm intelligence systems or in agent-based systems investigating the emergence of social norms. In biological models of evolution emergence as unpredictability is judged to be a fundamental fact (e.g., Mayr, 2000, p. 403). In ABM of evolution emergence as unpredictability is a by-product of the in silico experiments and as such the validation of ABM is nontrivial. For some authors like Bedau (1997) the main

characteristic of weak emergence is that "though macro-phenomena of complex systems are in principle ontologically and causally reducible to micro-phenomena, their reductive explanation is intractably complex, save by derivation through simulation of the system's microdynamics and external conditions" (Vintiadis, 2014, 2.ii). Thus weak emergence may be compatible with reduction. Therefore it may make sense to complement an ABM with a numerical one focusing on the system's view. Numerical methods based on nonlinear equation systems support the simulation of quantitative aspects of complex, discrete systems (Mainzer, 2007). In contrast, MAS (Woolridge, 2009) permit collective behavior to be modeled based on the local perspectives of individuals. Both approaches may complement each other. They can even be integrated to simulate both numerical, quantitative and qualitative, logical aspects, e.g. within one expressive temporal specification language (Bosse, Sharpanskykh, & Treur, 2008).

In strong emergence novelty means irreducibility and downward causation, i.e. that the emergent properties and laws supervene on their subvenient base. Whereas the so-called British emergentists in the nineteenth century were convinced that many cases of strong emergence exist (McLaughlin, 2008) today many scientists wonder whether examples (apart from consciousness) exist at all (Bedau & Humphreys, 2008).

In scenarios of distributed cognition where humans, software agents, and robots collaborate novel faculties may become manifest over time in a variant of weak emergence. Even the human controlling an avatar in a game may be affected. The relationship between player and avatar is a multi-faceted phenomenon since avatars simultaneously serve as characters in a simulated world, as a tool which extends the player's agency in the game activity and as props which can be used as a part of the player's presentation (Linderoth, 2005). Inspired by Mead (1934) and Blumer (1973) one could assume that the meaning of virtual objects, situations, and relationships result from the process of symbolic interaction and communication and that this participation in a virtual environment forms the virtual identity and influences the self. Mead distinguishes between three levels of role adoption in the process of identity formation: imitating role playing (play), rule-conforming cooperation (game), and universal cooperation and understanding. These levels can also be found in synthetic worlds which provide opportunities for role-playing, rule-governed games, involving cooperation and negotiation. Thus they allow for multiple virtual identities and new experiences in the virtual realm. On the other hand, they provide a basis for agency in virtual worlds offering novel experiences. These systems provoke us to ask questions about traditional categories such as "What kinds of relationships are appropriate to be had with machines?" (Turkle, 2010, p. 30) or more generally, how this technological progress will affect our interpersonal relationships (Turkle, 2011). Abstraction in mathematics does not challenge us in such a way.

11.6 The Potential for Social and Asocial Agency in Agent-Based Modeling

Agent-based approaches are especially suited to modeling and implementing open systems based on dynamically interacting entities pursuing individual potentially conflicting goals, without central control, using sophisticated approaches to communication and cooperation. This is exemplified in the ALADDIN project where a MAS toolbox was developed and employed to realize a demonstrator for the coordination of emergency response services in disaster management systems (Jennings, 2010). Agent-based systems are employed to simulate social norms (see, e.g., Savarimuthu & Cranefield, 2011). Current work focuses mostly on the detection of patterns in the behavior of crowds, like the phenomenon of standing ovations (Muldoon et al., 2014). The question arises under which circumstances the insights gained in the laboratory through social computing systems are transferable to real world scenarios.

In basic crowd simulation systems the pattern found in the simulations may be compared to the pattern found in real-world examples. In applications, where instrumental rationality is the sole basis of goal-oriented behavior such a transference is often possible. The most obvious case is that of agent-based CPS which are first simulated and then deployed to control processes in the material word. Examples include smart energy grids or distributed health monitoring systems. Even criminal behavior, deliberate misinterpretations of norms or negligence can be studied if it is based on bounded rationality.

Current MAS may be especially suited to modeling interacting egoists perceiving others only as social tools. This is due to the fact that current software agents resemble sociopaths rather than caring humans. This conviction is maintained, for example, by Noreen Herzfeld (2013) who cites M. E. Thomas' *Confessions of a Sociopath* (2013): "Remorse is alien to me… I am generally free of entangling and irrational emotions. I am strategic and canny, intelligent and confident, but I also struggle to react appropriately to other people's confusing and emotion-driven social cues."

The emotionality of humans is one indication that not all results gained in the laboratory via social computing systems are transferable to real world scenarios. Human capabilities and those of technical agents may differ widely. Their acts are based on different cognitive systems, different degrees of freedom and only partially overlapping spheres of experience. Current software agents possess at best synthesized emotions. Human drives and needs are (at least currently) alien to them.

Concerning commonalities and fundamental differences in unethical or illegal behavior in investigations into machine ethics and the treatment of artificial agents as legal subjects are very instructive. Books such as *The Law of Robots* (Pagallo, 2013) and *A Legal Theory for Autonomous Artificial Agents* (Chopra & White, 2011) demonstrate this. Chopra and White are convinced that "in principle artificial agents should be able to qualify for independent legal personality, since this is the closest legal analogue to the philosophical conception of a person" (Chopra & White,

2011, p. 182). In their view "artificial agents are more likely to be law-abiding than humans because of their superior capacity to recognize and remember legal rules" (Chopra & White, 2011, p. 166). If they do not abide the law "a realistic threat of punishment can be palpably weighed in the most mechanical of cost-benefit calculations" (Chopra & White, 2011, p. 168). Pagallo perceives the legal personhood of robots and their constitutional rights as an option only being relevant in the long term (Pagallo, 2013, p. 147). However, he discusses at length both human greediness, using robots as criminal accomplices, and artificial greediness. He states that "in certain fields of social interaction, 'intelligence' emerges from the rule of the game rather than individual choices" (Pagallo, 2013, p. 96). Moreover, investigations into potential ethical status of software agents have been undertaken (e.g., Moor, 2006) and propositions to teach "moral machines" to distinguish right from wrong have been developed (e.g., Wallach & Allen, 2008).

In order to clarify the state of the art in software agents' ethics Moor's distinctions between ethical-impact agents, implicit ethical agents, explicit ethical agents and full ethical agents may be employed (Moor, 2006). In social computing the three classes of lesser ethical agents may be found: software agents used as mere tools may have an ethical impact; electronic auctioning systems may be judged implicit ethical agents, if their "internal functions implicitly promote ethical behavior—or at least avoid unethical behavior" (Moor, 2006, p. 19); disaster management systems based on MAS systems (Jennings, 2010) may be exemplary explicit ethical agents if they "represent ethics explicitly, and then operate effectively on the basis of this knowledge" (Moor, 2006, p. 20). It is open to discussion whether any software agent will ever be a fully ethical agent which "can make explicit ethical judgments generally and is competent to reasonably justify them" (Moor, 2006). But the first variants of ethical (machine) behavior, i.e. proto-ethical systems, are already in place.

Analogous to this classification of ethical behavior displayed by software agents, a wide variety of amoral agents could be implemented. They could range from unethical impact agents or implicit unethical agents to explicit unethical agents, e.g. based on virtue ethics. They could be modeled for use in online games. Such games could provide sheer entertainment, edutainment or form part of the currently so popular serious games. The latter "have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement" (Abt, 1970, p. 6).

To conclude, ABM allow one to model a wide variety of asocial behavior. Yet when transferring the insights gained in the laboratory to real world scenarios, one must proceed with great care.

11.7 Conclusions and Future Directions

The proposed conceptual framework for agency and action offers a multidimensional gradual classification scheme for the observation and interpretation of scenarios where humans and non-humans interact. It may be applied to the analysis

of the potential of social computing systems and their virtual and real actualizations. The above-introduced approach may be used both by the software engineer and the philosopher when role-based interaction in socio-technical systems is to be defined and analyzed during execution. Proto-ethical agency in social computing systems may be explored by adapting (Moor, 2006). Profiting from work done by Darwall (2006), the framework could be expanded in order to potentially attribute commitments to diverse socio-technical actors. Shared agency, a "planning theory of acting together" as defined by Bratman (2014), could be investigated in socio-technical contexts where technical elements are not mere tools but interaction partners.

References

Abt, C. (1970). *Serious games*. New York, NY: Viking.

Agents of Change. (2010). Agents of change. *The Economist*. http://www.economist.com/node/16636121/print. Accessed 30 Jan 2015.

Anderson, P. (1972). More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science, 177*, 393–396.

Bedau, M. (1997). Weak emergence. In J. Tomberlin (Ed.), *Philosophical perspectives: Mind, causation and world* (Vol. 11, pp. 375–399). Malden, MA: Blackwell.

Bedau, M., & Humphreys, P. (2008). *Emergence: Contemporary readings in philosophy and science*. Cambridge, MA: MIT Press.

Blumer, H. (1973). Der methodologische Standort des symbolischen Interaktionismus. In Arbeitsgruppe Bielefelder Soziologen (Eds.), *Alltagswissen, Interaktion und gesellschaftliche Wirklichkeit, Bd. 1*. Reinbek bei Hamburg: Rowohlt.

Bosse, T., Sharpanskykh, A., & Treur, J. (2008). Integrating agent models and dynamical systems. In M. Baldoni, T. C. Son, M. van Riemsdijk, & M. Winikoff (Eds.), *Declarative agent languages and technologies V* (pp. 50–68). Heidelberg: Springer.

Bratman, M. (1992). Shared cooperative activity. *The Philosophical Review, 101*(2), 327–341.

Bratman, M. (2009). Modest sociality and the distinctiveness of intention. *Philosophical Studies, 144*, 149–165.

Bratman, M. (2014). *Shared agency: A planning theory of acting together*. Cambridge, MA: Harvard University Press.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*(5), 187–192.

Chliaoutakis, A., & Chalkiadakis, G. (2014). Utilizing agent-based modeling to gain new insights into the ancient Minoan civilization. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems* (pp. 1371–1372). Paris: International Foundation for Autonomous Agents and Multiagent Systems.

Chopra, S., & White, L. (2011). *A legal theory for autonomous artificial agents*. Ann Arbor: The University of Michigan Press.

Cowley, S. (2014). Cognition beyond the body: Looking differently at ABM. Invited lecture, *The Society for the Study of Artificial Intelligence and Simulation of Behavior (AISB)*. Workshop IV: Modeling Organizational Behavior and Social Agency, Bournemouth.

Darley, V. (1994). Emergent phenomena and complexity. In R. Brooks & P. Maes (Eds.), *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (pp. 411–416). Cambridge: MIT Press.

Darwall, S. (2006). *The second person standpoint: Morality, respect and accountability*. Cambridge, MA: Harvard University Press.

Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.
Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics – Part B, 26*(1), 29–41.
Eymann, T. (2003). *Digitale Geschäftsagenten*. Berlin: Springer.
Ferber, J. (1999). *Multi-agent system: An introduction to distributed artificial intelligence*. Harlow: Addison Wesley Longman.
Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines, 18*(3), 303–329.
Floridi, L. (2011). *The philosophy of information*. Oxford: Oxford University Press.
Goldstein, J. (1999). Emergence as a construct: History and issues. *Emergence: Complexity and Organization, 1*(1), 49–72.
Herzfeld, N. (2013). *When the second person is not a person*. Presented at the Second Person Perspective Conference, Oxford, unpublished manuscript.
Hewitt, C., Bishop, P., & Steiger, R. (1973). A universal modular actor formalism for artificial intelligence. In *Proceedings of the 3rd International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 235–245). Stanford.
Hubig, C. (2010). *Technik als Medium und "Technik" als Reflexionsbegriff*. Manuscript. http://www.philosophie.tu-darmstadt.de/institut/mitarbeiterinnen_1/professoren/a_hubig/downloadbereich/downloadsprofhubig.de.jsp. Accessed 30 Jan 2015.
Hubig, C. (2006). *Die Kunst des Möglichen I – Technikphilosophie als Reflexion der Medialität*. Bielefeld: Transcript Verlag.
Jennings, N. (2010). *ALADDIN End of Project Report*. http://www.aladdinproject.org/wp-content/uploads/2011/02/finalreport.pdf. Accessed 30 Jan 2015.
Kappelhoff, P. (2011). Emergenz und Konstitution in Mehrebenenselektionsmodellen. In J. Greve & A. Schnabel (Eds.), *Emergenz – Zur Analyse und Erklärung komplexer Strukturen* (pp. 319–345). Berlin: suhrkamp taschenbuch wissenschaft.
Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.
Leibnizcenter for Law. (2011). Multi-agent PhD position available. www.leibnitz.org/wp-content/uploads/2011/02/wervering.pdf. Accessed 30 July 2011.
Linderoth, J. (2005). Animated game pieces. Avatars as roles, tools and props. In *Aesthetics of Play Conference Online Proceedings*. http://www.aestheticsofplay.org/linderoth.php. Accessed 30 Jan 2015.
Mainzer, K. (2007). *Thinking in complexity. The complex dynamics of matter, mind, and mankind* (5th ed.). Heidelberg: Springer.
Mayr, E. (2000). *Das ist Biologie – Die Wissenschaft des Lebens*. Heidelberg: Spektrum Akademischer Verlag.
McLaughlin, B. (2008). The rise and fall of British emergentism. In M. Bedeau & P. Humphreys (Eds.), *Emergence: Contemporary readings in philosophy and science* (pp. 19–60). Cambridge, MA: MIT Press.
Mead, G. (1934). *Mind, self, and society from the standpoint of a social behaviourist*. Chicago, IL: The University of Chicago Press.
Moor, J. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems, 21*(4), 18–21.
Muldoon, R., Lisciandra, C., Bicchieri, C., Hartmann, S., & Sprenger, J. (2014). On the emergence of descriptive norms. *Politics, Philosophy & Economics, 13*(1), 3–22.
Nealon, J., & Moreno, A. (2003). Agent-based health care systems. In J. Nealon & A. Moreno (Eds.), *Applications of software agent technology in the health care domain* (pp. 3–18). Whitestein Series in Software Agent Technologies. Basel: Birkhäuser Verlag.
Pagallo, U. (2013). *The law of robots: Crimes, contracts, and torts*. Dordrecht: Springer.
Rammert, W. (2011). *Distributed agency and advanced technology or: How to analyse constellations of collective inter-agency*. The Technical University Technology Studies Working Papers, Berlin.

Ropohl, G. (1999). Philosophy of socio-technical systems. *Society for Philosophy and Technology, 4*(3). http://scholar.lib.vt.edu/ejournals/SPT/v4_n3html/ROPOHL. Accessed 20 Jan 2015.

Ruß, A., Müller, D., & Hesse, W. (2010). Metaphern für die Informatik und aus der Informatik. In M. Bölker, M. Gutmann, & W. Hesse (Eds.), *Menschenbilder und Metaphern im Informationszeitalter* (pp. 103–128). Berlin: LIT Verlag.

Savarimuthu, B., & Cranefield, S. (2011). Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems – An International Journal, 7*, 21–54.

Scheutz, M., Madey, G., & Boyd, S. (2005, April). tMANS – The Multi-Scale Agent-Based Networked Simulation for the study of multi-scale, multi-level biological and social phenomena. In *Proceedings of Spring Simulation Multiconference (SMC 05), Agent-Directed Simulation Symposium*. San Diego.

Teubner, G. (2006). Rights of non-humans? Electronic agents and animals as new actors. *Journal of Law and Society, 33*, 497–521.

Thomas, M. E. (2013). Confessions of a sociopath. *Psychology Today, 46*(3), 52–61.

Thürmel, S. (2012, July). A multi-dimensional agency concept for social computing systems. In G. Dodig-Crnkovic, A. Rotolo, G. Sartor, J. Simon, & C. Smith (Eds.), *Proceedings of the AISB/IACAP Word Congress Social Computing, Social Cognition, Social Networks and Multiagent Systems* (pp. 87–91). Birmingham.

Thürmel, S. (2013). *Die partizipative Wende: Ein multidimensionales, graduelles Konzept der Handlungsfähigkeit menschlicher und nichtmenschlicher Akteure*. Dissertation. Technische Universität München.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Turkle, S. (2010). In good company? On the threshold of robotic companions. In Y. Wilks (Ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues* (pp. 24–34). Amsterdam: John Benjamins.

Turkle, S. (2011). *Alone together, why we expect more from technology and less from each other*. New York, NY: Basic Books.

Vintiadis, E. (2014). *Emergence*. Internet Encyclopedia of Philosophy. http://www.iep.utm.edu/emergenc. Accessed 30 Jan 2015.

Wallach, W., & Allen, C. (2008). *Moral Machines: Teaching Robots Right from Wrong*, New York: University of Oxford Press.

Wedde, H., Lehnhoff, S., Rehtanz, Chr., & Krause, O. (2008). Bottom-up self-organization of unpredictable demand and supply under decentralized power management. In *Proceedings of the 2nd IEEE International Conference on Self-Adaptation and Self-Organization (SASO'08)* (pp. 10–20). Venice: IEEE Press.

Woolridge, M. (2009). *An introduction to multi-agent systems* (2nd ed.). New York, NY: John Wiley & Sons.