Klein, Dominik; Marx, Johannes; Fischbach, Kai

Agent-Based Modeling in Social Science, History, and Philosophy : An Introduction

Date of secondary publication: 03.07.2024

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-962960

Primary publication

Klein, Dominik; Marx, Johannes; Fischbach, Kai (2018): „Agent-Based Modeling in Social Science, History, and Philosophy : An Introduction". In: Historical social research : HSR = Historische Sozialforschung, Vol. 43, Nr. 1, pp. 7-27, Mannheim ; Köln ; Berlin: Gesis, doi: 10.12759/hsr.43.2018.1.7-27.

Agent-Based Modeling in Social Science, History, and Philosophy. An Introduction

Dominik Klein, Johannes Marx & Kai Fischbach *

Abstract: *»Agentenbasierte Modelle in den Sozialwissenschaften, Geschichte und Philosophie«* Agent-based modeling has become a common and well-established tool in the social sciences and certain of the humanities. Here, we aim to provide an overview of the different modeling approaches in current use. Our discussion unfolds in two parts: we first classify different aspects of the model-building process and identify a number of characteristics shared by most agent-based models in the humanities and social sciences; then we map relevant differences between the various modeling approaches. We classify these into different dimensions including the type of target systems addressed, the intended modeling goals, and the models' degree of abstraction. Along the way, we provide reference to related debates in contemporary philosophy of science.

Keywords: Agent-based Modeling, Methodology, Toy Models, Philosophy of Modeling.

1. Introduction

Agent-based modeling has become a common and well-established tool in the social sciences and certain of the humanities. Broadly speaking, it affords a way to study a social, economic, historic, or political phenomenon by examining the iterated interactions of individuals that give rise to the phenomenon. Agent-based models understood in this way are nothing new; Adam Smith's theory of the invisible hand (Smith 2005), first proposed in 1759 and which posits that individual self-interested actions may lead to overall social and economic benefits as non-intended consequences, bears important features of an agent-based model (see also Gavin 2018, in this HSR Special Issue). Much more recently, Schelling's segregation model (1969), which shows that segre-

* Dominik Klein, Political Science Department, University of Bamberg, Feldkirchenstraße 21, 96052 Bamberg, Germany; dominik.klein@uni-bamberg.de.
Johannes Marx, Political Science Department, University of Bamberg, Feldkirchenstraße 21, 96052 Bamberg, Germany; johannes.marx@uni-bamberg.de.
Kai Fischbach, Information Systems Department, University of Bamberg, An der Weberei 5, 96047 Bamberg, Germany; kai.fischbach@uni-bamberg.de.

Historical Social Research 43 (2018) 1, 7-27 │published by GESIS
DOI: 10.12759/hsr.43.2018.1.7-27

gation can occur even within a society of fairly liberal agents, qualifies as an agent-based model. The latter has an additional feature: All agents' behavior is specified by explicit mathematical rules, which allows the modeler to execute the dynamics of the model herself (on a chessboard, as Schelling invites the reader to do). This could, of course, also be done with a computer, giving rise to agent-based simulations.

With the growing availability of personal computers since the late 1980s, a worldwide community of scholars has been using simulations for various reasons. Within agent-based modeling, the use of simulation methods is so ubiquitous that "agent-based model" and "agent-based computational model" have become synonymous. In fact, we address only computational models herein.

Agent-based simulations have a number of distinct advantages over other methodologies. We point to four. First, simulations allow for numerical solutions of mathematical descriptions of social systems that are not tractable employing classical means, which is particularly important for agent-based models, where a large number of possibly heterogeneous agents can have complex interactions over an extended period. Second, agent-based models provide a way to bridge the micro-macro gap. While social processes are, in general, produced through the interaction of individuals, the emergent social patterns need not be related to these individual actions in any straightforward way (List and Spiekermann 2013). Agent-based models allow to map this micro-macro gap and thereby to demystify processes of emergence through scrutinizing the supervenience of macroscopic phenomena on micro level processes. By simulating the iterated interactions of agents over time, agent-based models allow the re-creation of the relevant processes of emergence step by step. *A fortiori*, by systematically varying individual parameters, agent-based simulations provide a way to gain a fine-grained understanding of how emergent patterns depend upon the exact types of interactions. In some cases, the phenomena of interest may be realized in multiple ways at the micro level making them micro-level robust. In other cases, the realization of the phenomena of interest is dependent on the presence of well-defined variables or results from random processes. Third, by controlling random effects and input parameters, agent-based models make it possible to study individual runs of non-deterministic processes. Such a degree of precision allows for the discovery and analysis of path dependencies and tipping points – phenomena that are ubiquitous in social processes, but are extremely difficult to address with classical, statistic means. Fourth, agent-based models often come with their own, distinctive representational tools. For instance, NetLogo, a software packet used for most simulations presented here, provides a graphic interface of the agents' landscapes that develop over time (Wilensky 1999). Such graphic tools have turned out to be extremely useful for the presentation and dissemination of simulation results. They can also play an important role in the context of discovery, pointing to new and unexpected patterns that call for an explanation.

Agent-based models have created a number of landmark contributions in philosophy and the social sciences. Space allows us to mention only a few. As early as 1981, Axelrod employed agent-based modeling to analyze the evolution of cooperation in collective action dilemmas (Axelrod 1981). In 1996, Epstein and Axtell presented their famous Sugarscape model, an agent-based social simulation to discuss social dynamics in an artificial society with a focus on the emerging distribution of wealth. Hegselmann and Krause (2002) and Deffuant et al. (2000) created formal models of belief dynamics showing that belief polarization – a pattern often linked to irrational behavior – can occur even among rational agents when they are faced with the problem of simultaneously determining who to take seriously and how to update their own beliefs. The role of high-order beliefs is discussed in Arthur (2006) with the El Farol model, named for a famous bar in Santa Fe, New Mexico. Modeling the patrons of said bar and their decisions on whether or not to attend at a given evening Arthur shows how solutions to game theoretic coordination problems can arise autonomously, through agents forming higher order beliefs about each other.

Within philosophy of science, two seminal contributions address the social and organizational makeup of science and how it influences scientific progress. Weisberg and Muldoon (2009) assess how the parallel existence of two types of scientists – mavericks exploring new and risky approaches and followers pursuing well-established theories – can advance science to levels that neither type alone could have achieved. Zollman (2010), in contrast, comes to the somewhat surprising conclusion that impeding scientists from exchanging intermediate results may be beneficial for the overall progress of science in the long run.

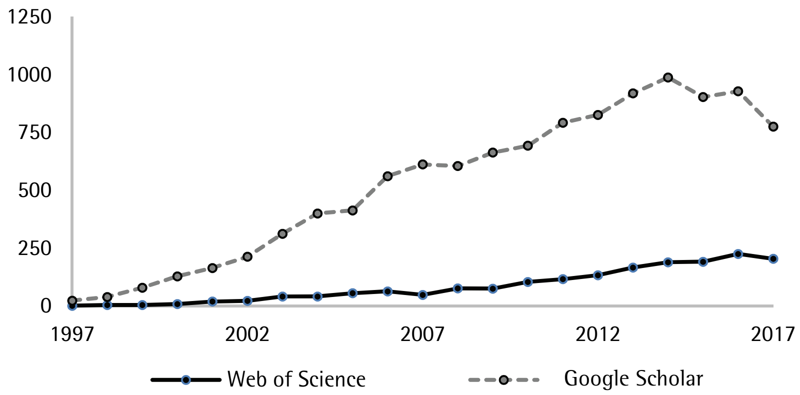
Until recently, agent-based models in large parts of the social sciences and the humanities had barely managed to enter the well-established, mainstream journals in their respective disciplines. Rather, agent-based modelers tended to "form a community of their own with their own journals and annual meetings," as Hedström and Manzo put it (2015, 180). They published their results in specialized journals such as the *Journal of Artificial Societies and Social Simulation* or *Computational and Mathematical Organization Theory*. In mainstream journals, agent-based models were typically found in special issues devoted to the method (e.g., *Revue Française de Sociologie* 2014, *American Journal of Sociology* 2005).

The impact of agent-based modeling on the general research in social science has traditionally been weak. One reason frequently cited to explain this is a discontinuity in methods and perspectives with more mainstream work. For one, the theoretical apparatus is non-standard in a variety of research fields. This holds especially true for fields such as history and philosophy, in which not only programming tools, but also the mathematical language used to express agents and their behavioral rules are not parts of the traditional toolbox.

Moreover, the representation of research problems in agent-based modeling differs from much mainstream work in the respective disciplines. Many of the agent-based models put forward in philosophy and the social sciences are highly abstract, omitting a variety of possibly relevant features of their target systems. The focus on abstract and highly idealized models, again, is far from standard in many disciplines and hence has contributed to deepening the gap between modelers and other practitioners.

Recently, however, the gap between agent-based modelers and standard practices in their broader disciplines seems to be closing. Researchers using simulation techniques no longer define themselves as modelers *per se* but instead refer to agent-based models as a useful tool for puzzle-solving within standard scientific discourse. As a result, more and more articles using agent-based simulations have been published in recent years.

Figure 1: Number of Articles Indexed in Google Scholar and the Social Science Citation Index (Web of Science) with "agent-based model," "agent-based modeling," or "agent-based modelling" in the Title, by Year.[1]



In light of these developments, we take it to be the right moment for an overview of current research in agent-based modeling. With the present HSR Special Issue, we aim to present a cross-section of current approaches to agent-based simulations. By taking stock and comparing the different approaches, our goal is to contribute to a deepened understanding of the assumptions, targets, opportunities, and pitfalls of agent-based modeling in the social sciences and humanities. The papers included in this issue come from a variety of disciplines, including history, literature studies, political science, and philosophy.

[1] Data sampled in late December 2017. Numbers for 2017 may not be complete, as it typically takes some time until papers are included in the databases.

Their broad range of purposes includes prediction, explanations of past events as well as general patterns, understanding of structural relationships, consistency checks of pre-formal theories, and, finally, robustness assessments of existing models. The models presented are aimed at a variety of target systems, including singular events in history as well as general event types, but also informal theories such as the works of John Stuart Mill and Adam Smith. Models discussed in this issue differ in their degrees of abstraction, ranging from empirically calibrated parameters to highly abstract representations. Finally, the contributions differ in how they relate to agent-based simulations. Some create new models aimed at a particular target system, whereas others apply existing models to new targets or discuss the virtues and limitations of earlier models. In short, the presented models cover a broad variety of aspects and features relevant in current agent-based modeling.

2. Characteristic Features of Agent-Based Models

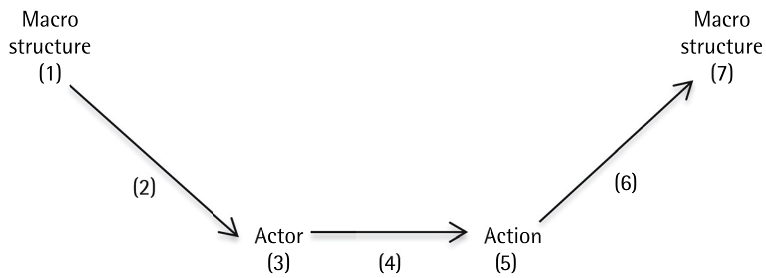
All the papers contributed to this special issue employ agent-based models for their argumentation. By representing the iterated adaptive interactions of agents, such models are particularly suitable for analyzing the dynamics of a social system. Within a model, the dynamics at the social level is emergent; that is, it occurs as the aggregated outcome of iterated interactions between its agents, which are the constituent units of the system (Bonabeau 2002). These agents can be individuals, organized groups, firms, or even states – indeed, any type of entity that owns so-called "actorness," that is, that forms perceptions of its surroundings and has the capacity to engage in autonomous decision-making and intentional action.

Within an agent-based model, social systems are modeled as a set of agents located within a suitable space, a set of behavioral rules for the agents and, lastly, a mechanism that guides the agents' interactions, including their actions' feedback on the structure. Thus, the workings of agent-based models can be reconstructed with Coleman's Bathtub, a methodological framework for explaining how macro-level structures influence micro-level behavior and vice versa.[2] The Coleman's Bathtub framework includes all the parts relevant for explaining social dynamics through individual interactions (see Figure 2): (1) the social structure within which an agent is placed; (2) a mechanism that describes how social structure impacts individual agents at the micro level; (3) the agents, with their perceptions and decision-making capacity; (4) a mechanism providing for how agents choose their actions; (5) the actions chosen; (6) a transformational mechanism that describes how the agents' behavior at the

2 For a discussion of the framework and its implications, see Udehn (2001), Hedström (2005).

micro level influences the macro structure; and (7) the updated macro level as the explanandum.

Figure 2: Coleman's Bathtub (see Coleman 1986: 1322)



Coleman's Bathtub was originally developed as a general framework for explanations in the social sciences. Its seven parts delineate those aspects an explanation in the social sciences should include to be considered complete. We use it here to reconstruct the basic and constituent parts of an agent-based simulation as a special form of a social explanation. Coleman's framework helps clarify the different levels at which modelers must decide how to construct their models as representations of their intended target domain, relative to their intended explanatory purposes. The discussion thus reveals common and basic design elements of all agent-based models in the social sciences and humanities.

1) In a first step, the framework demands that the researcher denote the features of the macro level that are held to be relevant for the phenomenon of interest. In many simulations, these features will create a space within which the agents are located and possibly navigate. The agents' location within the surrounding space may play a role in their perception of their environment, define the local social structure in which they are embedded, and affect their future actions. Some simulation packages for agent-based models, such as NetLogo, mirror this feature explicitly by placing agents on a two- or three dimensional grid and by allowing for networks or spatial inhomogeneities. Notably, the concept of space allows for several possible interpretations. Most naturally, space could be interpreted as physical space, as is done in Klein and Marx (2018). Space might, however, also be interpreted as structuring sets of opinions or belief profiles (Scheller 2018, Baumgaertner 2018), party positions (Schmitt and Franzmann 2018), or scientific theories (Fernández Pinto and Fernández Pinto 2018). Moreover, the modeler has to decide on macroscopic distributional features, and parameter sets supervening on the individual agents and their actions. The distribution of individual properties of the

agents, such as the initial allocation of opinions or resources, is a relevant macro feature of any model. Finally, the modeler must justify the modeling choices. At a minimum, she must provide reasons for her decisions regarding which properties to include, the spatial structure chosen, and the initial distribution of agents' properties. Part of this justification is to provide a mapping between the formal elements of the model and the social phenomena they represent.

2) The framework demands, in a second step, that the modeler provide a mechanism to bridge macro and micro level. Such mechanisms can be found in all agent-based models and, following Coleman, all explanations of social phenomena. Doing computational agent-based modeling forces the researcher to code these mechanisms explicitly. For a multi-agent model, the modeler must name at least one mechanism to explain how features of the social structure (e.g., the distribution of properties of the other agents) influence the agents (e.g., in their opinion formation). This step is also where the modeler needs to describe whether and how macroscopic features such as institutions, rules, or formal and informal norms influence the agents in their perceptions and available actions. Examples of such are institutional rules that limit the agents' available set of options or belief update mechanisms, delineating how the opinions of others shape an agent's mindset.
3) The agents come into play in the framework's next step. In social simulations, the agents are typically equipped with a basic form of perception and capabilities for learning and autonomous decision-making. Such agents may stand for a variety of different actor types, ranging from more or less rational agents of everyday life to highly strategic individuals in specialized contexts to collective actors such as political parties deciding on policy positions (Schmitt and Franzmann 2018). Naturally, the design of the agents should fit the intended research goal. If a model simulates the dynamic of a mass panic, for instance, it might be adequate to treat agents as particles governed by the laws of Newtonian mechanics. When explananda are more complex social phenomena, agents should be designed to include relevant features for their behavior in social life. Depending on the target phenomenon, memory, learning, and perception may all be important. Within a computational model, agents can be heterogeneous and may differ in various aspects; they can, for example, be of different types, equipped with different attributes, and display different behavioral patterns (Bianchi and Squazzoni 2015). Agents can also have their own histories of interactions that may guide their future behavior.
4) The agents' actions are driven by behavioral rules. Such rules could be to select the option that maximizes expected utility or to apply some simpler heuristics in line with the concept of bounded rationality. Moreover, agents may also be guided by social norms (Gavin 2018) or emotional

and psychological factors. In general, all kinds of economic, sociological, or psychological behavioral rules can be applied as long as they allow for a formal operationalization. It is part of the modeler's task to decide on and justify behavioral rules for the agents. Notably, agent-based models do not require that all agents follow the same behavioral rules. Rather, different agents may follow different behavioral policies if that is called for by the target domain. The ability to represent heterogeneous populations, is, in fact, often cited as a major advantage of agent-based models. The only restriction applicable is that all rules must be modeled explicitly.

5) To transform intention into physical action, an agent needs action resources. This seems obvious for human agents, but can be demanding for collective agents such as non-governmental organizations or states. Hence, a modeler may need to say something about the capacity of agents to perform the intended actions and about limitations thereof. While often left implicit when modeling individuals, such considerations frequently occur in contexts of collective agency.
6) Performing actions at the micro level generates consequences for the social structure at the macro level. In the sixth step, an aggregative mechanism is sought to determine how individual actions combine to affect social outcomes. In some cases, the action of a single person can generate enormous effects at the macro level. This, however, is far from usual. In most agent-based models, it is only the iteration and accumulation of individual actions, each with a relatively small effect, that drive the dynamic in the social system in the long run. In such cases, the aggregation mechanism is an endogenous component of the model. While the iterated interaction of the agents is a characteristic design element of simulations, the resulting collective pattern is an emergent consequence thereof. In such cases, the process can be reconstructed as a series of iterated and nested Coleman's Bathtubs.
7) Finally, if successful, the simulation provides an explanation of the phenomenon of interest by uncovering a set of conditions in the model that is sufficient to generate an aggregate pattern corresponding to the target phenomenon.[3] Explaining by means of an agent-based model hence implies uncovering the mechanisms that drive the dynamics of a social system over time. As Squazzoni puts it: According to the social simulation approach, explaining means generating, that is, specifying and showing the generative process through which interacting agents in a given environment combine to produce the macro-regularity of interest (Squazzoni

[3] Not all agent-based simulations aim to explain social phenomena. Some agent-based models are used to test the coherence of theoretical frameworks (Gavin 2018) or gain understanding of certain mechanisms and their interplay. See also the discussion in the following section.

> 2008, 5). The explanations afforded by agent-based models therefore fall into a special class of social explanations, for which Stegmüller coined the term "historically genetic explanations" (Stegmüller 1983).

There are two additional characteristic features of agent-based models to discuss. The first is that designing an agent-based simulation forces the modeler to be explicit about her design choices. Coding a model into a computer requires the researcher to specify all assumptions used in detail and to express all parameters and mechanisms employed within a highly precise formal framework. As a consequence, scientific arguments employing agent-based models tend to be comparatively explicit about their underlying assumptions and their exact scope. This arguably facilitates debates and comparisons between different approaches, as exemplified by Fernández Pinto and Fernández Pinto (2018) and Holman et al. (2018) in this volume. Moreover, the degree of precision that is required renders the task of transforming an informally described argument or theory into a formal model a non-trivial endeavor. In the process of translation, the modeler may encounter relevant aspects that are not sufficiently specified by the original theory, for instance, about the exact learning rules employed by agents in iterated games. Filling in these gaps may require substantial modeling choices. This has a variety of implications about learning from an agent-based model. On the one hand, the necessity to make modeling choices implies that the target theory cannot be tested in isolation, but only in conjunction with the choices made. On the other hand, the process of translation could be instructive: the attempt to transform an informal argument into a formal model may reveal hidden assumptions or gaps in an argument that had previously gone unnoticed. See Baumgaertner (2018) for an example.

Second, agent-based simulations in the humanities and the social sciences frequently share a further aspect: they are highly abstract representations of their target system.[4] Models distort their target systems in various possible ways. The most uncontroversial of these are Aristotelian idealizations (Weisberg 2007; Frigg and Hartmann 2006), sometimes also called abstractions (Cartwright 1989). Models frequently strip away aspects of the target system that are irrelevant for the modeling goal, such as the nationality of agents or their precise identity. This type of idealization is typically considered unproblematic. More controversial is a second type of distortion: Galilean idealizations, which omit, distort, or misrepresent aspects of the target system that are relevant to the phenomenon studied. In this special issue, for instance, Schmitt and Franzmann (2018) put forward a model on the strategic positioning of parties that completely abstracts away from any mobility restrictions parties

[4] Within certain parts of the social science, also large classes of highly complex, less abstract models can be found. For the present purpose, however, we restrict ourselves to highly abstract models as prevail in the humanities.

might face, derived for instance from their members' individual policy interests or considerations of long-term credibility. Even more to the extreme, De Langhe (2018) presents a model of the dynamic of scientific paradigms that omits almost any aspect of scientific theories such as predictive power, simplicity, fit to data, or even the very concept of a subject of science. It is such Galilean idealizations that pose serious puzzles as to what can be learned from highly abstract models, and how. While less abstract models might be evaluated by their fit to the data or the quality of their predictions (Friedman 1953), no such mechanism is available here. We return to the topic of validity in the next section.

3. Models and Validity

While agent-based models share many similarities, they differ widely in, for instance, the (a) type of target system addressed, (b) intended modeling goals, and (c) degrees of abstraction.

a) Let us begin by examining the types of target systems addressed in this special issue. Some of the models presented relate to singular events. Ewert and Sunder (2018), for example, aim to explain why a particular phenomenon, medieval sea trade in the Baltic region, developed as it did. Others address general classes of phenomena, such as the formation of centrist coalition governments (Schmitt and Franzmann 2018) or the editing history of Wikipedia articles (Rudas and Török 2018). Still other models relate not to any target system in the real world, but to counterfactual situations describing how the world might have been but is not. This, again, is reflected briefly in Ewert and Sunder (2018), who discuss why medieval trade in the Mediterranean region did not develop kinship network patterns similar to its Baltic counterpart.

   A further class of models is aimed not at social phenomena directly, but rather at preexisting theoretical accounts of social phenomena. Such informal theories may claim that some macroscopic phenomenon derives from certain individual actions or that some types of behavior uniformly lead to certain desirable outcomes. Here, agent-based models can provide a consistency check. By encoding the informally offered assumptions (e.g., on human behavior) into an agent-based simulation, one can test whether the proclaimed mechanism can actually generate the phenomenon of interest. Along these lines, De Langhe (2018) examines whether the progression of scientific paradigms described in Thomas Kuhn's *The Structure of Scientific Revolutions* (1970) emerges within a society of self-interested strategic scientists. Gavin (2018) uses agent-based models to determine whether two different theoretical approaches stemming from the two main works of Adam Smith are compatible and whether additional phenomena occur at their points of interplay. The process of translating an informal theory into a formal simulation is, notably, far from automatic. Within this process, the modeler may stumble across mechanisms that determine the agents' movements or learning speeds, or parameters that need to be controlled, yet have not been included or sufficiently specified in the original theory. Baumgaertner (2018), for instance, assesses John Stuart Mill's famous argument for opinion diversity in light of recent works on belief polarization. A central parameter for the latter is the agents' homophily. Baumgaertner examines whether empirically plausible levels of this parameter are such that Mill's argument remains sound.

b) A second major dimension is the type of modeling goal pursued. Here, too, there can be substantial differences. Some models aim at explanations of a macro-level phenomenon, the explanandum. The goal is to identify as explanans a certain set of micro-level entities, together with mechanisms, parameters, and interaction rules that jointly generate the target phenomenon in question, as the simulation is to prove. Model behavior is, in general, indeterminate and may exhibit path dependency. Hence, there are several possible outcomes of a simulation. A certain set of input parameters may always create a certain target phenomenon: it may usually create that phenomenon; or the simulation may merely show that the target phenomenon might possibly arise in the given model setting. Depending on the nature of the target phenomenon as well as the explanatory standards applied, either of these could qualify as an explanation. Ewert and Sunder (2018), for instance, assess explanatory power in their approach to medieval trade in the Baltics by considering average performance over a large number of model runs, while Klein and Marx (2018) merely show that certain self-enforcing bubbles of distrust might occur under low mobilities.

   Closely connected to explanation is the goal of prediction. Here, an agent-based model is to be used for predicting how the world will be or how potential choices or policy decisions translate into future states. Within the social sciences, such predictive use of models can be found, for instance, in the field of urban planning (Waddell 2002). This plays less of a role in the current special issue. None of the models presented here aim at detailed qualitative predictions. However, some are targeted or could serve to inform policy decisions, at least over the long run. Borg et al. (2018), for instance, study whether communication among scientists impedes or fosters the overall progress of science. As their model is highly abstract, it allows for qualitative predictions at best. Yet even qualitative answers can be used in discussing the ideal institutional setup of science. The same is true for Rudas and Török (2018), who study the impact of banning policies on the convergence of Wikipedia articles subjected to competing editing.

Still other models aim at *verstehen*, or enhancing the understanding of some mechanism's workings and effects. Take Klein and Marx's (2018) model on the emergence of generalized trust. The authors aim to understand one particular mechanism that plays a role in regulating whether agents are willing to trust hitherto unknown others: the mechanism of learning through direct, first-hand experience. While the model studies how certain parameters such as the agents' initial trust level or their mobility affect the dynamics of general trust, it explicitly omits a variety of relevant other factors. It does so to study direct experiences of trust and trustworthiness and their long-term effects in isolation. In particular, the model does not embed its agents in any type of social circles from which they could learn or where they could share their information. In real life, such indirect sources of information are arguably at least as relevant for the agents' propensity to trust as is their direct experience. The model hence can hardly generate any valid predictions, nor does it aim to do so. Rather, it aims to generate understanding of one particular mechanism that in reality is always conjoined with others. Similar considerations hold for De Langhe's model of Kuhnian scientific paradigms (2018), in which the scientific properties of a paradigm play no role, or for Holman et al.'s account of collaborative problem solving (2018).

Along a similar line, other models aim to create understanding of the interplay among various mechanisms. Rather than addressing a particular social phenomenon, such models seek to increase our understanding of how different well-known mechanisms interact. Scheller (2018), for instance, studies the interplay between deliberative decision making and voting in terms of expected accuracy and the average time it takes to reach a decision. Taken individually, both deliberation and voting are understood sufficiently well. Scheller, however, shows that this alone is not sufficient to understand complex collective decision making. By means of an agent-based model, he reveals that the interplay of both mechanisms gives rise to complex patterns that could hardly have been anticipated by addressing either mechanism in isolation.

c) The third relevant aspect is how agent-based models relate to their respective target systems. This aspect covers how similar model and target phenomenon are to each other, what aspects of the target phenomenon are included in the model and, crucially, how the model affords new insights about the target phenomenon in question. This topic is closely related to validation, that is, the question of how a model can be identified as correct, broadly speaking. Gräbner (2017, 6) and Tesfatsion (2017, 14) identify different forms of model verification and validation: the input parameters might be calibrated to available data (input validation); the modeler might provide arguments for the mechanisms encoded (process validation) and check the internal consistency of the model (model verification); the output might be calibrated against training data (descriptive output validation); and, finally, the model might be judged by the accuracy of its predictions for new data points (prescriptive output validation). It is an open question dating back at least to Friedman (1953) as to whether any subset of these is necessary or sufficient for model validity.

Some of these conditions, however, may be of limited applicability to a large class of agent-based models in philosophy and the social sciences. All the models developed in this special issue qualify as what are sometimes called "toy models": extremely simplified, highly abstract representations of a target system that possibly include a number of Galilean idealizations (Grüne-Yanoff 2009). With such models, quantitative point predictions are, in general, neither possible nor intended.

How, then, can we learn from such models? For one, other dimensions of validation still apply. When models cannot be compared by their accuracy or predictive power, it becomes even more crucial to justify the model mechanism by reference to the target phenomenon. Such justifications can be found in all the contributions to this special issue. Yet, providing justifications for the mechanisms included may not be sufficient. After all, toy models are highly idealized, omitting a variety of relevant factors and mechanisms. The risk, here, is that the result is driven or distorted precisely by what is not included, thus undermining the use of the simulation results for understanding the target phenomenon. This worry is reflected in, for instance, the contribution of Borg et al. (2018), who argue that previous models of communication in science omit relevant aspects of the scientific process, leading them to false conclusions.

Even within highly idealized models, some argument needs to be given for the value range of parameters included. There are two main strategies for doing so. The first is to cite empirical or theoretical evidence that justifies the choice of parameter values. In the present issue, this is done by Mayerhoffer (2018), who calibrates his discussion of tolerance dynamics among adolescents with empirical data. Similarly, Holman et al. (2018) base their criticism of the Hong and Page (2004) model on the claim that the latter misestimates the values of relevant parameters. The second strategy applies when empirical information about parameter values is lacking or incomplete. In this case, the modeler may resort to something akin to an existential quantification. They could show that the target phenomenon in question occurs under a large range of input parameters, broad enough to contain plausibly the unknown real values.

This second strategy is tightly connected to the topic of robustness. If the qualitative outcomes of a model are robust under changes in the model parameters, variations in the mechanisms incorporated, and the exact implementation of the model, these are generally taken as evidence of the model's validity. In this case, the micro-specifications in question are said to satisfy the criterion of "generative sufficiency" (Epstein and Axtell 1996, 6) with regard to the macro-regularity of interest.

A further relevant aspect in determining what can be learned from toy models is the distinction between "how-possible" and "how-actual" understanding, introduced by Reutlinger et al. (2017). In brief, the latter is about determining which mechanism drives a result in the actual world. How-possible understanding, in contrast, merely aims at identifying plausible mechanisms and showing that they are sufficient to produce the target phenomenon, without necessarily claiming that they are at work in the actual world. Having such how-possible understanding hence helps in charting the territory of possible mechanisms. It provides the modeler, at least, with additional insights about the connections between certain structural factors and mechanisms relevant to the target phenomenon. The highly idealized nature of toy models arguably squares well with the concept of how-possible understanding. Toy models, sometimes also referred to as "minimal idealizations" (Weisberg 2007), include only the bare minimum of mechanisms needed to create a desired output. On the one hand, this highly abstract nature makes it difficult to determine whether some real-world counterpart of the model mechanism drives the target phenomenon, that is, whether we acquire "how-actual" understanding. On the other hand, ontological sparseness facilitates determining how parameters and factors drive the target phenomenon in the model. In other words, it is exactly the ontological sparseness of toy models that allows the modeler to gain a high degree of understanding of the mechanisms involved, and hence affords "how-possible" understanding of the target system.

Finally, we point to a further way of learning from toy models: multiple model idealizations (Weisberg 2007). While every model simplifies and distorts the target system, different models will, in general, differ as to which factors to include. If a variety of models agree in some predictions they make, it should hence give greater credence to the idea that the phenomenon is real rather than an artifact of our modeling choices. This point is acknowledged by various contributions in this issue. In particular, Fernández Pinto and Fernández Pinto (2018) argue that replicating a classic model on epistemic landscapes in a different simulation environment and with slightly adapted decision rules lends additional reliability to the original results.

4. Conclusion

The twelve papers in this HSR Special Issue address the trends in agent-based modeling research just described. They cover classical simulation topics such as opinion dynamics, voting behavior, the formation of trust, and the social makeup of science, but provide a fresh perspective. Others put forward novel approaches to problems that have so far had little contact with agent-based models, such as medieval history of trade or Adam Smith and John Stuart Mill scholarship.

All the papers selected share some common characteristics that are, in our minds, typical for agent-based models. We have elaborated on these shared characteristics in Section 2, where we discussed common features in agent-based modeling using Coleman's Bathtub as a theoretical framework. In Section 3, we then offered a theoretical apparatus to categorize relevant differences between various agent-based models with respect to the quality of their target systems, intended modeling goals, and the degrees of abstraction of the models. Within our discussion, we have already pointed out selected aspects of each contribution. We conclude by providing brief summaries of all twelve papers included in this issue.

Rogier De Langhe presents "An Agent-Based Model of Thomas Kuhn's *The Structure of Scientific Revolutions.*" He shows that the patterns of normal science and scientific revolutions, and the corresponding succession of paradigms described by Kuhn, can emerge naturally within a society of strategic, self-interested scientists. By translating Kuhn's ideas into an agent-based model, he shows how mechanisms of self-organization can endogenously generate dominant paradigms, thereby providing additional support to Kuhn's analysis of the progress of science.

Manuela Fernández Pinto and *Daniel Fernández Pinto* criticize in "Epistemic Landscapes Reloaded: An Examination of Agent-Based Models in Social Epistemology" Weisberg and Muldoon's epistemic landscape model (Weisberg and Muldoon 2009). This highly abstract model studies the effects of mixing different types of scientist personalities, mavericks and followers, on the long-term success of science. Fernández Pinto and Fernández Pinto challenge the finding of the epistemic landscape model by showing that small changes in the rules determining the behavior of follower-type agents can lead to significant changes in simulation results. They go further to argue that the epistemic landscape model results are robust with respect to the computing mechanism, but not necessarily across agent realizations. Finally, they conclude by offering some thoughts on what this implies for general lessons drawn from this class of models.

Csilla Rudas and *János Török* address consensus-building processes in their paper on "Modeling the Wikipedia to Understand the Dynamics of Long Disputes and Biased Articles." Their analysis focuses on the dynamics of Wikipedia articles that are subjected to competing revisions by several editors, each with their own perspective on the matter. By means of an agent-based model, Rudas and Török study how the attitudes of editors affect the consensus-building process. They show that in most cases banning extreme agents from editing an article slows down the consensus-building process. Somewhat counterintuitively, they also find that having large groups of extremists who hold opinions far of the center accelerates the consensus-building process and leads to articles that are less biased.

Simon Scheller examines the claim that democratic decision making has an intrinsic epistemic value in "When Do Groups Get It Right? On the Epistemic Performance of Voting and Deliberation." The author presents an agent-based model that combines the two main modes of collective decision making: deliberation and voting. Under the assumption that there is an objective fact of the matter, the paper studies the circumstances under which groups are able to identify reliably the "correct alternative." The research scrutinizes the performance of groups under varying conditions on both, communication behavior and voting thresholds. Simulation results show that larger majority requirements in voting can increase expected adequacy in well-functioning groups, but can also award a veto power to closed-minded individuals. Further, the author concludes that when independent information acquisition is possible, reasonable skepticism regarding other people's opinions can provide a useful impediment against overly quick convergence on a false consensus.

Ulf Christian Ewert and *Marco Sunder*, in "Modelling Maritime Trade Systems: Agent-Based Simulation and Medieval History," adopt a fresh perspective on the medieval history of trade in the Baltic. In particular, they engage in an ongoing debate on the causes of both the formation of the Hanse's kinship network-based system of trade in Northern Europe and its later dissolution. They do so by applying a multi-agent model to the analysis. In addition to connecting their findings to ongoing discussions in institutional economics and economic history, the authors address the prospects and limitations of using agent-based simulation in historical research.

Daniel Mayerhoffer models adolescents in their formation of attitudes towards topics of sexual diversity. In "Raising Children to Be (In-)Tolerant" he shows that civil society has a strong impact on the level of tolerance these develop, while the church and state run educational institutions only have minor bearing. The agent-based model Mayerhoffer constructs is based on extensive survey data on the environments and attitudes of adolescents in Germany.

In "A Polarizing Dynamic by Center Cabinets? The Mechanism of Limited Contestation," *Johannes Schmitt* and *Simon Franzmann* analyze how patterns of government formation influence the polarization of party systems. By means of an agent-based simulation, they reveal the mechanisms that drive a polity to extremes. Their model focuses on the effects on polarization in party systems where the governing coalition is one of the ideological center.

In his paper on "Models of Opinion Dynamics and Mill-Style Arguments for Opinion Diversity," *Bert Baumgaertner* demonstrates how models of opinion dynamics can help us gain a better understanding of John Stuart Mill's argument for opinion diversity. He contrasts Mill's original argument with recent work on belief polarization driven by homophily, the tendency to preferably interact with like-minded individuals. Baumgaertner's findings show that realistic levels of homophily do not sufficiently explain decreased opinion diversity. The results provide insights into how arguments such as Mill's depend on hidden assumptions about the psychology and sociality of individuals. The results, thereby, increase our understanding of such arguments by bringing underlying assumptions out into the open, but also by restricting the scope of validity of said arguments.

Dominik Klein and *Johannes Marx* contribute with "Generalized Trust in the Mirror An Agent-Based Model on the Dynamics of Trust" to the debate about trust in larger societies. Generalized trust – the propensity to place trust in strangers – has attracted much attention in political science lately, as it has been identified as a central determinant for the economic and political success of modern states. In their analysis, the authors focus on understudied micro-level processes that determine the emergence and stability of generalized trust. They investigate conditions under which trust is likely to emerge and be sustained over an extended period. In contrast to predictions from the literature, Klein and Marx find that low degrees of geographic and social mobility are detrimental to both the emergence and stability of trust. They also identify a hidden, emergent link between trusting others and being trustworthy.

In "Diversity and Democracy: Agent-Based Modeling in Political Philosophy," *Bennett Holman, William J. Berger, Daniel J. Singer, Patrick Grim* and *Aaron Bramson* challenge the results of a highly cited paper by Hong and Page (2004) on the performance of epistemic groups. While Hong and Page conclude that diversity, rather than individual abilities, is the main driver of group competence, Holman et al. argue that some of their model's basic assumptions are due to misidentifications and the erroneous conceptualization of expertise. They argue that correct modeling leads to different results. Furthermore, the authors demonstrate how sensitive the finding is to changes in parameters and provide an analysis of when diversity trumps ability in terms of group performance.

AnneMarie Borg, Daniel Frey, Dunja Šešelja, and Christian Straßer investigate "Epistemic Effects of Scientific Interaction: Approaching the Question with an Argumentative Agent-Based Model". Their research contributes to a lively debate over whether increased communication among scientists is beneficial or harmful to the overall success of science. The authors argue that a variety of existing models on this question misrepresent or omit crucial aspects and parameters. On this basis, they call into question the relevance of some of these results for actual scientific practice. The authors then present a novel agent-based model that represents scientific interaction in terms of discovering and exchanging arguments for and against scientific theories. This allows them to propose novel hypotheses and assess the robustness of previously obtained results under different modeling choices.

In “An Agent-Based Computational Approach to ‘The Adam Smith Problem,’” *Michael Gavin* addresses the longstanding observation that the two main works of Adam Smith rest on seemingly contradictory assumptions about human nature: that we are self-interested, or that we are guided by social norms and a concern for others. By applying two agent-based models, Gavin demonstrates how both assumptions may be compatible with each other, after all. Moreover, he shows that the interplay between both factors entails significant consequences that Smith himself may have not anticipated. The author derives four propositions from his analysis, two of which offer a new perspective on the Adam Smith Problem. Here, the use of agent-based modeling may well be the starting point for a newly energized debate on a classic theme.

All contributions in this HSR Special Issue have gone through a thorough double-blind peer review process. We would like to thank the editors and the editorial office of the journal for giving us the opportunity to edit this issue, and for their support. We would also like to thank the numerous reviewers for their thorough and extremely helpful comments. And finally, of course, we’d like to thank all the authors who contributed to this special issue.

Special References

Baumgaertner, Bert. 2018. Models of Opinion Dynamics and Mill-Style Arguments for Opinion Diversity. *Historical Social Research* 43 (1): 210-33. doi: 10.12759/hsr.43.2018.1.210-233.

Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer. 2018. Epistemic Effects of Scientific Interaction: Approaching the Question with an Argumentative Agent-Based Model. *Historical Social Research* 43 (1): 285-307. doi: 10.12759/hsr.43.2018.1.285-307.

De Langhe, Rogier. 2018. An Agent-Based Model of Thomas Kuhn’s *The Structure of Scientific Revolutions*. *Historical Social Research* 43 (1): 28-47. doi: 10.12759/hsr.43.2018.1.28-47.

Ewert, Ulf Christian, and Marco Sunder. 2018. Modelling Maritime Trade Systems: Agent-Based Simulation and Medieval History. *Historical Social Research* 43 (1): 110-43. doi: 10.12759/hsr.43.2018.1.110-143.

Fernández Pinto, Manuela, and Daniel Fernández Pinto. 2018. Epistemic Landscapes Reloaded: An Examination of Agent-Based Models in Social Epistemology. *Historical Social Research* 43 (1): 48-71. doi: 10.12759/hsr.43.2018.1.48-71.

Gavin, Michael. 2018. An Agent-Based Computational Approach to ‘The Adam Smith Problem.’ *Historical Social Research* 43 (1): 308-336. doi: 10.12759/hsr.43.2018.1.308-336.

Holman, Bennett, William J. Berger, Daniel J.Singer, Patrick Grim, and Aaron Bramson. 2018. Diversity and Democracy: Agent-Based Modeling in Political Philosophy. *Historical Social Research* 43 (1): 259-84. doi: 10.12759/hsr.43.2018.1.259-284.

Klein, Dominik, and Johannes Marx. 2018. Generalized Trust in the Mirror. An Agent-Based Model on the Dynamics of Trust. *Historical Social Research* 43 (1): 234-58. doi: 10.12759/hsr.43.2018.1.234-258.

Mayerhoffer, Daniel. 2018. Raising Children to Be (In-)Tolerant. Influence of Church, Education, and Society on Adolescents' Stance towards Queer People in Germany. *Historical Social Research* 43 (1): 144-67. doi: 10.12759/hsr.43.2018.1.144-167.

Rudas, Csilla, and János Török. 2018. Modeling the Wikipedia to Understand the Dynamics of Long Disputes and Biased Articles. *Historical Social Research* 43 (1): 72-88. doi: 10.12759/hsr.43.2018.1.72-88.

Scheller, Simon. 2018. When Do Groups Get It Right? *Historical Social Research* 43 (1): 89-109. doi: 10.12759/hsr.43.2018.1.89-109.

Schmitt, Johannes, and Simon T. Franzmann. 2018. A Polarizing Dynamic by Center Cabinets? The Mechanism of Limited Contestation. *Historical Social Research* 43 (1): 168-209. doi: 10.12759/hsr.43.2018.1.168-209.

References

American Journal of Sociology. 2005. Special issue: *Social science computation*, ed. Nigel Gilbert and Andrew Abbott. Chicago: The University of Chicago Press.

Arthur, W. Brian. 2006. Out-of-Equilibrium Economics and Agent-Based Modeling. In *Handbook of Computational Economics*, ed. Leigh Tesfatsion and Kenneth L. Judd, 1551-64. Amsterdam: Elsevier.

Axelrod, Robert, and William Donald Hamilton. 1981. The evolution of cooperation. *Science* 211 (4489): 1390-6.

Bianchi, Federico, and Flaminio Squazzoni. 2015. Agent-based models in sociology. *Wiley Interdisciplinary Reviews: Computational Statistics* 7 (4): 284-306.

Bonabeau, Eric. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences* 99 (suppl 3): 7280-7.

Cartwright, Nancy. 1989. *Nature's Capacities and their Measurement*. Oxford: Oxford University Press.

Coleman, James. 1986. Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology* 91 (6): 1309-35.

Deffuant, Guillaume, David Neau, Frédéric Amblard, and Gérard Weisbuch. 2000. Mixing beliefs among interacting agents. *Advances in Complex Systems* 3: 87-98.

Epstein, Joshua M., and Axtell, Robert. 1996. *Growing artificial societies: social science from the bottom up*. Washington: Brookings Institution Press.

Friedman, Milton. 1953. *Essays in positive economics*. Chicago: The University of Chicago Press.

Frigg, Roman, and Stephan Hartmann. 2006. Models in science. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. <https://plato.stanford.edu/entries/models-science/> (Accessed January 27, 2018).

Gräbner, Claudius. 2017. *How to relate models to reality? An epistemological framework for the valida-tion and verification of computational models*. ICAE

Working Paper Series 63. <http://hdl.handle.net/10419/171438> [Accessed January 27, 2018].
Grüne-Yanoff, Till. 2009. Learning from minimal economic models. *Erkenntnis* 70 (1): 81-99.
Hedström, Peter, and Gianluca Manzo. 2015. Recent trends in agent-based computational research: A brief introduction. *Sociological Methods & Research* 44 (2): 179-85.
Hedström, Peter. 2005. *Dissecting the Social. On the Principles of Analytical Sociology*. Oxford: Oxford University Press.
Hegselmann, Rainer, and Ulrich Krause 2002. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of artificial societies and social simulation* 5 (3): 1-33.
Hong, Lu, and Scott E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Science*s 101: 16385-9.
Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
List, Christian, and Kai Spiekermann. 2013. Methodological individualism and holism in political science: A reconciliation*. American Political Science Review* 107 (4): 629-43.
Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann. Forthcoming. Understanding (with). Toy Models. *The British Journal for the Philosophy of Science.*
Revue Française de Sociologie. 2014. Special Issue. *Agent-based Simulation: Principles and Applications to Social Phenomena* 55 (4).
Schelling, Thomas C. 1969. Models of Segregation. *American Economic Review* 59 (2): 488-93.
Smith, Adam. 2005. *Wealth of Nations*. University of Chicago. Bookstore.
Squazzoni, Flaminio. 2008. The Micro-Macro Link in Social Simulation. *Sociologica* 18 (1): 1-26.
Squazzoni, Flaminio. 2010. The impact of agent-based models in the social sciences after 15 years of incursions. *History of Economic Ideas* 18 (2): 197-233.
Stegmüller, Wolfgang. 1983. *Probleme und Resultate der Wissenschaftstheorie und analytischen Philosophie. 1,C: Erklärung, Begründung, Kausalität: Historische, psychologische und rationale Erklärung. Verstehendes Erklären.* Berlin, Heidelberg, New York: Springer.
Tesfatsion, Leigh. 2017. *Modeling Economic Systems as Locally-Constructive Sequential Games*. Economics Working Paper 17022. [*Journal of Economic Methodology* 24 (4): 384-409]
Udehn, Lars. 2001. *Methodological Individualism: Background, History and Meaning*. London, New York: Routledge.
Waddell, Paul. 2002. UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning. *Journal of the American Planning Association* 68 (3): 297-314.
Weisberg, Michael, and Ryan Muldoon. 2009. Epistemic Landscapes and the Division of Cognitive Labor. *Philosophy of Science* 76 (2): 225-52.
Weisberg, Michael. 2007. Three kinds of idealization. *The Journal of Philosophy* 104 (12): 639-59.

Wilensky, Uri. 1999. *NetLogo*. <http://ccl.northwestern.edu/netlogo/> [Accessed January 27, 2018]. Center for Connected Learning and Computer-Based Modeling: Northwestern University, Evanston, IL.
Zollman, Kevin JS. 2010. The epistemic benefit of transient diversity. *Erkenntnis* 72 (1): 17-35.