

Agent-Based Simulation and Sociological Understanding

Petri Ylikoski

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-148454>

N.B.: When citing this work, cite the original publication.

Ylikoski, P., (2014), Agent-Based Simulation and Sociological Understanding, *Perspectives on Science*, 22, 318-335. https://doi.org/10.1162/POSC_a_00136

Original publication available at:

https://doi.org/10.1162/POSC_a_00136

Copyright: Massachusetts Institute of Technology Press (MIT Press) (6 month embargo)

<https://mitpress.mit.edu/>



Agent-Based Simulation and Sociological Understanding

Petri Ylikoski

Department of Social Research

University of Helsinki

petri.ylikoski@helsinki.fi

Abstract

The article discusses agent-based simulation as a tool of sociological understanding. Based on an inferential account of understanding, it argues that computer simulations increase our explanatory understanding both by expanding our ability to make *what-if* inferences about social processes and by making these inferences more reliable. However, our ability to understand simulations limits our ability to understand real world phenomena through them. Thomas Schelling's checkerboard model of ethnic segregation is used to demonstrate the important role played by abstract *how-possibly* models in the process of building a mechanistic understanding of social phenomena. 1. **Introduction**

This article discusses agent-based simulation (ABS) as a tool of sociological understanding. I argue that agent-based simulations can play an important role in the expansion of explanatory understanding in the social sciences. The argument is based on an inferential account of

understanding (Ylikoski 2009, Ylikoski & Kuorikoski 2010), according to which computer simulations increase our explanatory understanding by expanding our ability to make *what-if* inferences about social processes and by making these inferences more reliable. The inferential account also suggests a number of ways in which the use of simulation methodology might give rise to illusory understanding.

The structure of the article is as follows. I begin by describing the challenge of explaining macro social phenomena in terms of causal mechanisms. I then introduce the idea of ABS and explain why an increasing number of social scientists find it a promising tool for a mechanistic understanding of social phenomena. The sociological opportunities raised by ABS will be illustrated by Thomas Schelling's famous checkerboard model of ethnic segregation. The latter part of the paper will concentrate on giving a philosophical account of the contribution agent-based simulation models can make to sociological understanding. Then, employing the inferential account of explanatory understanding, I show how computer simulation can be understood as a case of extended cognition. I conclude the article with two applications of the proposed account of understanding. First, I articulate a possible source of illusory understanding in simulation studies: the confusion between understanding a simulation and understanding the target phenomenon using the simulation. Second, I reconstruct some reasons why highly abstract *how-possibly* explanations – such as Schelling's checkerboard model – are seen as providing explanatory insight.

2. Social Mechanisms and Agent-Based Simulation

One of the central challenges of sociology is to make sense of macro social facts. Such facts concern for example the distribution of individuals into different classes and status groups, social

networks existing between entrepreneurs, the social norms characterizing social groups, and the segregation of members of various groups into different social activities or residential areas. While there is no general sociological agreement about the nature of sociological explanation, it has become increasingly popular to think that social explanation should explicate the causal mechanisms that generate the phenomenon to be explained (Hedström & Ylikoski 2010). As this view is almost unanimously accepted by social scientists who advocate ABS, I will use it as the starting point of my discussion.

One of the basic premises of the mechanistic approach is that a proper understanding of collective processes requires paying attention to the entities that mechanisms are made of (the agents, their properties, actions, and relations) rather than treating them as black boxes. To account for macro-level phenomena it is not sufficient to simply relate macro-level phenomena to one another. For example, to explain macro-level change, a theory must specify the set of causal mechanisms that can bring about the change, and this requires demonstrating how macro states at one point in time influence individuals, and how their actions generate new macro-states at a later point in time. In other words, the mechanistic ambition requires that a full explanation of social phenomenon spell out three kinds of mechanisms. The *situational mechanisms* describe how social structures constrain individuals' action and how cultural environments shape their desires and beliefs, the *action-formation mechanisms* describe how individuals choose their preferred courses of action from the feasible alternatives, and *the transformational mechanisms* describe how individual actions produce various intended and unintended social outcomes (Coleman 1990; Hedström & Ylikoski 2010).

While both situational and action-formation mechanisms certainly pose their own challenges, it is generally agreed that the transformational mechanisms are the greatest bottleneck in sociological theory (Coleman 1990). Social scientists know a great deal about how individuals' desires, beliefs, opportunities, and so on are influenced by the social contexts in which they are embedded (the situational mechanisms), and they also know a great deal about how these desires, beliefs, and opportunities influence actions (the action-transformation mechanisms); but when it comes to the link between individual actions and social outcomes, they are often forced to resort to hand-waving. Macro-level patterns are often difficult to predict from the individual-level descriptions, and the way in which individual actions produce social patterns is rarely a simple process of aggregation.

One of the key challenges is that of generative sufficiency. Social scientists must be able to demonstrate that the suggested explanatory factors are at least *in principle capable* of bringing about the outcome within the intended circumstances. However, demonstrating generative sufficiency is not easy. The social outcomes of interest are typically results of numerous individuals acting and interacting with one another over extended periods of time. Furthermore, the aggregate behavior of these complex dynamic systems is extremely difficult to understand and to predict without the aid of analytical tools. Without such tools it is difficult to recognize, and even more difficult to convince others, that the large-scale phenomena that one seeks to explain may be the result of a particular type of mechanism.

ABS naturally combines with the idea of generative explanation. A natural affinity exists between the components of mechanism-based explanations and ABS models. As with society, a social scientific ABS is composed of agents having goals and beliefs. These agents interact with

each other and they possess the resources needed for attaining their goals. Finally their behaviors affect each other's possibilities to act. With the goal of generative explanations being to understand how macro social patterns are created and maintained as the intended and unintended outcomes of individual actions, it is easy to see why social scientists regard the emergence of macro patterns in cellular automata as analogical to the processes they study. An ABS allows us to see how the phenomenon to be explained could have been generated and how changes in agents' attributes or relational structures are likely to change the macro outcome. Similarly, a well-designed simulation can expand the reach of sociological theory and raise new and well-defined problems for empirical research.

The strongest formulation of the idea of generative explanation employing ABS has been presented by Joshua Epstein (2006). According to him, "*if you didn't grow it, you didn't explain it*". His idea is that growing the macro-level outcome by ABS is a necessary condition for its explanation. The strong claim for this necessity can be challenged by pointing out that there have been satisfactory explanations in the social sciences before the invention of ABS. Second, simply "growing" the phenomenon of interest is not sufficient for a proper understanding of it. The crucial challenge is to understand *how* the specified micro-configuration produces the phenomenon and what the background conditions for this production are. For this reason, Macy and Flache (2009, 263) are right in slightly changing Epstein's slogan. According to them, it should read: "*If you don't know how you grew it, you didn't explain it.*" Finally, it is important to remember that simulations have at most the status of a *how-possibly* explanation, as there might be alternative ways to generate the same result. Generative sufficiency is an essential step, but it

does not imply that the empirical outcome to be explained was actually brought about in that fashion.

Despite these reservations, Epstein's slogan touches on something important: if we are interested in mechanism-based bottom-up explanations of social phenomena, ABS provides a much needed tool for theoretically oriented social scientists. Much of the development of mechanism-based knowledge consists of developing how-possibly explanation schemes (Ylikoski & Aydinonat 2011). The purpose of these is not to explain any particular empirical facts directly; rather, they can be regarded as dynamic thought experiments whose purpose is to provide a general understanding of how things could work in principle.

Such thought experiments are not mere fairy tales when used as a part of systematic theoretical research that explores a series of *what-if* questions. Given the limitations of experimental methods and the complexity of social phenomena, computer simulations are important for this kind of endeavor. They allow a systematic exploration of the consequences of modeling assumptions and make it possible to model much more complex phenomena than was earlier possible.

When contrasted with more traditional formal modeling tools, the advantages of ABS are numerous. With ABS one need not limit one's attention to equilibrium outcomes: one can also study the dynamics of the system. This makes it possible to study cyclical processes, path dependence, evolution, and the like. Similarly, one need not limit one's attention to cases with either one, two, or an infinite number of agents, as is the case with standard economic models. Being free of this restriction makes room for a much more realistic modeling of social processes. ABS does not impose *a priori* constraints on the mechanisms assumed to be operating. Unlike

rational choice theory, ABS modeling is not based on a specific theory of action or interaction. It is a methodology for deriving the social outcomes that groups of interacting actors are likely to bring about, whatever the properties of the agents or structures governing their interaction may be. This advantage is based on the flexibility of ABS with respect to the characteristics of the agents. One does not have to assume the hyper-rational all-knowing agents of economic theory. (Miller & Page 2007: 80-87.)

Agents in ABS models have both cognitively and socially interesting properties. From the point of view of cognition, the agents are heuristic and adaptive. They are *heuristic* in the sense that they follow simple behavioral rules that can be interpreted as habits, rituals, routines, and norms. Their *adaptiveness* is based on the fact that they respond to feedback from their environment through learning and evolution, and as a consequence this can alter the probability that a given behavior will recur. As well, the social properties of agents are interesting. First, agents are *autonomous* in the sense that they have their own goals, behaviors, and internal states and can change aspects of their environment to attain the goals. Second, *interdependence* constrains agents' autonomy. Agents are behaviorally and strategically interdependent in the sense that the perceived local environment to which they react is partially dependent on the beliefs, goals, and behaviors of other agents. In other words, any agent's behavior may change the environment of other agents. ABS methodology also allows that agents are *heterogeneous*. In contrast to more traditional game-theoretical and system dynamics models, agent-based models make it possible to have agents that differ in their beliefs, goals, and decision procedures. Finally, agents can be *embedded* in networks. This means that population dynamics are an emergent property of local interactions. ABS also allows that agents change their structural locations or break off relations

with some neighbors and seek out others. (Miller & Page 2007: 80-87; Macy & Flache 2009: 247.)

One advantage of traditional analytical models is that their results are expressed in the form of algebraic formulas that provide information for all possible values of variables and parameters, thus making them more general than simulation models. In other words, they provide a more concise way of understanding the crucial relationships in the model (Lehtinen & Kuorikoski 2007: 320). This is a clear advantage; such general understanding can easily be extrapolated to novel cases. Thus it would be an advantage if such an understanding of the dependencies underlying the simulation were possible. However, as there are strict limitations on what can be analytically modeled, one cannot argue that the understanding provided by agent-based simulations reduces to the understanding provided by analytical models. The former allow modeling of systems whose analytical treatment is either very difficult or impossible. For this reason it is legitimate to argue that agent-based simulation *expands* the scope of understanding and the two modeling strategies should be treated as complementary rather than competing.

Agent-based simulations should not be regarded only as a tool of theoretical exploration, however. Empirically calibrated ABS models make it possible to integrate theoretical ideas with the results of empirical research (Hedström 2005, Epstein 2006). For example, nothing prevents imputing empirically observed preferences to agents, or incorporating the complexity of real-world social phenomena. However, these complications present many challenges to explanatory understanding. As Macy and Flache (2009: 263) point out, when a model becomes so complicated that a researcher can only report input-output covariance with a long list of parameters, its ability to provide understanding is very limited. This implies that the pursuit of a

mechanism-based understanding of complex social phenomena can only proceed in small steps. One should only add such elements to the model that are considered essential, and one should always make certain that one understands what is being added to the model.

3. Simulating Segregation

Thomas Schelling's (1971, 1978) checkerboard model of segregation is one of the best known – and probably the most explored – ABS model in the social sciences. It addresses the origins of residential segregation by race or ethnicity. Residential segregation is an important phenomenon, because of its far-reaching social consequences. The neighborhood in which people live affects their quality of life, future opportunities, social networks, occupations, and how they are judged by others. However, the significance of Schelling's model has not been limited to studies in urban geography. It has also attracted much theoretical attention and has been granted exemplary status in both sociology and economics (Aydinonat 2008).

In the original two-dimensional checkerboard model, the world consists of agents inhabiting the squares of a checkerboard. The agents are divided into two classes that represent any binary social division that could affect the distribution of agents in space (e.g., whites and non-whites, or humanists and engineers). While the model says nothing about the origin of the agents' preferences, it assumes that each agent has a threshold for tolerating members of the other group in its neighborhood. For example, an agent might prefer not to be in the minority, or might require that at least a third of its neighbors be from the same group. Initially, agents are randomly distributed across the board and some squares are left unoccupied. Agents can observe their Moore neighborhood (the 8-cell combination of 4 adjacent cells and 4 diagonal cells), and can

change their location if the number of agents of the other type exceeds a certain threshold. When this happens, the dissatisfied agent randomly relocates to a new position on the board.

One of the striking results of this model is that even when agents are highly tolerant, segregation is still likely to emerge. Segregation arises due to the phenomenon of tipping, in which the first moves of even a few dissatisfied agents can create an incentive for others to move. This results in a cascade of movement that only ends when the whole board has become highly segregated. This is the core feature of the model. Agents' attempts to avoid being in a minority by moving to a new location change the composition of both old and new locations in a way that precipitates further movement. The neighborhood they leave becomes less attractive to members of their own group and the composition of the new neighborhood becomes less attractive for the members of the other group. Ultimately, over successive iterations, segregated neighborhoods emerge.

Now, many factors could cause real-world residential segregation, for example:

- ¥ strong discriminatory preferences (e.g. straightforward racism)
- ¥ economic disparity between groups (combined with price differences in housing)
- ¥ institutional discrimination (e.g. redlining, racial steering by real estate agents)
- ¥ structural features associated with group membership (e.g. the location of employment, channels of information flow about available housing).

The central contribution of Schelling's checkerboard model was to demonstrate that a mild discriminatory preference – for example, the wish to avoid being surrounded by too many members of the other group – can also create and sustain residential segregation. This finding has important policy implications. Intuitively one might think that the removal of the above causes of segregation would ultimately put an end to residential segregation, but Schelling's model shows

that this might not occur. Individuals' preferences might be incompatible with the integrated outcome. As Schelling's finding turned out to be quite robust with respect to changes in its background assumptions, achieving residential integration appears much more difficult than previously thought (Fossett 2006; Aydinonat 2008).

Schelling's simple and stylized model also teaches a more general theoretical lesson. It demonstrates that macro-level outcomes and relationships often tell us very little about their origins. Without the model, one could easily make two fallacious inferences. First, after observing the macro-level segregation, one might infer that a significant amount of prejudice must underlie it. Second, when being informed of the mild discriminatory preferences, one would probably expect that this would not lead to a significant amount of segregation. However, small and seemingly trivial differences at the micro-level can create huge differences at the macro level. Only by explicitly considering the individual-level attributes and the mechanisms of transformation can the crucial factors affecting macro-level outcomes be identified.

It is important to recognize the generality of the ideas underlying the checkerboard model. The reason it is regarded as insightful is not that it provides a convincing explanation of some empirical cases of residential segregation. The case of racial segregation is only an illustration of a more general idea. For this reason, it does not matter if 'mild discriminatory preferences' do not account for many real world examples of ethnic segregation. As Schelling (1978: Chapter 4) himself quite aptly shows, the model could be applied to other types of segregation such as that among friends at parties, or among guests at a wedding reception, etc. More generally, segregation can be defined as: "... the nonrandom allocation of people who belong to different groups into social positions and associated social and physical distances between groups" (Bruch

& Mare 2009: 270). What the checkerboard model provides is a template for thinking about the important subclass of segregation processes where the segregation is an unintended result of individual choices that influence the choice situations of other agents.

Schelling's model has turned out to be a fruitful template for *in silico* testing of theoretical hypotheses. Due to the flexibility of the ABS methodology, social scientists have been able to engage in an extensive exploration of theoretical possibilities. ABS models allow researchers to perform virtual experiments showing how specified sets of initial conditions and behavioral rules are likely to give rise to particular kinds of residential patterns. A look at the recent literature on the checkerboard model (e.g., Benard & Willer 2007; Bruch & Mare 2006; Clark & Fossett 2008; Fossett 2006) shows that the research has focused mainly on the exploration of series of *what-if* questions. ABS has made it possible for social scientists to ask what would happen if we:

- ¥ change the size of the neighborhood the agents observe
- ¥ change the number of groups (or their relative sizes)
- ¥ change the ratio of vacancies (or their accessibility)
- ¥ change the nonlinear choice function to a linear one
- ¥ allow different groups to have different preferences
- ¥ allow variation in preferences within the groups
- ¥ allow errors in individual choices (and observations)
- ¥ introduce other factors (e.g. wealth, social status)
- ¥ assume that some agents have a preference for integration.

These *what-if* questions represent modifications of the original idea presented by Schelling, and their point is to see how the changes influence the segregation outcomes and the working of the mechanism. One of the most interesting recent developments has been the incorporation of empirical information about people's preferences into simulations (see Bruch & Mare 2006; Clark & Fossett 2008). While much remains to be done to bridge simulation results with the empirical observations of residential segregation, these ABSs show some of the potential of simulation methodology.

4. Understanding and Simulation

How does ABS contribute to a scientific understanding of social phenomena? To answer this question, three things are required. First, we need an explication of the notion of scientific understanding. Second, the way in which computer simulation can contribute to understanding must be made intelligible. Finally, a response is needed for those who dismiss highly abstract *how-possibly* models – like Schelling's checkerboard model – as mere storytelling that provides little understanding of real-world phenomena. In other words, an account of the epistemic contribution of the very stylized *how-possibly* explanations is required. I will proceed accordingly in this order.

4.1 Scientific understanding

In my view, the best way to analyze the notion of understanding is inferential. According to this view, understanding is the ability to make *what-if* inferences about a phenomenon (Ylikoski 2009). In this view, understanding is a kind of ability that comes in degrees. When one understands something, one can make various correct inferences about it. When one understands

something better, one can make more – or more precise – correct *what-if* inferences about it. Accordingly, the degree of understanding conveyed by an explanation can be defined as the number of counterfactual inferences that the explanatory information makes possible (Ylikoski & Kuorikoski 2010). Of course, the mere counting of inferences is not the point. (There would also be the tricky problem of individuating the inferences.) It is much more important to be able to make relevant inferences about the things one is interested in. Thus the key question is whether a piece of information facilitates making inferences of interest about the phenomenon. This can be evaluated qualitatively. If the scope of these inferences expands, the understanding has also expanded.

Understanding is based on knowledge about dependencies. This knowledge provides an ability to predict and explain. A prediction is a straightforward case of *what-if* inference: it tells how a system behaves under specific circumstances. When one can forecast the behavior of the system, one has at least some understanding of it. Scientists are also interested in answering how- and why-questions about the phenomena they study. Answering these questions, which amounts to explanation – can be conceived in terms of *what-if* inferences. An answer to an explanation-seeking question tells us about the counterfactual dependence between the *explanans* and the *explanandum* – it specifies how the *explanandum* would have been different if the *explanans* had been different (Woodward 2003, Ylikoski & Kuorikoski 2010). In most cases, this presupposes knowledge of the *causal mechanisms* that govern the process (Hedström & Ylikoski 2010).

The inferential account of understanding does not conceive of understanding as any kind of special mental state. It is not a privately accessible sensation or experience, nor is it a state of

mind. The criteria for attributing understanding are public. To judge whether someone understands something, one does not look into her mind, but observes whether he or she can make relevant counterfactual inferences about it. Having an appropriate mental model might be a precondition of understanding, but unlike some accounts of understanding inspired by cognitive science (Waskan 2006), the inferential approach does not equate understanding with having the right kind of mental model (Ylikoski 2009; Kuorikoski 2011).

Having a behavioral and non-mentalistic notion of understanding has two clear advantages. The first is that it makes it possible to consider cases of extended and collective understanding. Recent work on cognition has increasingly emphasized the role of external cognitive tools (extended cognition) and the importance of the social division of epistemic labor (distributed cognition) in our cognitive lives. As these phenomena are particularly important in scientific cognition, it is an advantage if one's notion of understanding can accommodate these aspects of cognition rather than rendering them conceptually impossible. I will return to the idea of extended cognition shortly.

The second advantage of a non-mentalistic concept of understanding is that one can distinguish between actually understanding, and thinking that one understands (Ylikoski 2009; Kuorikoski 2011): one of the key ideas of the inferential account is the distinction between understanding and the *sense of understanding*. The latter is a feeling that tells us when we have understood or grasped something. This sense of confidence can be easily confused with the ability that it indicates. Ideally, understanding and a sense of understanding would go hand in hand. However, empirical studies show that a sense of understanding is a highly fallible indicator of understanding. People often overestimate the detail, coherence, and depth of their understanding

(Rozenblit and Keil 2002, Keil 2003). This illusion of depth of understanding can play an important role in science (Ylikoski 2009). As we will see, it also is a constant danger in social simulation research.

4.2 Understanding with computer simulation

The next challenge is to demonstrate how the use of computer simulation contributes to understanding. I will begin by showing how the inferential view can accommodate the use of external aids of understanding. Then I will articulate certain ways in which ABS enhances our understanding of social phenomena. I will conclude the section by considering a particularly important potential source of illusory understanding in simulation studies.

In the inferential view, it does not matter whether the model is a mental model, a physical object, a set of equations or a computer program. What matters is the ability to use it to make correct *what-if* inferences about the object of interest. As the inferential account is non-mentalist, it can easily accommodate computer simulation as a form of extended cognition (Kuorikoski 2011). Accordingly, we should regard the computer-scientist pair as the most relevant cognitive system. In this view, figuring out the extra understanding provided by the computer simulation is easy. We just compare which correct *what-if* inferences the scientists are able to make with and without the computer simulation. The difference is the contribution the simulation makes. Similarly, we can study whether the use of computer simulation creates illusory understanding by determining whether it increases (or decreases) the number of false beliefs about our inferential abilities.

An ABS can enhance understanding in at least three ways. First, building the simulation forces theorists to be more explicit about their assumptions. Second, the use of a computer makes the

inferences more reliable. Third, the computer simulation expands the number of *what-if* inferences that one is able to make. Let us look at these enhancements in more detail.

Making the assumptions explicit. In contrast to verbal theorizing, computational modeling does not allow implicit assumptions within a model (Miller & Page 2007; Macy & Flache 2009). Everything that goes into the model must be explicit, which makes it easier to evaluate the assumptions from the point of view of consistency, realism, and relevance. However, it is important to recognize that building an agent-based simulation does not completely do away with implicit assumptions. The assumptions that guide the inferences from the simulation to the target system are not part of the simulation, so they might remain implicit. For example, the simulation model of residential segregation might not include the activities of housing agencies, while in the real world they may play an important role. However, the (implicit or explicit) assumptions about the effects of housing agencies will guide the inferences that the theorist makes from the simulation to the real-world housing patterns.

It is also important to note that increasingly the simulators do not build their simulations from the ground up. Rather, they employ ready-made software libraries or modify existing simulations built by others. Under these circumstances it is possible that the scientist may not be familiar with all of the assumptions underlying his or her simulation. For this reason, it is important to distinguish between merely using a simulation and fully understanding how it works. When a simulation is used merely as an inferential aid, the user has a very limited understanding of the underlying mechanisms. So the scientist is not able to infer what would happen if some parts of the simulation were changed, nor would the scientist be able to determine what is the proper

domain of the simulation's application as he or she would not know what assumptions the simulation has been making about the simulated phenomenon.

The increased reliability of inference. Both psychological science and everyday experience show that unaided human reasoning is not very reliable. We are prone to limitations in our working memory, various biases, and, of course, errors. As well, the frequency in which the results of simulations are found surprising and unintuitive demonstrates the fallibility of our reasoning abilities. By externalizing some of the inferences to the computer, we can increase their reliability.

Of course, computerization is not a magic bullet for reliable inferences. While we can almost always trust that the machine executes its code correctly, simulations have other sources of unreliability as inferential aids. Bugs in the code, idiosyncratic programming habits, difficulties in comparing simulations built in different programming languages (or platforms), and a lack of sufficient commentary all present challenges to simulation-aided scientific inferences. For these reasons agent-based simulation in the social sciences has a long way to go before it can realize its full potential. In this respect, the simulation community must reach some level of standardization in programming languages and software packages. It also needs conventions both for reporting results and for commenting on and sharing code.

Computer simulation expands our ability to understand the world, but it does not make it easy. It is possible to distinguish between two separate challenges in understanding something with the assistance of ABS. The first is to understand the simulation itself. One must understand how the changes in simulation outcomes depend on changes in its assumptions. One also must know how the simulation behaves under various circumstances; in other words, one must know the tipping

points and areas of robust results, and their respective background conditions. Finally, one must be able to recognize and track programming errors. Only when one more fully understands the simulation, can one use it as a tool for understanding real-world phenomena.

However, knowing the tool is not enough. The second challenge is to have enough background knowledge to know which aspects of the phenomenon are comparable to the simulation and what kinds of evidence would justify these comparisons. One should also be able to judge what types of inferences the simulation allows about the target. Are quantitative inferences possible, or are the results only qualitative? What are the margins of error for the predictions? Can one make inferences about singular cases or only about the distributions of cases?

The confusion concerning these two challenges is a dangerous source of illusion of understanding in simulation studies. Building the simulation from the bottom up requires much effort, and it easily creates an intuitive feeling that one truly understands how it works. This feeling of understanding can be easily misattributed to the phenomenon that is the target of the simulation: one *intends* to understand the phenomenon, and the *activity produces understanding*, so one unreflectively infers that one has understood the phenomenon. Of course, the correctness of this inference is contingent upon the assumption that one knows enough to meet the second challenge as well.

Expanding the inferential scope. As suggested above, the use of computer simulation removes some of the restrictions that our verbal reasoning skills and available analytical methods place on our inferential ability. With an ABS model we can systematically study how the changes in the assumptions change the outcomes. In the case of social sciences, ABS methodology has enabled a whole new way of studying micro-macro relations: it is now possible to explore the kinds of

consequences that various micro-level changes have on macro processes. There are some cases where the reach of more traditional analytical methods and ABS overlap, but it is clear that agent-based methodology has expanded the scope of theoretical inferences in the social sciences. While it sometimes makes sense to discuss which modeling approach provides a better understanding (Lehtinen & Kuorikoski 2007), it is a mistake to treat them as exclusive alternatives. The use of ABS in the social sciences is in its early stages, so we cannot say with any certainty to what extent it could expand the scope of theoretically relevant *what-if* inferences; but given that it removes some of the theoretically relevant restrictions on analytical modeling, it is to be expected that the expansion of the inferential scope will be substantial.

4.3 The storytelling challenge

The highly abstract ABS models – such as Schelling’s checkerboard model – provide *how-possibly* explanations. These models are sometimes dismissively characterized as toy models that have little to contribute to an understanding of real-world phenomena. This criticism is misguided. While ABS is no replacement for empirical case studies – or any other traditional method of social research – it can play an important role in developing an explanatory understanding of complex real-world phenomena.

It is important to recognize that most of the ABS models are not intended to be explanations of particular empirical facts. Their purpose is to specify a causal mechanism scheme that is intended to capture the crucial features of the true causal scenario that could in principle explain the empirically observable facts. Confusing such a mechanism scheme with a simplistic causal hypothesis is the most common cause of a dismissive attitude. Once this fallacy is recognized, we can begin to think about the real contribution of these causal mechanism schemes.

The epistemic importance of *how-possibly* causal scenarios is based on the fact that the search for causal explanations often has the structure of an inference to the best – or, more precisely, to the remaining – explanation. Explanatory hypotheses are not confirmed only by finding supporting evidence, but also by ruling out competing hypotheses. The alternative causal scenarios play an important role in this process. Reducing the number of possible causal scenarios presupposes well-specified causal scenarios that may inform us of the traces of their working that should be found in the evidence. For this reason, the causal scenarios guide the search for the new evidence that discriminates between the competing scenarios. (Ylikoski & Aydinonat 2011.)

This process of scenario culling assumes that the alternative causal scenarios are known. A demonstration that unexplored, but viable, alternatives exists is quite damaging to the inference to the best explanation. This makes the knowledge about the range of possible explanations crucial: we can only be confident in having the right explanation if we have good reasons to believe that our list of alternatives is comprehensive. As a consequence, any contribution that revises *the menu of possible explanations* is of crucial epistemic value. This observation helps us to understand why Schelling's checkerboard model has made an important contribution to the explanatory understanding of residential segregation: its mere introduction changed the evidential landscape against which the explanatory hypotheses are evaluated (Ylikoski & Aydinonat 2011).

The introduction of a new causal scenario (or the elimination of an old one) changes the evidential requirements set for accepting one causal scenario as the correct one. A change in the menu of possible explanations brings with it a new set of challenges that the winning scenario

must overcome because the purpose of the evidence is to discriminate between the alternatives. On the one hand, the demand for evidence also arises: in order to justify a claim about actual causal contribution one has to rule out one more competing explanation. This forces researchers to search for more stringent evidence. This is clear in the case of Schelling's checkerboard: its main contribution in the studies of segregation has not been in the area of explaining particular phenomena but in the requirements for a satisfactory explanation. After Schelling, it can be expected that social scientists are not satisfied with purely verbal explanations of segregation. Now they can require a detailed dynamical model with explicitly articulated underlying mechanisms – even when the actual explanatory mechanism is different from the one Schelling suggested (Ylikoski & Aydinonat 2011).

For these reasons it is a mistake to regard abstract *how-possibly* explanations as idle storytelling. They refine, systematize and expand the menu of possible explanations, and this is the basis of their epistemic value. Rather than providing an understanding of particular empirical facts, they provide modal understanding: they put forward how things could be. Similarly, they also make clear what cannot happen by showing that certain mechanisms cannot generate the specified outcomes. Further, they help to articulate the essential elements and preconditions of causal mechanism schemes. This makes it easier to adapt them to particular causal scenarios, thus extending our ability to explain particular cases. These contributions explain (and at least partially justify) the modelers' feeling that they are contributing to scientific understanding by creating and examining abstract ABS models.

5. Conclusion

In this article I have described the use of ABS models in sociology as a tool for theoretical exploration. When analyzing their contribution to sociological understanding, I made use of the inferential account of understanding (Ylikoski 2009, Ylikoski & Kuorikoski 2010, Kuorikoski 2011). In this account, the use of ABS models can be understood as extended cognition that enhances sociological understanding in three ways. First, the use of ABS methodology forces theoreticians to make their assumptions explicit, which may increase the consistency, realism and relevance of the theoretical model. Second, the use of ABS methods can increase the reliability of the *what-if* inferences that their users are making. Thirdly, and most importantly, the ABS models increase the scope of theoretically relevant inferences that social scientists are able to make.

While this paper has focused mostly on presenting the positive contribution that ABS methodology can make to social sciences, I have also pointed out some challenges and possible limitations of it. First, if the simulator is employing ready-made simulation components, he or she might not recognize all of the assumptions included in the model. Second, the reliability of ABS-aided theoretical inferences is challenged by such things as bugs in the code, a lack of commentary or documentation, and idiosyncratic programming habits, which all make the comparison and interpretation of simulations difficult. Third, the experience of building and using a simulation can give rise to an illusion of understanding where the understanding of the simulation is confused with understanding the simulated phenomenon. Finally, a better understanding of the epistemic contribution of theoretical ABS models presupposes that we understand the nature of scientific *how-possibly* reasoning (Ylikoski & Aydinonat 2011); otherwise these simulations are too easily dismissed as abstract forms of idle storytelling.

References

- Aydinonat, N. Emrah. 2008. *The Invisible Hand in Economics: How Economists Explain Unintended Social Consequences*. London: Routledge.
- Benard, Stephen and Robb Willer. 2007. "A Wealth and Status-Based Model of Residential Segregation." *Journal of Mathematical Sociology* 31: 149-174.
- Bruch, Elizabeth and Robert Mare. 2006. "Neighborhood Choice and Neighborhood Change." *American Journal of Sociology* 112: 667-709.
- Bruch, Elizabeth and Robert Mare. 2009. "Segregation Dynamics." Pp 269-293 in *The Oxford Handbook of Analytical Sociology*. Edited by Peter Hedström and Peter Bearman. Oxford: Oxford University Press.
- Clark, William AV and Mark Fossett. 2008: "Understanding the social context of the Schelling segregation model." *PNAS* 105 (11): 4109-4114.
- Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: The Belknap Press.
- Epstein, Joshua M. 2006. *Generative Social Science. Studies in Agent-Based Computational Modeling*. Princeton: Princeton University Press.
- Fossett, Mark. 2006. "Ethnic Preferences, Social Distance Dynamics, and Residential Segregation: Theoretical Explorations Using Simulation Analysis." *Journal of Mathematical Sociology* 30: 185-274.
- Hedström, Peter. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
- Hedström, Peter and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36: 49-67.
- Kuorikoski, Jaakko. 2011. "Simulation and the Sense of Understanding." Pp. 168-187 in *Models, Simulations and Representations*. Edited by Paul Humphreys and Cyrille Imbert, London: Routledge.
- Lehtinen, Aki and Jaakko Kuorikoski. 2007. "Computing the Perfect Model: Why Do Economists Shun Simulation?" *Philosophy of Science* 74: 304-329.

- Macy, Michael and Andreas Flache. 2009. "Social Dynamics from the Bottom Up. Agent-Based Models of Social Interaction." Pp 245-268 in *The Oxford Handbook of Analytical Sociology*. Edited by Peter Hedström and Peter Bearman. Oxford: Oxford University Press.
- Miller, John H. and Scott E. Page. 2007. *Complex Adaptive Systems. An Introduction to Computational Models of Social Life*. Princeton: Princeton University Press.
- Pancs, Romans and Nicolaas J. Vriend. 2007. "Schelling's spatial proximity model of residential segregation revisited." *Journal of Public Econometrics* 91: 1-24.
- Schelling, Thomas C. 1971. "Dynamic models of segregation." *Journal of Mathematical Sociology* 1: 143-86.
- Schelling, Thomas C. 1978. *Micromotives and Macrobehavior*. London & New York: W. W. Norton.
- Waskan, Jonathan. 2006. *Models and cognition*. Cambridge, MA: Massachusetts Institute of Technology Press.
- Woodward, James. 2003. *Making things happen: A theory of causal explanation*. New York: Oxford University Press.
- Zhang, Junfu. 2004. "A Dynamical Model of Residential Segregation." *Journal of Mathematical Sociology* 28: 147-170.
- Ylikoski, Petri. 2009. "The Illusion of Depth of Understanding in Science." Pp 100-119 in *Scientific Understanding: Philosophical Perspectives*. Edited by H. De Regt, S. Leonelli, and K. Eigner. Pittsburgh: Pittsburgh University Press.
- Ylikoski, Petri and Emrah Aydinonat 2011. "Understanding with Theoretical Models." manuscript under review.
- Ylikoski, Petri and Jaakko Kuorikoski. 2010. "Dissecting Explanatory Power." *Philosophical Studies* 148: 201–219.