What's wrong with social simulations?

Eckhart Arnold

Institute for Philosophy

Heinrich Heine University of Düsseldorf

September 2013; last revision: March 2016

Abstrakt

This paper tries to answer the question why the epistemic value of so many social simulations is questionable. I consider the epistemic value of a social simulation as questionable if it contributes neither directly nor indirectly to the understanding of empirical reality. In order to justify this allegation I rely mostly but not entirely on the survey by Heath, Hill and Ciarallo (2009) according to which 2/3 of all agent-based-simulations are not properly empirically validated. In order to understand the reasons why so many social simulations are of questionable epistemic value, two classical social simulations are analyzed with respect to their possible epistemic justification: Schelling's neighborhood segregation model (Schelling, 1971) and Axelrod's reiterated Prisoner's Dilemma simulations of the evolution of cooperation (Axelrod, 1984). It is argued that Schelling's simulation is useful, because it can be related to empirical reality, while Axelrod's simulations and those of his followers cannot be related to empirical reality and therefore their scientific value remains doubtful. Finally, I critically discuss some of the typical epistemological background beliefs of modelers as expressed in Joshua Epsteins's keynote address "Why model?" (Epstein, 2008). Underestimating the importance of empirical validation is identified as one major cause of failure for social simulations.

Table of contents

1   Introduction

In this paper I will try to answer the question: Why is the epistemic value of so many social simulations questionable? Under social simulations I understand computer simulations of human interaction as it is studied in the social sciences. The reason why I consider the epistemic value of many social simulations as questionable is that many simulation studies cannot give an answer to the most salient question that any scientific study should be ready to answer: "How do we know it's true?" or, if specifically directed to simulation studies: "How do we know that the simulation simulates the phenomenon correctly that it simulates?" Answering this question requires some kind of empirical validation of the simulation. The requirement of empirical validation is in line with the widely accepted notion that science is demarcated from non-science by its empirical testability or falsifiability. Many simulation studies, however, do not offer any suggestion how they could possibly be validated empirically.

A frequent reply by simulation scientists is that no simulation of empirical phenomena was intended, but that the simulation only serves a "theoretical" purpose. Then, however, another equally salient question should be answered: "Why should we care about the results?" It is my strong impression that many social simulation studies cannot answer either this or the first question. This is not to say that the use of computer programs for answering purely theoretical questions is generally or

necessarily devoid of value. The computer assisted proofs of the four color theorem (Wilson, 2002) are an important counterexample. But in the social sciences it is hard to find similarly useful examples of the use of computers for purely theoretical purposes. In any case, the social sciences are empirical sciences. Therefore, social simulations should contribute either directly or indirectly to our understanding of social phenomena in the empirical world.

There exist many different types of simulations but I will restrict myself to agent-based and game theoretical simulations. I do not make a sharp difference between models and simulations. For the purpose of this paper I identify computer simulations just with programmed models. Most of my criticism of the practice of these simulation types can probably be generalized to other types of simulations or models in the social sciences and maybe also to some instances of the simulation practice in the natural sciences. It would lead too far afield to examine these connections here, but it should be easy to determine in other cases whether the particulars of bad simulation practice against which my criticism is directed are present or not.

In order to bring my point home, I rely on the survey by Heath, Hill and Ciarallo (2009) on agent-based modeling practice for a general overview and on two example cases that I examine in detail. I start by discussing the survey which reveals that in an important sub-field of social simulations, namely, agent based simulations, empirical validation is commonly lacking. After that I first discuss Thomas Schelling's well-known neighborhood segregation model. This is a model that I do not consider as being devoid of epistemic value. For, unlike most social simulations, it can be empirically falsified. The discussion of the particular features that make this model scientifically valuable will help us to understand why the simulation models discussed in the following fail to be so.

The simulation models that I discuss in the following are simulations in the tradition of Robert Axelrod's "Evolution of Cooperation" (Axelrod, 1984). Although the modeling tradition initiated by Axelrod has delivered hardly any tenable and empirically applicable results, it still continues to thrive today. By some, Axelrod's approach is still taken as a role model (Rendell et al., 2010, 208-209), although there has been severe criticism by others (Arnold, 2008; Binmore, 1994, 1998).

Finally, the question remains why scientists continue to produce such an abun-

dance of simulation studies that fail to be empirically applicable. Leaving possible sociological explanations like the momentum of scientific traditions, the cohesion of peer groups, the necessity of justifying the investment in acquiring particular skills (e.g. math and programming) aside, I confine myself to the ideological background of simulation scientists. In my opinion the failure to produce useful results has a lot to do with the positivist attitude prevailing in this field of the social sciences. This attitude includes the dogmatic belief in the superiority of the methods of natural sciences like physics in any area of science. Therefore, despite frequent failure, many scientists continue to believe that formal modeling is just the right method for the social sciences. The attitude is well described in Shapiro (2005). Such attitudes are less often expressed explicitly in the scientific papers. Rather they form a background of shared convictions that, if not simply taken for granted as "unspoken assumptions", find their expression in informal texts, conversations, blogs, keynote speeches. I discuss Joshua Epstein's keynote lecture "Why Model?" (Epstein, 2008) as an example.

2   Simulation without validation in agent-based models

In this section I give my interpretation of a survey by Heath, Hill and Ciarallo (2009) on agent-based-simulations. I do so with the intention of substantiating my claim that many social simulations are indeed useless. This is neither the aim nor the precise conclusion that Heath, Hill and Ciarallo (2009) draw, but their study does reveal that two thirds of the surveyed simulation studies are not completely validated and the authors of the study consider this state of affairs as "not acceptable" (Heath, Hill and Ciarallo, 2009, 4.11). Thus my reading does not run counter the results of the survey. And it follows as a natural conclusion, if one accepts that a) an unvalidated simulation is - in most of the cases - a useless one and b) agent-based simulations make up a substantial part of social simulations.

The survey by Heath, Hill and Ciarallo (2009) examines agent-based modeling practices between 1998 and 2008. It encompasses "279 articles from 92 unique publication outlets in which the authors had constructed and analyzed an agent-

based model" (Heath, Hill and Ciarallo, 2009, abstract). The articles stem from different fields of the social sciences including, business, economics, public policy, social science, traffic, military and also biology. The authors are not only interested in verification and validation practices, but the results concerning these are the results that I am interested in here. Verification and validation concern two separate aspects of securing the correctness of a simulation model. Verification, as the term is used in the social simualtions community, roughly concerns the question whether the simulation software is bug-free and correctly implements the intended simulation model. Validation concerns the question whether the simulation model represents the simulated empirical target system adequately (for the intended purpose).

Regarding verification, Heath, Hill and Ciarallo notice that "Only 44 (15.8%) of the articles surveyed gave a reference for the reader to access or replicate the model. This indicates that the majority of the authors, publication outlets and reviewers did not deem it necessary to allow independent access to the models. This trend appears consistently over the last 10 years" (Heath, Hill and Ciarallo, 2009, 3.6). This astonishingly low figure can in part be explained by the fact that as long as the model is described with sufficient detail in the paper, it can also be replicated by re-programming it from the model description. It must not be forgotten that the replication of computer simulation results does not have the same epistemological importance as the replication of experimental results. While the replication of experiments adds additional inductive support to the experimental results, the replication of simulation results is merely a means for checking the simulation software for programming errors ("bugs"). Hence the possibility of precise replication is not an advantage that simulations enjoy over material experiments, as for example Reiss (2011, 248) argues. Obviously, if the same simulation software is run in the same system environment the same results will be produced, no matter whether this is done by a different team of researchers at a different time and place with different computers. Even if the model is re-implemented the results must necessarily be the same provided that both the model and the system environment are fully specified and no programming errors have been made in the original implementation or the re-implementation.[1] Replication

---

[1] A possible exception concerns the frequent use of random numbers. As long as only pseudo

or reimplementation can, however, help to reveal such errors.[2] It can therefore be considered as one of several possible means for the verification (but not validation) of a computer simulation. Error detection becomes much more laborious if no reference to the source code is provided. And it does happen that simulation models are not specified with sufficient detail to replicate them (Will and Hegselmann, 2008). Therefore, the rather low proportion of articles that provide a reference to access or replicate the simulation is worrisome.

More important than the results concerning verification is what Heath, Hill and Ciarallo find out about validation or, rather, the lack of validation:

> Without validation a model cannot be said to be representative of anything real. However, 65% of the surveyed articles were not completely validated. This is a practice that is not acceptable in other sciences and should no longer be acceptable in ABM practice and in publications associated with ABM. (Heath, Hill and Ciarallo, 2009, 4.11)

This conclusion needs a little further commentary. The figure of 65% of not completely validated simulations is an average value over the whole period of study. In the earlier years that are covered by the survey hardly any simulation was completely validated. Later this figure decreases, but a ratio of less than 45% of completely validated simulation studies remains constant during the last 4 yours of the period covered (Heath, Hill and Ciarallo, 2009, 3.10).

Furthermore it needs to be qualified what Heath, Hill and Ciarallo mean when they speak of complete validation. The authors make a distinction between conceptual validation and operational validation. Conceptual validation concerns the question whether the mechanisms built into the model represent the mechanisms that drive the modeled real system. An "invalid conceptual model indicates the model may not be an appropriate representation of reality." Operational validation then "validates results of the simulation against results from the real system." (Heath, Hill and Ciarallo, 2009, 2.13). The demand for complete validation is well

---

random numbers with the same random number generator and the same "seed" are used, the simulation is still completely deterministic. This not to say that sticking to the same "seeds" is good practice other than for debugging.

[2]I am indebted to Paul Humphreys for pointing this out to me.

motivated: "If a model is only conceptually validated, then it [is] unknown if that model will produce correct output results." (Heath, Hill and Ciarallo, 2009, 4.12). For even if the driving mechanisms of the real system are represented in the model, it remains – without operational validation – unclear whether the representation is good enough to produce correct output results. On the other hand, a model that has been operationally validated only, may be based on a false or unrealistic mechanism and thus fail to explain the simulated phenomenon, even if the data matches. Heath, Hill and Ciarallo do not go into much detail concerning how exactly conceptual and operational validation are done in practice and under what conditions a validation attempt is to be considered as successful or as a failure.

But do really all simulations need to be validated both conceptually and operationally as Heath, Hill and Ciarallo demand? After all, some simulations may – just like thought experiments – have been intended to merely prove conceptual possibilities. One would usually not demand an empirical (i.e. operational) validation from a thought experiment. Heath, Hill and Ciarallo themselves make a distinction between the generator, mediator and predictor role of a simulation (Heath, Hill and Ciarallo, 2009, 2.16). In the generator role simulations are merely meant to generate hypotheses. Simulations in the mediator role "capture certain behaviors of the system and [..] characterize how the system may behave under certain scenarios" (3.4) and only simulations in the predictor role are actually calculating a real system. All of the surveyed studies fall into the first two categories. Obviously, the authors require complete validation even from these types of simulations.

This can be disputed. As stated in the introduction, in order to be useful, a simulation study should make a contribution to answering some relevant question of empirical science. This contribution can be direct or indirect. The contribution is direct if the model can be applied to some empirical process and if it can be tested empirically whether the model is correct. The model's contribution is indirect, if the model cannot be applied empirically, but if we can learn something from the model which helps us to answer an empirical question, the answer to which we would not have known otherwise. The latter kind of simulations can be said to function as thought experiments. It would be asking too much to demand complete empirical validation from a thought experiment.

But does this mean that the figures from Heath, Hill and Ciarallo concerning the

validation of simulations need to be interpreted differently by taking into account that some simulations may not require complete validation in the first place? This objection would miss the point, because the scenario just discussed is the exception rather than the rule. Classical thought experiments like Schrödinger's cat usually touch upon important theoretical disputes. However, as will become apparent from the discussion of simulations of the evolution of cooperation, below, computer simulation studies all too easily lose the contact to relevant scientific questions. We just do not need all those digital thought experiments on conceivable variants of one and the same game theoretical model of cooperation. And the same surely applies to many other traditions of social modeling as well. But if this is true, then the figure of 65% of not completely validated simulation studies in the field of agent-based simulations is alarming indeed.[3]

Given how important empirical validation is, "because it is the only means that provides some evidence that a model can be used for a particular purpose." (Heath, Hill and Ciarallo, 2009, 4.11), it is surprising how little discussion this important topic finds in the textbook literature on social simulations. Gilbert and Troitzsch (2005) mention validation as an important part of the activity of conducting computer simulations in the social sciences, but then they dedicate only a few pages to it (22-25). Šalamon (2011, 98) also mentions it as an important question without giving any satisfactory answer to this question and without providing readers with so much as a hint concerning how simulations must be constructed so that their validity can be empirically tested. Railsback and Grimm (2011) dedicate many pages to describing the ODD-protocol, a protocol that is meant to standardize agent-based simulations and thus to facilitate the construction, comparison and evaluation of agent-based simulations. Arguably the most important topic, empirical validation of agent-based simulations, is not an explicit part of this protocol. One could argue that this is simply a different matter, but then, given the importance of this topic it is slightly disappointing that Railsback and Grimm do not treat it more explicitly in their book.

Summing it up, the survey by Heath, Hill and Ciarallo shows that an increasin-

---

[3]For a detailed discussion of the cases in which even unvalidated simulations can be considered as useful, see Arnold (2013). There are such cases, but the conditions under which this is possible appear to be quite restrictive.

gly important sub-discipline of social simulations, namely the field of agent-based simulations faces the serious problem that a large part of its scientific literature consists of unvalidated and therefore most probably useless computer simulations. Moreover, considering the textbook literature on agent-based simulations one can get the impression that the scientific community is not at all sufficiently aware of this problem.

3 How a model works that works: Schelling's neighborhood segregation model

Moving from the general finding to particular examples, I now turn to the discussion of Thomas Schelling's neighborhood segregation model. Schelling's neighborhood segregation model (Schelling, 1971) is widely known and has been amply discussed not only among economists but also among philosophers of science as a role model for linking micro-motifs with macro-outcomes. I will therefore say little about the model itself, but concentrate on the questions if and, if so, how it fulfills my criteria for epistemically valuable simulations.

Schelling's model was meant to investigate the role of individual choice in bringing about the segregation of neighborhoods that are either predominantly inhabited by blacks or by whites. Schelling considered the role of preference based individual choice as one of many possible causes of this phenomenon – and probably not even the most important, at least not in comparison to organized action and economic factors as two other possible causes (Schelling, 1971, 144).

In order to investigate the phenomenon, Schelling used a checkerboard model where the fields of the checkerboard would represent houses. The skin color of the inhabitants can be represented for example by pennies that are turned either heads or tails.[4] Schelling assumed a certain tolerance threshold concerning the number of differently colored inhabitants in the neighborhood, before a household would move to another place. A result that was relatively stable among the different variants of the model he examined was that segregated neighborhoods would emerge –

---

[4]Schelling's article was published before personal computers existed. Today one would of course use a computer. A simple version of Schelling's model can be found in the netlogo models library (Wilensky, 1999).

even if the threshold preference for equally colored neighbors was far below 50%, which means that segregation emerged even if the inhabitants would have been perfectly happy to live in an integrated environment with a mixed population. As Aydinonat (2007) reports, the robustness of this result has been confirmed by many subsequent studies that employed variants of Schelling's model. At the end of his paper Schelling discusses "tipping" that occurs when the entrance of a new minority starts to cause the evacuation of an area by its former inhabitants. In this connection Schelling also mentions an alternative hypothesis according to which inhabitants do not react to the frequency of similar or differently colored neighbors but on their on expectation about the future ratio of differently colored inhabitants. He assumes that this would aggravate the segregation process, but he does not investigate this hypothesis further (Schelling, 1971, 185-186) and his model is built on the assumption that individuals react to the actual and not the future ratio of skin colors.

Is this model scientifically valuable? Can we draw conclusions from this model with respect to empirical reality and can we check whether these conclusions are true? Concerning these questions the following features of this model are important:

1. The assumptions on which the model rests can be tested empirically. The most important assumption is that individuals have a threshold for how many neighbors of a different color they tolerate and that they move to another neighborhood if this threshold is passed. This assumption can be tested empirically with the usual methods of empirical social research (and, of course, within the confinements of these methods). Also, the question whether people base their decision to move on the frequency of differently colored neighbors or on their on expectation concerning future changes of the neighborhood can be tested empirically.

2. The model is highly robust. Changes of the basic setting and even fairly large variations of its input parameters, e.g. tolerance threshold, population size, do not lead to a significantly different outcome. Therefore even if the empirical measurement of, say, the tolerance threshold, is inaccurate, the model can still be applied. Robustness in this sense is directly linked to empirical testability. It should best be understood as a relational property

between the measurement (in-)accuracy of the input parameters and the stability of the output values of a simulation.[5]

3. The model captures only one of many possible causes of neighborhood segregation. Before one can claim that the model explains or, rather, contributes to an explanation of neighborhood segregation, it is necessary to identify the modeled mechanism empirically and to estimate its relative weight in comparison with other actual causes. While the model shows that even a preference for integrated neighborhoods (if still combined with a tolerance limit) can lead to segregation, it may in reality still be the case that latent or manifest racism causes segregation. The model alone is not an explanation. (Schelling was aware of this.)

4. Besides empirical explanation another possible use of the model would be policy advice. In this respect the model could be useful even if it does not capture an actual cause. For public policy must also be concerned about possible future causes.

   Assume for example, that manifest racism was a cause of neighborhood segregation, but that due to increasing public awareness racism is on the decline. Then the model can demonstrate that even if all further possible causes, e.g. economic causes, be removed as well, this might still not result in desegregated neighborhoods[6] - provided, of course, that the basic assumption about a tolerance threshold is true.

   Thus, for the purpose of policy advice a model does not need to capture actual causes. It can be counter-factual, but it must still be realistic in the

---

[5]There are of course different concepts of robustness. I consider this relational concept of robustness as the most important concept. An important non-relational concept of robustness is that of derivational robustness analysis (Kuorikoski and Lehtinen, 2009). See below.

[6]But then, would we really worry about segregated neighborhoods, if the issue wasn't tied to racial discrimination and social injustice? After all, ethnic or religious groups in Canada also often live in segregated areas ("Canadian mosaic"). But other than in the U.S. this is hardly an issue. Therefore, Schelling's model – for all its epistemological merits that are discussed here – really seems to miss the point in terms of scientific relevance. Discrimination is the important point here, not segregation. But Schelling's model induces us to frame the question in a way that makes us miss the point. (*This comment has been added later as the result of some discussions I had on this point. E.A., March 25th 2016.*)

sense that its basic assumptions can be empirically validated. Therefore, while the purpose of policy advice justifies certain counter-factual assumptions in a model, it cannot justify unrealistic and unvalidated models. This generally holds for models that are meant to describe possible instead of actual scenarios.

Schelling did not validate his model empirically. But for classifying the model as useful it is sufficient that it can be validated. Now, the interesting question is: Can the model be validated and is it valid? Recent empirical research on the topic of neighborhood segregation suggests that inhabitants react to anticipated future changes in the frequency of differently colored neighbors rather than the frequency itself (Ellen, 2000, 124-125). An important role is played by the fear of whites that they might end up in an all-black neighborhood. Thus, the basic assumption of the model that individuals react upon the ratio of differently colored inhabitants in their neighborhood is wrong and one can say that the model is in this sense falsified.[7]

The strong emphasis that is placed on empirical validation here stands in contrast to some of the epistemological literature on simulations and models. Robert Sugden, noticing that "authors typically say very little about how their models relate to the real world", treats models like that of Schelling (which is one of his examples (Sugden, 2000, 6-8)) as "credible counterfactual worlds" (Sugden, 2009, 3) which are not intended to raise any particular empirical claims. Even though the particular relation to the real world is not clear, Sugden believes that such models can inform us about the real world. His account suffers from the fact that he remains unclear about how we can tell a counter-factual world that is credible from one that is incredible, if there is no empirical validation.

A possible candidate for stepping in this gap of Sugden's account is Kuorikoski's and Lehtinen's concept of "derivational robustness analysis" (Kuorikoski and Lehtinen, 2009). According to this concept conclusions from unrealistic mo-

---

[7]There are two senses in which a model (or more precisely: a model-based explanation) can be falsified: a) if the model's assumptions are empirically not valid as in this case and b) if the causes the model captures are (i) either blocked by factors not taken into account in the model or (ii) cannot be disentangled from other possible causes or (iii) turn out to be irrelevant in comparison with other, stronger or otherwise more important causes for the same phenomenon. The connection between the model's assumptions and its output, being a logical one, can, of course, not be empirically falsified.

dels to reality might be vindicated if the model remains robust under variations of its unrealistic assumptions. For example, in Schelling's model the checkerboard topography could be replaced by other different topographies (Aydinonat, 2007, 441). If the model still yields the same results about segregation, we are – if we follow the idea of "derivational robustness analysis" – entitled to draw the inductive conclusion that the model's results would still be the same if the unrealistic topographies were exchanged by the topography of some real city, even though we have not tested it with a real topography. A problem with this account is that it requires an inductive leap of a potentially dangerous kind: How can we be sure that the inductive conclusion derived from varying unrealistic assumptions holds for the conditions in reality which differ from any of these assumptions?

Some philosophers also dwell on the analogy between simulations and experiments and consider simulations as "isolating devices" similar to experiments (Mäki, 2009). But the analogy between simulations and experiments is rather fragile, because other than experiments simulations are not empirical and do not allow us to learn anything about the world apart from what is implied in the premises of the simulation. In particular, we can – without some kind of empirical validation – never be sure whether the causal mechanism modeled in the simulation represents a real cause isolated in the model or does not exist in reality at all.

Summing it up, it is difficult, if not impossible, to claim that models can inform us about reality without any kind of empirical validation. Schelling's model, however, appears to be a scientifically useful model, at least in the sense that it can be validated (or falsified for that matter). The most decisive features of the model in this respect are its robustness and the practical feasibility of identifying the modeled cause in empirical reality. Next we will see how models fare when these features are not present.

4   How models fail: The Reiterated Prisoner's Dilemma model

Robert Axelrod's computer simulations of the Reiterated Prisoner's Dilemma (RPD) (Axelrod, 1984) are well known and still considered by some as a role model

for successful simulation research (Rendell et al., 2010, 408-409). What is not so widely known is that the simulation research tradition initiated by Axelrod has remained entirely unsuccessful in terms of generating explanations for empirical instances of cooperation. What are the reasons for this lack of explanatory success? And how come that Axelrod's research design is none the less considered as a role model today?

Axelrod had the ingenious idea to advertise a public computer tournament where participation was open to everybody. Participants were asked to hand in their guess at a best strategy in the reiterated two person Prisoner's Dilemma in the form of an algorithmic description or computer program. This provided Axelrod with a rich, though naturally very contingent set of diverse strategies and it had the, surely welcome, side-effect of generating attention for Axelrod's research project. Axelrod ran a sequence of two tournaments. As is well known the rather simplistic strategy *Tit For Tat* won both tournaments.

In the Prisoner's Dilemma Game the players can decide whether to cooperate or not to cooperate. Mutual cooperation yields a higher payoff than mutual non-cooperation, but it is best to cheat by letting the other player cooperate while not cooperating oneself. And it is worst to be cheated, i.e. to cooperate while the other player does not. *Tit For Tat* cooperates in the first round of the Repeated Prisoner's Dilemma, but if the other player cheats, then *Tit For Tat* will punish the other player by not cooperating in the following round.[8] Axelrod analyzed the course of the tournament in order to understand just why *Tit For Tat* was such a successful strategy. He concluded that it is a number of characteristics that determine the success of a strategy in the Reiterated Prisoner's Dilemma (Axelrod, 1984, chapter 6): Successful strategies are (1) "friendly", i.e. they start with cooperative moves, (2) envy-free, (3) punishing, but also (4) forgiving. Axelrod furthermore believed that repeated interaction is a necessary requirement for cooperation to evolve and that, of course, *Tit For Tat* is generally quite a good strategy in Reiterated Prisoner's Dilemma situations.

Unfortunately for Axelrod, the Reiterated Prisoner's Dilemma model is anything but robust. For each of his conclusions, variations of the RPD-model can

---

[8]For a detailed description RPD-model and the tournament see Axelrod (1984). An open-source implementation is available from: `www.eckhartarnold.de/apppages/coopsim`.

be constructed where the conclusion becomes invalid (Arnold, 2013, 107). It is even possible to construct a variant that allows strategies to break off the repeated interaction at will and that does not lead to the breakdown of cooperation (Schüßler, 1997). The failure to derive any robust results highlights the danger of drawing generalizing conclusions from models and of relying on models as a tool of theoretical investigation. This point has most strongly been emphasized by Ken Binmore, who describes the popularity that Axelrod's model enjoyed derogatorily as the "The Tit-For-Tat Bubble" (Binmore, 1994, 194). Because the folk theorem from game theory implies that there are infinitely many equilibria in the Reiterated Prisoner's Dilemma, there is not much reason to assign of all things the *Tit For Tat*-equilibrium a special place (Binmore, 1994, 313-317). If one follows Binmore's criticism then it is not the reiterated Prisoner's Dilemma that explains why *Tit For Tat* is such a good strategy, but rather the fact that *Tit For Tat* is a very salient and easily understood mode of behavior in many areas of life that explains why people so easily believed in the superiority of the *Tit For Tat* strategy in the RPD game.

It is not only its lack of robustness that troubles Axelrod's model. It is also the difficulty of relating it to any concrete empirical subject matter – a problem that Axelrod shares with many game theoretical explanations.[9] Axelrod himself had offered a very impressive example of empirical application by relating the RPD model to the silent "Live and Let Live" agreement that emerged between enemy soldiers on some of the quieter stretches of the western front in the First World War. However, as critics were quick to point out (Battermann, D'Arms and Kryzstof, 1998; Schüßler, 1997), it is not at all clear whether this situation really is a Prisoner's Dilemma situation, let alone how the numerical values of the payoff parameters could be assessed. But precise numerical payoff values would be necessary since Axelrod's model is not robust against changes of the numerical values of the payoff parameters within the boundaries that the Prisoner's Dilemma game allows (Arnold, 2008, 80). Also, Axelrod's model could not explain why "Live and Let Live" occurred only on some stretches of the front line (Arnold, 2008, 180). Therefore, Axelrod's theory of the evolution of cooperation could not

---

[9]This is very frankly admitted by the leading game theorist Rubinstein (2013) in a newspaper article. Rubinstein resorts to an aesthetic vindication of game theory ("flowers in the garden of God").

really add anything substantial to the historical explanation of the "Live and Let Live" by Tony Ashworth (1980).

The chapter from Axelrod's book on the "Live and Let Live"-system shows that he did not understand his model only as a normative model, but at least also as an explanatory model. And the model was certainly understood as potentially explanatory by the biologists who were trying to apply it to cooperative behavior among animals (see below). The distinction is important, because the validation requirements for normative models are somewhat relaxed in comparison to explanatory models. After all, we would not expect from a model that is meant to generate advice for rationally adequate behavior to correctly predict the behavior of unadvised and potentially irrational agents. Still, even normative models must capture the essentials of the empirical situations to which they are meant to be applied well enough to generate credible advice. Here, too, robustness is an important issue. For similar reasons as in the descriptive case it would be dangerous to trust the advice given on the basis of a non-robust model.

Thus, in contrast to Schelling's model Axelrod's model is neither robust nor can the postulated driving factors of the emergent phenomenon (stable cooperation) easily be identified empirically. In Schelling's case the driving factor was the assumed tolerance threshold, in Axelrod's case it is the payoff parameters of the Prisoner's Dilemma. Therefore, two important prerequisites (robustness and empirical identifiability) for the application of a formal model to a social process appear to be absent in Axelrod's case.

The popularity of Axelrod's computer tournaments had the consequence that it became a role model for much of the subsequent simulation research on the evolution of cooperation. It spawned myriads of similar simulation studies on the evolution of cooperation (Dugatkin, 1997; Hoffmann, 2000). Unfortunately, most of these simulation studies remained unconnected to empirical research. Axelrod had – most probably without intending it – initiated a self-sustaining modeling tradition where modelers would orientate their next research project on the models that they or others had published before without paying much attention to what kind of models might be useful from an empirical perspective. Instead it was more or less silently assumed that because of the generality of the model investigations of the reiterated Prisoner's Dilemma model would surely be useful.

How little contact the modeling tradition initiated by Axelrod had to empirical research becomes very obvious in a survey of empirical research on the evolution of cooperation in biology by Dugatkin (1997). In the beginning, Dugatkin lists several dozens of game theoretical simulation models of the evolution of cooperation, an approach to which Dugatkin himself is very favorable. However, none of the models can be related to particular instances of cooperation in animal wildlife. A seemingly insurmountable obstacle in this respect is that payoff parameters usually cannot be measured. It is just very difficult to measure precisely the increased reproductive success, say, that apes that reciprocate grooming enjoy over apes that don't.

The most serious attempt to apply Axelrod's model was undertaken by Milinski (1987) in a study on predator inspection behavior in shoal fishes like sticklebacks. When a predator approaches, it happens that one or two sticklebacks leave the shoal and carefully swim closer to the predator. The hypothesis was that if two sticklebacks approach the predator they play a Reiterated Prisoner's Dilemma and make the decision to turn back based on a *Tit For Tat* strategy taking into account whether the partner fish stays back or not. This was tested experimentally by Milinski (1987) as well as others (Dugatkin and Reeve, 1998, 59-69). While in his 1987-paper Milinski himself believed that the hypothesis could be confirmed, it was after a long controversy ultimately abandoned. In a joint paper on the same topic that appeared ten years later Milinski and Parker (1997) do not draw on the RPD model any more. In fact they treat it as an unresolved question whether the observed behavior is cooperative at all.

In a later discussion, Dugatkin explained the problem when linking the model research about cooperation to the empirical research in biology by the difficulty of establishing a feedback-loop between model research and empirical research (Dugatkin, 1998, 57-58). The empirical results were never fed back into the model building process and the obstacles when trying to apply the models were never considered by the modelers. Without a feedback-loop between theoretical and empirical research, however, the model-building process soon reaches a stalemate where models remain detached from reality.

The frustration about this kind of pure model research is well expressed in a polemical article by Peter Hammerstein (2003). "Why is there such a discrepancy

between theory and facts?" asks Hammerstein (2003, 83) and continues: "A look at the best known examples of reciprocity shows that simple models of repeated games do not properly reflect the natural circumstances under which evolution takes place. Most repeated animal interactions do not even correspond to repeated games." In saying so, Hammerstein is by no means opposed to employing game theory in biology. It's just that in the aftermath of Axelrod most simulation studies on the evolution of cooperation focused on the Reiterated Prisoner's Dilemma or similar repeated games. This shows that the demand for empirical validation has an important side effect besides allowing to judge the truth and falsehood of the models themselves: It forces the modelers to concern themselves seriously with the empirical literature and the empirical phenomena that their models address. If they do so, there is hope that this will lead quite naturally to the choice of simulation models that address relevant questions of empirical research. Or, as Hammerstein (2003, 92) nicely puts it: "Most certainly, if we invested the same amount of energy in the resolution of all problems raised in this discourse, as we do in the publishing of toy models with limited applicability, we would be further along in our understanding of cooperation."

Just how little model researchers care for the empirical content of their research is inadvertently demonstrated by a research report on the evolution of cooperation that appeared roughly 20 years after the publication of Axelrod's first paper about his computer tournament (Hoffmann, 2000). There is only one brief passage where the author of this research report talks about empirical applications of the theory of the evolution of cooperation. And in this passage there is but one piece of empirical literature that the author quotes, the study on predator inspection in sticklebacks by Milinski (1987)! Nevertheless, Hoffmann believes that the "general framework is applicable to a host of realistic scenarios both in the social and natural worlds" (Hoffmann, 2000, 4.3). Much more believable is Dugatkin's summary of the situation: "Despite the fact that game theory has a long standing tradition in the social sciences, and was incorporated in behavioral ecology 20 years ago, controlled tests of game theory models of cooperation are still relatively rare. It might be argued that this is not the fault of the empiricists, but rather due to the fact that much of the theory developed is unconnected to natural systems and thus may be mathematically intriguing but biologically meaningless" (Dugatkin, 1998,

57). That this fact could escape the attention of the modelers tells a lot about the prevailing attitude of modelers towards empirical research.

5   An ideology of modeling

The examples discussed previously indicate that simulation models can be a valuable tool to study some of the possible causes of some social phenomena. However, the examples also show that a) modeling approaches in the social sciences can easily fail to deliver resilient results, that b) social simulations are not yet generally embedded in a research culture where the critical assessment of the (empirical) validity of the simulation models is a salient part of the research process and that c) the significance of pure simulation results is likely to be overrated.

Unsurprisingly, simulation models in the social sciences excel when studying those causes that can be represented by a mathematical model as in the case of Schelling's neighborhood segregation model. Part of the secret of Schelling's success is surely that he had a good intuition for picking those example cases where mathematical models really work. But many of the causal connections that are of interest in the social science cannot be described mathematically. For example, the question how the proliferation and easy accessibility of adult content in the internet shapes the attitude of youngsters towards love, sex and relationships, is hardly a question that could be answered with mathematical models. Or, if we want to understand what makes people follow orders to slaughter other people even in contradiction to their acquired moral codes (Browning, 1992), then any reasonable answer to this question will hardly have the form of a mathematical model.[10]

Unfortunately, the field of social simulations has by now become so much of a specialized field that modelers are hardly aware of the strong limitations of their approach in comparison with conventional, model-free methods in the social sciences. There is a widespread, though not necessarily always outspoken belief that more or less everything can – somehow – be cast into a simulation model. Part of the reason for this belief may be the fact that with computers the power of modeling techniques has indeed greatly increased. This belief has found explicit

---

[10]A good discussion of the respective merits and limitations of different research paradigms in the social sciences can be found in Moses and rn L. Knutsen (2012).

expression in Joshua Epstein's keynote address to the Second World Congress of Social Simulation under the title "Why model?" (Epstein, 2008).

In the following I am going to discuss Epstein's arguments and point out the misconceptions underlying this belief. In my opinion these misconceptions are to no small degree responsible for the misguided practices in the field of social simulations. Epstein sets out by arguing that it is never wrong to model, because – as he believes – there exists only the choice between explicit and implicit models, anyway:

> The first question that arises frequently – sometimes innocently and sometimes not – is simply, "Why model?"Imagining a rhetorical (non-innocent) inquisitor, my favorite retort is, "You are a modeler."Anyone who ventures a projection, or imagines how a social dynamic – an epidemic, war, or migration – would unfold is running some model. But typically, it is an implicit model in which the assumptions are hidden, their internal consistency is untested, their logical con- sequences are unknown, and their relation to data is unknown. But, when you close your eyes and imagine an epidemic spreading, or any other social dynamic, you are running some model or other. It is just an implicit model that you haven't written down (see Epstein 2007).
>
> ...
>
> The choice, then, is not whether to build models; it's whether to build explicit ones. In explicit models, assumptions are laid out in detail, so we can study exactly what they entail. On these assumptions, this sort of thing happens. When you alter the assumptions that is what happens. By writing explicit models, you let others replicate your results. (Epstein, 2008, 1.2-1.5)

It is not entirely clear whether Epstein restricts his arguments to projections, but even in this case it is most likely false. It is simply not possible to cast anything that can be described in natural language into the form of a mathematical or computer model. But then we also cannot assume that this must be possible, if projections to the future are concerned. It is of course always commendable to make one's own assumptions explicit. But this does not require modeling.

In addition, there are certain dangers associated with mathematical and compu-
tational modeling:

1. the danger of underrating or ignoring those causal connections that do not
   lend themselves to formal descriptions.

2. the danger of arbitrary ad hoc decisions when modeling causes of which
   we only have a vague empirical understanding. The necessity to specify
   everything precisely easily leads to the sin of false precision, which consists
   in assuming detailed knowledge where in fact there is none.

3. the danger of conferring a deceptive impression of understanding even if the
   model is not validated.

4. the shaping and selection of scientific questions by the requirements of
   modeling, rather than by other, arguably more important, criteria of relevance
   as, for example, the social impact or relevance for public policy.

That Epstein mentions replicability as another advantage of explicit modeling
is ironic given that it is still quite uncommon in published simulation studies to give
a reference for the reader to access and replicate the model (as described further
above). More worrisome, however, is Epstein's attitude towards validation:

> ... I am always amused when these same people challenge me with the
> question, "Can you validate your model?"The appropriate retort, of
> course, is, "Can you validate yours?"At least I can write mine down so
> that it can, in principle, be calibrated to data, if that is what you mean
> by "validate,"a term I assiduously avoid (good Popperian that I am).
> (Epstein, 2008, 1.4)

Calibration (i.e. fitting a model to data) is of course neither the same nor a
proper substitute for validation (testing a model against data), as Epstein knows.
Validation in the sense of empirical testing of a model, hypothesis or theory is
a common standard in almost all sciences, including those sciences mentioned
earlier that usually do not rely on formal models like history, ethnology, sociology,

21

political science. It is obviously not the case that validation presupposes explicit modeling, for otherwise history as an empirical science would be impossible.

Epstein furthermore advances 16 reasons for building models other than prediction (Epstein, 2008, 1.9-1.17). None of these reasons is exclusively a reason for employing models, though. The functions, for example, of guiding data collection or discovering new questions can be fulfilled by models and also by any other kind of theoretical reasoning. Nor is it an exclusive virtue of the modeling approach "that it enforces a scientific habit of mind" (Epstein, 2008, 1.6). Here Epstein is merely articulating the positivistic stock prejudice of the superiority, if only of a didactic kind, of formal methods. Given what Heath, Hill and Ciarallo (2009) have found out about the lack of proper validation of many agent-based simulations one might even be inclined to believe the opposite about the simulation method's aptitude to encourage a scientific habit of mind.

It fits into the picture of a somewhat dogmatic belief in the power of modeling approaches that modelers consider the lack of acceptance of their method often as more of a psychological problem on the side of the recipients to be addressed by better propaganda (Barth, Meyer and Spitzner, 2012, 2.11-2.12, 3.22-3.26), rather than a consequence of the still immature methodological basis of many agent-based simulation studies. This attitude runs the risk of self-deception, because one of the major reasons why non-modelers tend to be skeptical of agent-based simulations is that they perceive such simulations as highly speculative. As we have seen, the skeptics have good reason to do so.

6   Conclusions

It is in my opinion not least because of the abundance of simulations with low empirical impact that "social simulation is not yet recognized in the social science mainstream" (Squazzoni and Casnici, 2013, abstract). Why should a mainstream social scientist take simulation studies seriously, if he or she cannot be sure about the reliability of the results, because the simulations have never been validated? If modelers started to take the requirement of empirical validation more seriously, I expect two changes to occur – both of them beneficial: 1) Social simulations will become more focused in scope. Scientists will not attempt to cast anything into the

form of a computer simulation from classical social contract philosophy (Skyrms, 1996, 2004) to, well, the whole world (Fut, 2012; Liv, 2012), but they will develop a better feeling for when simulations can be empirically validated and when not, and they will mostly leave out those problems where computer simulations cannot be applied with some hope of producing empirically applicable results. 2) Yet, while the simulation method will become more focused in scope, it will at the same time become much more useful in practice, because simulations will more frequently yield results that other scientists can rely on without needing to worry about their speculative character and potential lack of reliability.

Reference

Arnold, Eckhart. 2008. *Explaining Altruism. A Simulation-Based Approach and its Limits*. Heusenstamm: ontos Verlag.

Arnold, Eckhart. 2013. "Simulation Models of the Evolution of Cooperation as Proofs of Logical Possibilities. How Useful Are They?" *Ethics & Politics* 2(XV):101–138.
URL: *http://www2.units.it/etica/2013_2/ARNOLD.pdf*

Ashworth, Tony. 1980. *Trench Warfare 1914-1918. The Live and Let Live System*. MacMillan Press Ltd.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. Basic Books.

Aydinonat, N. Emrah. 2007. "Models, conjectures and exploration: an analysis of Schelling's checkerboard model of residential segregation." *Journal of Economic Methodology* 14(4):429–454.
URL: *http://www.informaworld.com/smpp/ftinterface~content=a-787026382~fulltext=713240930*

Barth, Rolf, Matthias Meyer and Jan Spitzner. 2012. "Typical Pitfalls of Simulation Modeling - Lessons Learned from Armed Forces and Business." *Journal of Artifical Societies and Social Simulation (JASSS)* 15(5):2.
URL: *http://jasss.soc.surrey.ac.uk/15/2/5.html*

Battermann, Robert, Justin D'Arms and Górny Kryzstof. 1998. "Game Theoretic Explanations and the Evolution of Justice." *Philosophy of Science* 65:76–102.

Binmore, Ken. 1994. *Game Theory and the Social Contract I. Playing Fair*. Fourth printing (2000) ed. Cambridge, Massachusetts / London, England: MIT Press.

Binmore, Ken. 1998. *Game Theory and the Social Contract II. Just Playing*. Cambridge, Massachusetts / London, England: MIT Press.

Browning, Christopher R. 1992. *Ordinary men: Reserve Police Battalion 101 and the final solution in Poland*. New York, N.Y.: HarperCollins.

Dugatkin, Lee Alan. 1997. *Cooperation among Animals*. Oxford University Press.

Dugatkin, Lee Alan. 1998. Game Theory and Cooperation. In *Game Theory and Animal Behavior*. Oxford University Press chapter 3, pp. 38–63.

Dugatkin, Lee Alan and Hudson Kern Reeve. 1998. *Game Theory and Animal Behavior*. Oxford University Press.

Ellen, Ingrid Gould. 2000. *Sharing America's neighborhoods: the prospects for stable racial integration*. Cambridge, Mass.: Harvard University Press.

Epstein, Joshua M. 2008. "Why Model?". Based on the author's 2008 Bastille Day keynote address to the Second World Congress on Social Simulation, George Mason University, and earlier addresses at the Institute of Medicine, the University of Michigan, and the Santa Fe Institute.
URL: *http://www.santafe.edu/research/publications/workingpapers/08-09-040.-pdf*

Fut. 2012. "| FuturICT FET Flagship.".
URL: *http://www.futurict.eu/*

Gilbert, Nigel and Klaus G. Troitzsch. 2005. *Simulation for the Social Scientist*. McGraw-Hill.

Hammerstein, Peter. 2003. Why Is Reciprocity So Rare in Social Animals? A Protestant Appeal. In *Genetic and Cultural Evolution*, ed. Peter Hammerstein. Cambridge, Massachusetts / London, England: MIT Press in cooperation with Dahlem University Press chapter 5, pp. 83–94.

Heath, Brian, Raymond Hill and Frank Ciarallo. 2009. "A Survey of Agent-Based Modeling Practices (January 1998 to July 2008)." *Journal of Artificial Societies and Social Simulation (JASSS)* 12(4):9.
URL: *http://jasss.soc.surrey.ac.uk/12/4/9.html*

Hoffmann, Robert. 2000. "Twenty Years on: The Evolution of Cooperation Revisited." *Journal of Artificial Societies and Social Simulation* Volume 3, No. 2.
URL: *http://jasss.soc.surrey.ac.uk/3/2/forum/1.html*

Kuorikoski, Jaakko and Aki Lehtinen. 2009. "Incredible Worlds, Credible Results." *Erkenntnis* 70:119–131.
URL: *http://dx.doi.org/10.1007/s10670-008-9140-z*

Liv. 2012. "Living Earth Simulator will simulate the entire world | ExtremeTech.".
URL: *http://www.extremetech.com/extreme/108025-living-earth-simulator-will-simulate-the-entire-world*

Milinski, Manfred. 1987. "TIT FOR TAT in sticklebacks and the evolution of cooperation." *nature* 325, January:433–435.

Milinski, Manfred and Geoffrey A. Parker. 1997. "Cooperation under predation risk: a data-based ESS analysis." *Proceedings of the Royal Society* 264:1239–1247.

Moses, Jonthon W. and Torbjørn L. Knutsen. 2012. *Ways of Knowing. Competing Methodologies in Social and Political Research.* 2nd (first edition 2007) ed. London: palgrave macmillen.

Mäki, Uskali. 2009. "MISSing the World. Models are Isolations and Credible Surrogate Systems." *Erkenntnis* 70:29–43.
URL: *http://dx.doi.org/10.1007/s10670-008-9135-9*

Railsback, Steven F. and Volker Grimm. 2011. *Agent-Based and Individual-Based Modeling: A Practical Introduction.* Princeton University Press.

Reiss, Julian. 2011. "A Plea for (Good) Simulations: Nudging Economics Toward an Experimental Science." *Simulation & Gaming* 42(2):243–264.
URL: *http://sag.sagepub.com/content/42/2/243*

Rendell, Luke, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M.W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap and Kevin N. Laland. 2010. "Why Copy Others? Insights from the Social Learning Strategies Tournament." *Science* 328:208–213.
URL: *http://www.sciencemag.org/cgi/content/abstract/328/5975/208*

Rubinstein, Ariel. 2013. "Kann die Spieltheorie die Probleme der Eurozone lösen und das iranische Atomprogramm aufhalten?" Daily Newspaper.
URL: *http://www.faz.net/aktuell/feuilleton/debatten/spieltheorie-kann-die-spieltheorie-die-probleme-der-eurozone-loesen-und-das-iranische-atomprogramm-aufhalten-12129126.html*

Schelling, Thomas C. 1971. "Dynamic models of segregation†." *The Journal of Mathematical Sociology* 1(2):143–186.
URL: *http://www.tandfonline.com/doi/abs/10.1080/0022250X.1971.9989794*

Schüßler, Rudolf. 1997. *Kooperation unter Egoisten: Vier Dilemmata.* 2nd (first:1990) ed. München: R. Oldenbourg Verlag.

Shapiro, Ian. 2005. *The Flight from Reality in the Human Sciences.* Princeton and Oxford: Princeton University Press.

Skyrms, Brian. 1996. *Evolution of the Social Contract.* Cambridge: Cambridge University Press.

Skyrms, Brian. 2004. *The Stag Hunt Game and the Evolution of Social Structure.* Cambridge: Cambridge University Press.

Squazzoni, Flaminio and Niccolò Casnici. 2013. "Is Social Simulation a Social Science Outstation? A Bibliometric Analysis of the Impact of JASSS." *Journal of Artificial Societies and Social Simulation* 16(1):10.
URL: *http://jasss.soc.surrey.ac.uk/16/1/10.html*

Sugden, Robert. 2000. "Credible worlds: the status of theoretical models in economics." *Journal of Economic Methodology* 7(1):1–31.

Sugden, Robert. 2009. "Credible Worlds, Capacities and Mechanisms." *Erkenntnis* 70:3–27.
URL: *http://dx.doi.org/10.1007/s10670-008-9134-x*

Šalamon, Tomáš. 2011. *Design of Agent-Based Models. Developing Computer Simulations for a Better Understanding of Social Processes.* Řepin-Živonín: Tomáš Bruckner.

Wilensky, Uri. 1999. "netlogo.".
URL: *http://ccl.northwestern.edu/netlogo/*

Will, Oliver and Rainer Hegselmann. 2008. "A Replication That Failed – on the Computational Model in Michael W. Macy and Yoshimichi Sato: Trust, Cooperation and Market Formation in the U.S. and Japan. Proceedings of the National Academy of Sciences, May 2002?" *Journal of Artificial Societies and Social Simulation* 11(3):3.
URL: *http://jasss.soc.surrey.ac.uk/11/3/3.html*

Wilson, Robin J. 2002. *Four Colors Suffice: How the Map Problem was Solved.* Princeton University Press.