How Models Fail

A Critical Look at the History of Computer Simulations of the Evolution of Cooperation

Eckhart Arnold

Bavarian Academy of Sciences

Digital Humanities Dpt.

April 2015

Abstract

Simulation models of the Reiterated Prisoner's Dilemma have been popular for studying the evolution of cooperation since more than 30 years now. However, there have been practically no successful instances of empirical application of any of these models. At the same time this lack of empirical testing and confirmation has almost entirely been ignored by the modelers community. In this paper, I examine some of the typical narratives and standard arguments with which these models are justified by their authors despite the lack of empirical validation. I find that most of the narratives and arguments are not at all compelling. None the less they seem to serve an important function in keeping the simulation business running despite its empirical shortcomings.

Keywords: Evolution of Cooperation, Social Simulations, History of Simulations

Contents

1 Introduction

Simulation models of the Reiterated Prisoner's Dilemma (in the following: RPD-models) are since 30 years considered as one of the standard tools to study the evolution of cooperation (Rangoni, 2013) (Hoffmann, 2000). A considerable number of such simulation models has been produced by scientists. Unfortunately, though, none of these models has empirically been verified and there exists no example of empirical research where any of the RPD-models has successfully been employed to a particular instance of cooperation. Surprisingly, this has not kept

scientists from continuing to produce simulation models in the same tradition and from writing their own history as a history of success. In a recent simulation study – which does not make use of the RPD but otherwise follows the same pattern of research – Robert Axelrod's (Axelrod, 1984) original role model for this kind of simulation studies is praised as "an extremely effective means for investigating the evolution of cooperation" and considered as "widely credited with invigorating that field" (Rendell et al., 2010*b*, 208-209).

According to a very widespread philosophy of science that is usually associated with the name of Karl Popper (1971) science is distinguished from non-science by its empirical testability and right theories from wrong theories by their actual empirical success. Probably, most scientists in the field of social simulations would even agree to this philosophy of science at least in its general outlines.[1] However, RPD models of the evolution of cooperation have not been empirically successful. So, how come that they are still considered as valuable?

In this paper I am going to examine the question, why the continuous lack of empirical success did not lead the scientists working with these simulation models to reconsider their approach. In the first part I explain what RPD-models of the evolution of cooperation are about. I show that these models failed to produce empirically applicable and tenable results. This will be done by referring to research reports and meta-studies, none of which comes up with an example of successful empirical application.

In the second part of the paper, I highlight a few example cases that show why these models fail. In this context I examine the framing narratives with which

---

[1]A referee pointed out to me that there is a tension in my paper between the reliance on a Popperian falsificationism and the implicit use of Kuhn's paradigm concept. However, both can be reconciled if the former is understood in a normative and the latter in a descriptive sense. Popper's falsificationism requires, though, that paradigms are not completely incommensurable. But then, there are many good reasons that speak against a strong reading of the incommensurability-thesis, anyway.

Similarly, the Duhem-Quine-thesis does not result in a fatal problem for a Popperian epistemology, if one admits that in most concrete contexts there exist further clues which allow to decide which particular elements of a falsified set of propositions are more likely to be responsible than others. For example, if an experiment falsifies a well established physical theory then it is prima facie more likely that a loose wire was the cause than a failure of the theory. Only after this has been checked carefully one would assign different probabilities to the experimental setup or the theory being false, respectively. See the very enlightening remarks about Kuhn and Duhem-Quine in the case study by Zacharias (2013, 11ff., 305ff.).

scientists justify their method. Such framing narratives form an integral part of any scientific enterprise. My point is not to criticize simulation scientists for employing narratives to justify their method, but I believe that the typical framing narratives that RPD modelers in the tradition of Axelrod employ of are badly founded and I show that in each case there are good arguments against accepting the narrative.

In the third part of this paper I take this analysis one step further by discussing typical arguments with which scientists justify the production and use of unvalidated "theoretical" simulations. Most of the arguments discussed here do usually not form the central topic of scientific papers. Rather, they appear in the less formal communication of scientists, in oral discussions, in small talk, eventually in keynote addresses (Epstein, 2008). One may object that if these arguments are never explicitly spelled out, they may not be worth discussing. After all they have never been cast into their strongest imaginable form. Why discuss dinner table talk, anyway? But then, it is often this kind of communication where the deeper convictions of a community are expressed. And it is by no means true that these convictions are without effect on the scientific judgments of the community members. Quite to the contrary, general agreement with the underlying convictions is silently presupposed by one's scientific peers and adherence to them is usually taken for granted by supervisors from their PhD students and often expected by referees from the authors of the papers they review. Therefore, the informal side-talk of science should not at all be exempt from rational criticism.

In the last part of the paper, I relate my criticism to similar discussions in a neighboring (if not overlapping) science, namely political science. It seems that there exist structural similarities in the way scientific schools or research traditions deal with failures of their paradigm. Rather than admitting such a fundamental failure (which, as it touches one's own scientific world view, is obviously much harder than admitting the failure of a particular research enterprise within a paradigm) they retreat by adjusting their goals. In the worst case they become so modest in their achievements (which they, by an equal adjustment of their self-perception, continue to celebrate as successes) that they reach the verge of irrelevance. Green and Shapiro (1994, 44f.) have described this process of clandestine retreat for the case of rational choice theory in political science.

4

2 The Empirical Failure of Simulations of the Evolution of Cooperation

2.1 Axelrod's Evolution of Cooperation

One of the most important initiators of the research on the RPD-model was Robert Axelrod. The publication of his book "The Evolution of Cooperation" popularized the simulation approach to studying the evolution of cooperation. At the core of Axelrod's simulation lies the two person's Prisoner's Dilemma game. The two person's Prisoner's Dilemma is a game, where two players are asked to contribute to the production of a public good. Each player can choose to either contribute, that is, to cooperate, or not to contribute, that is to defect. If both players cooperate they both receive a reasonably high payoff. If neither player cooperates, they both receive a low payoff. If one player tries to cooperate while the other player defects, the player who tried to cooperate receives a zero payoff, while the successful cheater receives the highest possible payoff in the game, which at the same time is more than the cooperative payoff. Since, no matter what the other player does, it is always more advantageous for each individual player not to cooperate. Therefore, both players, if they are rational egoists, end up with the low non-cooperation payoff – at least as long as the game is not repeated. The reiterated Prisoner's Dilemma (RPD), in which the same players play through a sequence of Prisoner's Dilemmas, changes the situation, because defecting players can be punished with non-cooperation in the following rounds.

It can be shown that in the reiterated Prisoner's Dilemma there is no single best strategy. In order to find out if there exist certain strategies that are by and large more successful than other strategies and whether there are certain characteristics that successful strategies share, Robert Axelrod (1984) conducted a computer tournament with different strategies. Axelrod also fed the results of the tournament simulation into a population dynamical simulation, where more successful strategies would gradually out-compete less successful strategies in a quasi-evolutionary race. Famously, TIT FOR TAT emerged as the winner in Axelrod's tournament.[2]

---

[2]More detailed descriptions of the RPD-model and Axelrod's tournament can be found in

The way Axelrod employed his model as a research tool was by running simulations and then generalizing from the results he obtained. These included recommendations such as that TIT FOR TAT usually is good choice for a strategy or that a strategy should not defect unmotivated itself, but should punish defections and should also be forgiving etc. As subsequent research revealed, however, almost none of these conclusions was in fact generalizable (see Arnold (2013, 106ff., 126f.) with further references). For each of them there exist variations of the RPD-model where it does not hold and where following Axelrod's recommendations could be a bad mistake. The only exception is Axelrod's result about the collective stability of TIT FOR TAT, which he proved mathematically.

The central flaw of Axelrod's research design is that it relies strongly on impressionistic conclusions and inductive generalizations from what are in fact contingent simulation results. This deficiency of Axelrod's model has convincingly been criticized by Binmore (1998, 313ff.). To give just one example: In Axelrod's tournament TIT FOR TAT won in two subsequent rounds. Axelrod concluded that TIT FOR TAT is a good strategy and that it is advisable to be forgiving. However, if one chooses the set of all 2-state automata as strategy set – which is a reasonable choice because it contains all strategies up to a certain complexity level – then the unforgiving strategy GRIM emerges as the winner (Binmore, 1994, 295ff.).

Axelrod's followers would usually be much more cautious about drawing general conclusions from simulations, but they did not completely refrain from generalizing. In the ensuing research a historical pattern emerged where researchers would pick up existing models, investigate variants of these models, and eventually demonstrate that the previous results could not be generalized (Schüßler's and Arnold's simulations, which are discussed below, are examples for this pattern). Thus, Axelrod's research design became – despite its great deficiencies – a role model for simulation studies until today. As a justification for publishing yet another model, it would usually suffice to relate to the previous research. No reference to empirical research or just empirical applicability would be considered necessary. For example, Rangoni (2013) introduces his study of a variant of the RPD-model by mentioning that "Axelrod's work on the prisoner's dilemma is one of the most discussed models of social cooperation" and declaring that "After more

---

Axelrod (1984), Binmore (1994, 1998) or Arnold (2008).

than thirty years from the publication of its early results, Axelrod's prisoner's dilemma tournament remains a cornerstone of evolutionary explanation of social cooperation", although – as will be discussed in the following – this "cornerstone of evolutionary explanation" has not been confirmed empirically in a single instance.

2.2 The empirical failure of the RPD-model

Axelrod himself was confident that simulation studies like his yield knowledge that can be applied in the context of empirical application. In his book "The Evolution of Cooperation" (Axelrod, 1984) he provided two case studies. One of these concerned biology. It was of highly speculative character as Axelrod honestly admitted and it has indeed never been confirmed since. Therefore, I am not going to discuss this particular case study here. Further biological research on the evolution of cooperation will briefly be outlined below.

The other one of Axelrod's case studies was a highly dramatic case study concerning the Live and Let Live System which emerged on some stretches of the deadlocked western front between enemy soldiers in World War One. However, as acknowledged by Axelrod his case study relies entirely on the prior historiographic work by Tony Ashworth (1980). Based on an extensive study of the historical sources, Ashworth had crafted a careful and highly differentiated explanation for the emergence, sustainment and eventual breakdown of the Live and Let Live on the western front. Axelrod's recasting of this story in game theoretical terms has nothing to add in terms of explanatory power, because the RPD-model is far too simple to account for the complicated network of causes for the Live and Let Live that Ashworth study had revealed (Arnold, 2008, 180ff.). Even among game theorists it was disputed, whether there existed any straight-forward way to interpret the situation as a Prisoner's Dilemma at all (Schüßler, 1997, 33ff.). Thus, if any particular scientific approach is to be credited with the successful explanation of the Live and Let Live in World War One, then it is not game theoretical modeling or computer simulations but the well-established methods of traditional historiography. Interestingly, though, this dramatic case-study did a lot to increase the popularity of Axelrod's simulation approach.

If Axelrod's attempts to apply his model to empirical case studies weren't

7

particularly successful, then subsequent research could still demonstrate that the empirical application of these models is possible. The most noteworthy attempt to apply the RPD empirically was undertaken by Manfred Milinski (1987), who sought to explain the seemingly cooperative behavior that shoal fishes show when inspecting a predator. This paper is quoted time and again when it comes to giving an example for the empirical applicability of the RPD model. For example, Hoffman maintains in a research report about Axelrod's RPD framework that "This general framework is applicable to a host of realistic scenarios both in the social and natural worlds (e.g. Milinski 1987)." (Hoffmann, 2000, 4.3). Milinski's 1987-paper, however, remains the sole example for the "host of realistic scenarios" to which this framework is supposedly applicable. The same paper by Milinski is quoted in Osborne's "Introduction to Game Theory" as an example for the empirical applicability of game theory (Osborne, 2004, 445). Unfortunately, it was already by the late 1990s clear that Milinski's explanation of the predator inspection behavior did not work (Dugatkin, 1997, 1998). The reason is that it is not possible to obtain the necessary empirical data to either confirm or disconfirm the RPD model in the case of the predator inspection behavior of sticklebacks. This is also more or less the conclusion at which Milinski and Parker arrive in a joint paper on the same topic that they published 10 years after the initial study by Milinski (Milinski and Parker, 1997, 1245).

In a broad meta-study on the research on "Cooperation among Animals" Lee Allan Dugatkin (1997) does not find a single instance of animal cooperation where any of the many variants of the RPD model (Dugatkin lists more than two dozens of them in the beginning of his study) can successfully be applied. He summarizes the situation in a very thoughtful article as follows: "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 and Reeve, 1998, 57). The same frustration about empirically ungrounded model research is expressed by Peter Hammerstein: "Why is there such a discrepancy between theory and facts? 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." (Hammerstein, 2003, 83). It is safe to say that there exist no successful empirical application cases for the RPD in biology. But the fact that the modeling community still entertains the believe that there are such successful application cases, if not "a host of" them, clearly demonstrates how little, in fact, the community occupies itself with empirical matters.

3 Justificatory narratives

If the simulations studies in this research tradition do not bear any explanatory value for empirical research, then the question naturally arises what they are good for. Some authors present explanations that are meant to justify the method. I will go through some of them before entering on the discussion of the general arguments in favor for the simulation method.

3.1 Axelrod's narrative

Axelrod motivated the use of the Prisoner's Dilemma mostly by the fact that it already was an extremely popular game theoretical model that had already been used in experimental economic research. He compares the Prisoner's Dilemma to the E.coli in biological research. Comparisons to the ever successful natural sciences are quite typical for the justificatory discourse of the modeling approaches in the social sciences. With the benefit of hindsight it can, however, be said that this comparison was slightly misleading. E.coli is a great object of study in biology, because what one learns when studying E.coli can often directly be transferred to other bacteria. Many bacteria are similar to E.coli in important respects. The same is unfortunately not true for the RPD model, which is not at all a robust model (Arnold, 2013, 127f.). Change the parameters of the simulation, the initial set of participating strategies or other aspects of the model only a bit and you can get qualitatively different results. Most likely, another strategy than TIT FOR TAT would turn out as winner, and maybe not even a friendly or cooperative strategy

(Binmore, 1994, 315).

One part of Axelrod's motivation is also a supposed advantage of the simulation approach to experimental approaches. Axelrod relates to the notorious problem of economic experimental research that the laboratory setting is usually highly artificial and that, therefore, any obtained results cannot easily be transferred to real life situations. He omits to mention, however, that computer simulations based on highly stylized models like the RPD share the same problem.

3.2   Schüßler's narrative

Several years after Axelrod, Rudolf Schüßler (1997) published a book with game theoretical simulations. One part of this book directly relates to Axelrod. This part of Schüßler's book follows the pattern: Pick a well-known simulation, change the settings or other details of this simulations, produce "surprising" results and publish. If Axelrod had demonstrated with his simulation that the shadow of the future is crucial for the evolution of cooperation, Schüßler demonstrates with a modified simulation that this does not need to mean that the same partners must expect to meet again and again in order to sustain cooperation. In Schüßler's simulation cooperators succeed although cheaters can decide to break off the interaction at any time, thus avoiding punishment.[3]

Given Axelrod's previous simulations and conjectures this can appear surprising. But what is surprising? That a different simulation produces different results is prima facie anything but surprising. Given the almost complete modeling freedom – remember, there are no empirical constraints to be honored – and the volatility of the original model it would be surprising if no surprises could be produced. So why should we be interested in the results of another arbitrary simulation?

At this point Schüßler's narrative steps in. As Schüßler (1997, 91) writes "One of the central, classical assumptions of the normativistic sociology says that in an exchange society of rational egoists no stable cooperation can emerge (see Durkheim 1977, Parsons 1949). Alleged proofs for this thesis try to show that

---

[3]The details of this simulation are described in Schüßler (1997, 61ff.) and in a simpler form in Arnold (2008, 291ff.). For the curious: Schüßler achieves his effect, because the non-cooperators that break off the interaction are forced to pick a new partner from a pool that mostly contains non-cooperators from which it is impossible to rip a high payoff.

already simple analytical considerations suffice to draw this conclusion. The present simulation should be able to shake this firm conviction."[4] One may wonder whether this means that the simulation serves more than a purely didactic purpose. But be that as it may. It is in any case questionable whether the premises are correct. Do normativistic sociologists really rely on simple analytical considerations? Sociologists like Durkheim usually argue on the basis of thick narratives supported by empirical research. Highly abstract computer simulations like Rudolf Schüßler's simulations can at best prove logical possibilities. However, it is unlikely that this kind of discourse is vulnerable to proofs of logical possibilities. After all, a normativistic sociologist can easily claim that the seeming possibility of rational egoists to cooperate is an artifact of the simulations that strips away all concrete features of human nature, especially those of a psychological kind which make cooperation of egoists impossible in reality (Arnold, 2013, 128ff.). (Generally, proofs of logical possibilities cannot disprove real impossibilities; e.g. a perpetuum mobile is logically possible but impossible in reality, because it contradicts the laws of nature. See Arnold (2013) for a detailed discussion of the category of logical possibility.)

Schüßler, who seems to be quite aware of the weaknesses of his argument, follows up with the remark that ultimately it is up to the scientist to decide whether this is sufficient or not (Schüßler, 1997, 91). But as we have seen, proofs of logical possibility are simply not sufficient. And then again, it is an indefeasible claim that scientific knowledge is objective and that its validity is independent from the opinions and discretion of any particular person. If it were up to the discretion of the scientist to decide whether some theory or model is sufficient to decide a scientific question, we would not call that science any more.

It is noteworthy that Schüßler criticizes Axelrod quite strongly in the beginning of his book (Schüßler, 1997, 33ff.), but then presents computer simulations of exactly the same brand as Axelrod's simulations. The same kind of performative

---

[4]This is my translation. The German original reads: "Eine der zentralen, klassischen Annahmen der normativistischen Soziologie besagt, daß in einer Austauschgesellschaft rationaler Egoisten keine stabilen Kooperationsverhältnisse entstehen können (vgl. Durkheim 1977, Parsons 1949). Angebliche Nachweise für diese These versuchen zu zeigen, daß bereits einfache, analystische Überlegungen zu diesem Schluß ausreichen. Die vorliegende Simulation sollte geeignet sein, diese Sicherheit zu erschüttern." (Schüßler, 1997, 91)

self-contradiction is even more obvious in the following example.

3.3   The story of "slip stream altruism"

Although RPD simulations already fell out of fashion, I have myself published a book with RPD simulations as late as 2008. I felt uneasy about it at the time of writing the book and today I am even more convinced that the scientific method that I describe (but also criticize) in this book is fundamentally flawed. But the book was my PhD-thesis and I was not really given the free choice of topic – which is, of course, a widespread grievance of PhD-theses. So, I figured that the best I could make out of this situation was to follow the established pattern of research in this field, but also to examine it from an epistemological point of view and point out its deficiencies. The research pattern is that of producing a variant of an existing simulation model, finding "interesting" results and embedding them in a narrative that makes them appear "new", "surprising" or at least somehow noteworthy.

In the series of population dynamical simulations of the RPD that I conducted, there are quite a few simulations where naive cooperators, i.e. strategies that cooperate but other than TIT FOR TAT do not retaliate when the partner fails to reciprocate, can still survive with a low share of the population or – even more "surprising" – come out on top, i.e. with larger population share than even the retaliating cooperators (Arnold, 2008, 109ff.). I used the term "slip stream altruism" as a catch phrase to describe this phenomenon, because the simulations prove the logical possibility that unconditional altruism (which some moralists consider to be the only form of altruism that deserves its name) can develop in the "slip stream" of tough, reciprocating strategies.

But is this phenomenon really surprising and did we really need a series of computer simulations to get the idea? As mentioned earlier, with unrestricted modeling freedom and a volatile base model like the RPD, one is liable to find all kinds of phenomena. There are not really any surprises. And just as in Schüßler's case there is a simple explanation for the phenomenon: Unconditional cooperators can come out on top, if the conditional cooperators that drive the non-cooperators to extinction are badly coordinated so that they inadvertently hurt each other (Arnold, 2008, 113). So, the phenomenon that my simulation series yields acquires the

12

appearance of being interesting, surprising or relevant mostly by the narrative and the rhetoric of "slip stream altruism" in which it is embedded.

I never took the story of slip stream altruism very seriously and, as I said earlier, I was already convinced that the simulation method as practiced by Axelrod and his followers leads to nothing at the time when I wrote the book down. (See, for example, my talk at the Models & Simulations in Paris 2006, some time before I wrote down the book (Arnold, 2006).) Given how strongly I criticize Axelrod-style simulations in the book, it may appear odd to the readers that I even bothered to conduct computer simulations of the same brand and describe them in the book. As mentioned earlier, this was a tribute that I had to pay to the circumstances. Somewhat to my distress I later found that some readers liked the simulation series much better than my criticism of the method (Schurz, 2011, 344, 356). Others, at least, have understood that the main purpose of the book is a critical one (Zollman, 2009). In my (biased) opinion, however, I believe that the criticism or, what amounts to the same, the deficiencies of the simulation method as practiced by the adherents of Axelrod have not yet been taken seriously enough.

3.4   The social learning strategies tournament

The last example of a justificatory narrative does not concern the RPD model, but a simulation enterprise that is similar in spirit to Axelrod's. The authors of this study explicitly refer to Axelrod for the justification of their approach (Rendell et al., 2010*b*, 208-209). The model at the basis of the "Social Learning Strategies tournament" is a 100-armed bandit model (Rendell et al., 2010*a*, 30ff.). Just like the RPD it is a highly stylized and very sparse model: The model assumes an environment with 100 cells representing foraging opportunities. The payoff from foraging is distributed exponentially: few high payoffs, many low or even zero payoffs. In each round of the game the players can choose between three possible moves: INNOVATE where they receive information about the payoff opportunity in a randomly picked cell; EXPLOIT where players forage one of their known cells to receive a payoff; OBSERVE where a player receives slightly imprecise information about the foraging opportunities that other players are exploiting. Arbitrarily many players can occupy one cell. The resources never expire, but

the environment changes over time so that the players' information about good foraging opportunities gets outdated after a while. The payoffs drive a population dynamical model where players live and die and are replaced by new players depending on the success of the existing players.

The most important result of the tournament was that – under the conditions of this specific model – the best strategies relied almost entirely on social learning, i.e. playing OBSERVE. It almost did not make any sense at all to play INNOVATE.[5] Other than that the ratio between OBSERVE moves and EXPLOIT moves was crucial to success. Too few OBSERVE moves would lead to sticking with poor payoffs. Too many OBSERVE moves would mean that payoffs would not be gathered often enough which results in a lower average payoff. Finally, the right estimate of expected payoffs was important. The winning strategy and the second best strategy used the same probabilistic standard formula to estimate the expected payoff values (Rendell et al., 2010*b*, 211).

The authors themselves make every effort to present their findings as a sort of scientific novelty. For that purpose they employ a framing narrative that links their model with an important research question, prior research and successful (or believed to be successful) past role models. The broader research question, mentioned in the beginning of the paper, to which the model is related is how cultural learning has contributed to the success of humans as a species: "Cultural processes facilitate the spread of adaptive knowledge, accumulated over generations, allowing individuals to acquire vital life skills. One of the foundations of culture is social learning,..." (Rendell et al., 2010*b*, 208). Surely, this is a worthwhile scientific question.

As to the prior research they refer to theoretical studies. These, however, only "have explored a small number of plausible learning strategies" (Rendell et al., 2010*b*). Therefore, the tournament was conducted which gathers a contingent but large selection of strategies. The tournament's results are then described as "surprising results, given that the error-prone nature of social learning is widely thought to be a weakness of this form of learning ... These findings are particularly

---

[5]This was partly due to an inadvertency in the design of the model, where OBSERVE moves could – due to random errors – serve much the same function as INNOVATE moves. The authors of the study did, however, verify that their results are not just due to this particular effect (Rendell et al., 2010*a*, 21f.).

unexpected in the light of previous theoretical analyzes ..., virtually all of which have posited some structural cost to asocial learning and errors in social learning." (Rendell et al., 2010*b*, 212).

Thus, the results of the tournament constitute a novelty, even a surprising novelty. The surprising character of the results is strongly underlined by the authors of the study: "The most important outcome of the tournament is the remarkable success of strategies that rely heavily on copying when learning in spite of the absence of a structural cost to asocial learning, an observation evocative of human culture. This outcome was not anticipated by the tournament organizers, nor by the committee of experts established to oversee the tournament, nor, judging by the high variance in reliance on social learning ..., by most of the tournament entrants." (Rendell et al., 2010*b*, 212) Again, however, it is not surprising, but to be expected that one reaches results that differ form previous research if one uses a different model.

Axelrod's tournament plays an important role as historical paragon in the framing narrative: "The organization of similar tournaments by Robert Axelrod in the 1980s proved an extremely effective means for investigating the evolution of cooperation and is widely credited with invigorating that field."(Rendell et al., 2010*b*, 208). But as mentioned earlier, the general conclusions that Axelrod drew from his tournament had already turned out not to be tenable and the research tradition he initiated did not really yield any empirically applicable simulation models. Nonetheless, the author's seem to consider it as an advantage that: "Axelrod's cooperation tournaments were based on a widely accepted theoretical framework for the study of cooperation: the Prisoner's Dilemma." (Rendell et al., 2010*b*, 209). However, the wide acceptance of the Prisoner's Dilemma model says more about fashions in science than about the explanatory power of this model. Although not as widely accepted as the Prisoner's Dilemma, the authors are confident that "the basic generality of the multi-armed bandit problem we posed lends confidence that the insights derived from the tournament may be quite general." (Rendell et al., 2010*b*, 212). But the generality of the problem does not guarantee that the conclusions are generalizable beyond the particular model that was used to describe the problem. Quite the contrary, the highly stylized and abstract character of the model raises doubts whether it will be applicable without ambiguity in many

empirical instances. The generality of the model does not imply – nor should it, as I believe, lend any confidence in that direction to the cautious scientist – that it is of general relevance for the explanation of empirical instances of social and asocial learning. This simply remains to be seen. If anything at all then it is its robustness with respect to changes of the parameter values that lends some confidence in the applicability of the tournament's results. Robustness is of course only one of several necessary prerequisites for the empirical applicability of a model.

Summing it up, it is mostly in virtue of its framing narrative that the tournament's results appear as a novel, important or surprising theoretical achievement. If one follows the line of argument given here, however, then the model – being hardly empirically grounded and not at all empirically validated – represents just one among many other possible ways of modeling social learning. In this respect it is merely another grain of dust in the inexhaustible space of logical possibilities.

4   Discussion of Standard Arguments for Modeling

While the narratives discussed so far could be traced to their specific sources in the papers and books in which they appear, the following standard arguments for the supposed superiority of the simulation approach to studying the "evolution of cooperation" or for the use of formal models crop up in discussions and the less formal forms of scientific communication, but not so often in scientific papers. I have heard all of these arguments in discussions about the RPD simulation model more than once, but I cannot easily trace them back to printed sources. As I explained in the introduction, these arguments seem to me none the less to represent an attitude that effects the scientific work. Therefore, I believe that they deserve discussion.

4.1   "Our knowledge is limited, anyway"

Argument: Our ability to gain knowledge is limited in the social sciences, anyway. Therefore, we have to be content with the kind of computer simulations we can make, even if they are not sufficient to generate empirical explanations.

16

Response: No one says that we have to use computer simulations in the social sciences. If computer simulations do not work, other methods may still work. As explained earlier, the "Live and Let Live" in World War One cannot really be explained by RPD models, but historiographic methods still work perfectly well in this case.

Even if there exist no alternative methods, we should not accept the existing methods no matter how bad they are. The use of a particular scientific method is justified only, if the results it yields are better than mere speculation and by and large as good as or better than what can be achieved with alternative methods.

Moreover, we should not mistake the failure of a paradigm – say, agent-based simulations or RPD-simulations of cooperation or rational choice theory or sociobiology – for the failure of a science. It is only from the keyhole perspective of the strict adherents to one particular paradigm that the limits of the paradigm appear as the limits of the science or of human cognition as such. In this respect the argument resembles the strategy of silent retreat to false modesty mentioned in the introduction. While it is laudable for a scientist to be modest about one's own claims of knowledge, scientific modesty becomes inappropriate when it gives up any claim of generating empirically falsifiable knowledge.

4.2 "One can always learn something from failure"

Argument: Even if Axelrod's approach ultimately turned out to be a failure, we can still learn important lessons from it. Failure is at least as important for the progress of science as success.

Response: Unfortunately, it is not clear, whether the necessary lessons have already been learned. If Axelrod's computer tournament is still remembered as an "extremely effective means for investigating the evolution of cooperation" (Rendell et al., 2010*b*, 208) by the scientific community then it seems that the lessons have not been learned. And even if the lessons have been learned (by some) then the many dozens of inapplicable simulations that have kept scientists busy in the aftermath of Axelrod's book have surely been a rather long detour.

4.3 "Models always rely on simplification"

Argument: Models, by their very definition, rely on simplifications of reality. If a model wouldn't simplify it would be useless as a model. After all, the best map of a landscape would be the landscape itself, but then it would be useless as a map. (A typical example is Zollman (2009) who relies on this argument in his criticism of mine. See also Green and Shapiro (1994, 191) who discuss a similar argument in the context of rational choice theory.)

Response: On the other hand it is obvious that there must be some limit to how strongly a model may simplify reality. For otherwise any model could be a model for anything. So, where is the borderline between legitimate simplification and illegitimate oversimplification? A possible answer could be that a model is not oversimplified as long as it captures with sufficient precision all causally relevant factors of the modeled phenomenon with respect to a specific research question, i.e. all factors that are liable to determine the outcome of this question. In all other cases we should be very careful to trust an explanation based on that model alone.

At this point two replies are common: 1) That no one claims such an explanatory power for his or her own models. But then, what is the point of modeling, if models do not help us to explain anything? 2) That the research question did not require that all causally relevant factors have to be captured by one and the same model. However, if a model concentrates only on some causal factors, then these must at least be discernible empirically from other factors at work. Unfortunately, this is often not possible and certainly not with most RPD models. (See also Arnold (2014, 367f.).)

As far as RPD-simulations are concerned it appears clear to me that these are far too simplified to be acceptable representations of reality. One could object that they help us to understand the mechanism of reciprocal altruism as such. This is already one step back from claiming that RPD-models are an effective means for investigating the evolution of cooperation, because now it is merely claimed that they are illustrating a mechanism. However, for this purpose a single model would be sufficient. One does not need dozens of them. Plus, how and why reciprocal altruism works in principle has perfectly well been conceptualized by

Robert Trivers (1971) many years earlier with a single simple equation.

4.4 "There are no alternatives to modeling"

Argument: There is no real alternative to modeling, anyway. If you try to do without models, merely relying on verbal explanations, you are just making use of implicit models that are never fully articulated. Surely, explicit modeling is better than relying on implicit models. Without models nothing could be explained. (See also Epstein (2008), who employs a variant of this argument.)

Response: It is at least for the time being (the distant future of science may of course prove me wrong) practically impossible to express everything that can be expressed verbally in mathematical terms or with formal logic. This includes many of the causal connections that we are interested in when doing social sciences. Otherwise, how come that among the many books published about the causes, course and consequences of the First World War these days, there is no game theoretical or otherwise model-based study that could rival the conventional historical treatments? Otherwise, how come that lawyers, attorneys and judges – their job being to a large part one of logical reasoning, as one should think – do not use formal logic to express the legal connections they ponder over?

4.5 "Modeling promotes a scientific habit of mind"

Argument: "To me, however, the most important contribution of the modeling enterprise – as distinct from any particular model, or modeling technique – is that it enforces a scientific habit of mind, which I would characterize as one of militant ignorance – an iron commitment to 'I don't know.' That is, all scientific knowledge is uncertain, contingent, subject to revision, and falsifiable in principle. (This, of course, does not mean readily falsified. It means that one can in principle specify observations that, if made, would falsify it). One does not base beliefs on authority, but ultimately on evidence. This, of course, is a very dangerous idea. It levels the playing field, and permits the lowliest peasant to challenge the most exalted ruler – obviously an intolerable risk." (Epstein, 2008, 1.16)

Response: Unfortunately, the modeling tradition discussed in this paper failed completely with respect to all the virtues that Epstein naively believes to be virtues promoted by modeling: It did not readily submit its results to empirical falsification. Where the few and far between attempts of empirical application have been made and failed, the modelers did not learn from failure. (The empirical scientists did learn from the failure by turning away from the RPD-models.) The commitment to "I do not know" becomes a joke if modelers do not dare to come up with concrete empirical explanations or predictions any more. And as far as authority goes, the appeal to "scientific authority" in more or less subtle forms is a common rhetoric device in the modeler's discourse. (See also Moses and Knutsen (2012, 157), Green and Shapiro (1994, 195) and argument 7 below).

Generally, the scientific habit of mind does not at all depend on the use of models. Also, secondary virtues like clarity, explicitness and the like are by no means a prerogative of modelers. Computer simulation studies in particular can become dangerously unclear if the source code is not published or not well structured or not well commented.

4.6 "Division of labor in science exempts theoreticians from empirical work"

Argument: There exists division of labor in science. Model builders are not responsible for the empirical application of their models, but they are mere suppliers. If the empirical scientists fail to test or otherwise make use of models, it is not the modelers that should be blamed.

Response: But modelers need to take into account the conditions and restrictions that empirical research imposes, otherwise they run the danger of producing models that can never, not even under the most favorable circumstances, be applied empirically. In the case of the Axelrod-tradition it is clearly the modelers that must take the blame, because they failed to learn from the failures of early attempts at empirical application like Milinski's (1987). And they never worried about the restrictions under which empirical work struggles in the potential application fields of their models.

Now, one might say that this is also true for much of mathematics, and still mathematics has often proven to be applicable, even in cases that no one had guessed before. But surely it is not a good research strategy to rely on later to come historical coincidences of science. Plus, there is an important difference between mathematics and models. Mathematics deals with general structures, while simulation-models like the RPD represent particular example cases (comparable to a concrete calculation in mathematics). From a technical point of view most models in the Axelrod tradition remain fairly trivial, while mathematics could – if worst comes to worst – still be justified by its high intellectual level which allows to ascribe an innate value to it.

4.7 "Success within the scientific community proves scientific validity"

Argument: The scientific value of computer simulations in the social sciences cannot be disputed. There is a growing number of research projects, journals, institutes that is dedicated to social simulations. (Variants of this argument are: This book has been quoted so many times, it cannot be all wrong! Or, this article has been published in *Science*, the authors surely know what they are doing. See also Green and Shapiro (1994, 195), who discuss a similar argument.)

Response: The scientific value of a method, theory, model or simulation is to be judged exclusively on the basis of its scientific merits, i.e. logical reasoning and empirical evidence, and not at all on the basis of its social success. As far as computer simulations are concerned, a survey by Heath, Hill and Ciarallo (2009) on agent-based simulations revealed that the empirical validation of computer simulations is still badly lacking.

There is one grain of truth in this argument. For those questions, about which one does not know enough to judge the scientific arguments it is best to rely on the judgment of the socially approved specialists. But social success can never be used as an argument within a scientific dispute. After all, it is just the question whether the social success of a theory, model or paradigm was deserved from a scientific point of view.

4.8 "Natural sciences do it just the same way"

Argument: The use of models is pervasive throughout the natural sciences and in particular physics. Now, the natural sciences have been extremely successful and continuously progressing since their very inception in early modern times. Why should not social sciences learn from the successful methods of the natural sciences and employ models?

Response: So physicists do it just the same way? Nay, they don't! Throughout the natural sciences it is common practice to test models and theories rigorously against experiments and empirical observations. The success of the natural sciences is not only due to mathematical modeling alone, but rather to the co-evolution of mathematical theory and measurement technology.

However, even if social modelers were to apply the same standards of empirical rigor as natural scientists, success is not at all guaranteed. For, it may be the case that social life just does not follow any mathematical laws that are simple enough for us to understand or any mathematical laws at all. It is a contingent fact that physical nature follows laws that can be described mathematically. But there is no necessity that this will turn out to be the case for all realms of being. God has never promised that it would.

4.9 Concluding Remarks

None of the arguments discussed above appear to be particularly pervasive in the first place. Never the less I believe they are worth being discussed, because – like the previously described narratives – they help to keep the spirits of the scientists up even in face of apparent failure. Just like social prejudices they need to be made explicit to be overcome.

5 History repeats itself: Comparison with similar criticisms of naturalistic or scientistic approaches

Although this paper was mostly dedicated to the case of RPD-simulations of the evolution of cooperation, much of the criticism uttered here does not only concern this specific research tradition. In some points it overlaps with like-minded criticism of model oriented or "naturalistic" approaches in the social sciences. In this last part, I'd like to point out some of these overlaps.

In a fundamental, though still constructive criticism Green and Shapiro (1994) have described what they call the "Pathologies of Rational Choice Theory". The idea that people are by and large rational actors is in itself not necessarily connected to using mathematical models or simulations. But many of the pathologies that Green and Shapiro describe seem to be tied to a particular complex of ontological and methodological convictions lying at the base of the rational choice creed. Among these is a strong commitment to mathematical methods, which are prima facie considered to be more scientific than other methods. What is of interest in this context is what happens when these convictions are frustrated, which they must be, if on the basis of these convictions it is not possible to generate that amount of solid and empirically supported scientific results that had been promised and expected. Will the adherents of the school start to weaken or revise their fundamental convictions? Green and Shapiro (1994, 33ff.) found out that, rather then doing this, adherents of the school applied about any immunization strategy imaginable to protect their theoretical commitment. These strategies ranged form post-hoc theory development over projecting evidence from theory or searching exclusively for confirming evidence to arbitrary domain restrictions. The latter is of particular interest here, because it suggests a historical pattern that is analogous to the one observed in the history of the evolution of cooperation and which I have described as a retreat to false modesty.

According to Green and Shapiro (1994, 45) scientifically legitimate domain restriction is distinguished from arbitrary domain restriction by "specifying the relevant domain in advance by reference to limiting conditions", rather than "specifying as the relevant domain: 'wherever the theory seems to work' ". This problem has – according to their analysis – been particularly acute in the so called "paradox

23

of voter turnout", which consists in the fact that people vote at political elections even though the individual influence on the result is so marginal that any cost, even that of leaving the house for voting, should exceed the expected benefit. Now, rational choice theorists have never advanced any convincing explanation for this alleged paradox. Rather, they moved from the question of why people vote to much less ambitious explanations for turnout rate changes (Green and Shapiro, 1994, 59). And even here they did not manage to advance more than quite unoriginal hypotheses concerning, for example, the relation between education and the inclination to vote.

In two respects this resembles my results about the scientific tradition of the evolution of cooperation. First of all with regards to the triviality of the results that the simulation-based approach produced in its later stage (like my own "slip stream altruism"-story quoted above). Secondly, with respect to the stepping down from great scientific promises to such humble results. Had Axelrod believed that his simulation models have considerable explanatory power, many of his later followers (e.g. Schüßler) were so careful not to promise too much that one wonders what the simulation method is good for in the context of finding explanations for cooperative behavior, anyway. These coincidences between rational choice theory and RPD-simulations are not surprising, if one assumes that they represent typical immunization strategies of failing paradigms. One difference should be mentioned, though. In the case of rational choice it was largely an empirical failure of the theory, while in the case of the "evolution of cooperation" its was already the failure not to compare the models to empirical research.

Another connection can be pointed out between the criticism launched here and a more recent criticism of the naturalistic paradigm in the political sciences as part of the textbook on competing methodologies in social and political research by Moses and Knutsen (2012, 145-168). Moses and Knutson describe and (modestly) criticize the interconnected complex of ontological and methodological beliefs that makes up the naturalistic paradigm. This complex is composed of elements which are not unlike those that I have discussed as arguments and narratives in the two previous sections. One important element of these is the play with an assumed scientific authority (Moses and Knutsen, 2012, 157ff.). Given the many imponderables that surround any theory in the social sciences, including those

that profess to employ strictly scientific methods like formal models, Moses and Knutsen come to a similar result as I have: Namely, that this kind of professed scientism is largely a bluff.

References

Arnold, Eckhart. 2006. The Dark Side of the Force. When computer simulations lead us astray and model think narrows our imagination. (Pre conference draft for the Models and Simulations Conference, Paris, June 12-14.). In *Homepage Eckhart Arnold*, ed. Eckhart Arnold.
URL: *http://www.eckhartarnold.de/papers/2006_simulations/Simulations_-preconference.html* 13

Arnold, Eckhart. 2008. *Explaining Altruism. A Simulation-Based Approach and its Limits*. Heusenstamm: ontos Verlag. 6, 7, 10, 12

Arnold, Eckhart. 2013. "Simulation Models of the Evolution of Cooperation as Proofs of Logical Possibilities. How Useful Are They?" *Etica & Politica / Ethics & Politics* XV(2):101–138. 6, 9, 11

Arnold, Eckhart. 2014. "What's wrong with Social Simulations?" *The Monist* 97(3):361–379. 18

Ashworth, Tony. 1980. *Trench Warfare 1914-1918. The Live and Let Live System*. MacMillan Press Ltd. 7

Axelrod, Robert. 1984. *The Evolution of Cooperation*. Basic Books. 3, 5, 6, 7

Binmore, Ken. 1994. *Game Theory and the Social Contract I. Playing Fair*. Fourth printing (2000) ed. Cambridge, Massachusetts / London, England: MIT Press. 6, 10

Binmore, Ken. 1998. *Game Theory and the Social Contract II. Just Playing*. Cambridge, Massachusetts / London, England: MIT Press. 6

Dugatkin, Lee Alan. 1997. *Cooperation among Animals*. Oxford University Press. 8

Dugatkin, Lee Alan. 1998. Game Theory and Cooperation. In *Game Theory and Animal Behavior*. Oxford University Press chapter 3, pp. 38–63. 8

Dugatkin, Lee Alan and Hudson Kern Reeve. 1998. *Game Theory and Animal Behavior*. Oxford University Press. 8

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* 4, 19

Green, Donald P. and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory. A Critique of Applications in Political Science*. New Haven and London: Yale University Press. 4, 18, 20, 21, 23, 24

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. 9

Heath, Brian, Raymond Hill and Frank Ciarallo. 2009. "A Survey of Agent-Based Modeling Practices (January 1998 to July 2008)." *Journal of Artifical Societies and Social Simulation (JASSS)* 12(4):9.
URL: *http://jasss.soc.surrey.ac.uk/12/4/9.html* 21

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* 2, 8

Milinski, Manfred. 1987. "TIT FOR TAT in sticklebacks and the evolution of cooperation." *nature* 325, January:433–435. 8

Milinski, Manfred and Geoffrey A. Parker. 1997. "Cooperation under predation risk: a data-based ESS analysis." *Proceedings of the Royal Society* 264:1239–1247. 8

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. 20, 24

Osborne, Martin J. 2004. *An Introduction to Game Theory*. Oxford University Press. 8

Parsons, Talcott. 1949. *The Structure of Social Action*. first published: 1937 ed. Glencoe: Free Press.

Popper, Karl R. 1971. *Logik der Forschung*. Erstauflage: 1934 ed. Tübingen: Mohr. 3

Rangoni, Ruggero. 2013. "Heterogeneous Strategy Learning in the Iterated Prisoner's Dilemma." *Etica & Politica / Ethics & Politics* XV(2):42–57. 2, 6

Rendell, Luke, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M.W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap and Kevin N. Laland. 2010*a*. "Supporting Online Material for: Why Copy Others? Insights from the Social Learning Strategies Tournament." *Science* 328:2–53.
URL: *http://www.sciencemag.org/cgi/content/full/328/5975/208/DC1* 13, 14

Rendell, Luke, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M.W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap and Kevin N. Laland. 2010*b*. "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* 3, 13, 14, 15, 17

Schurz, Gerhard. 2011. *Evolution in Natur und Kultur*. Heidelberg: Spektrum Akademischer Verlag. 13

Schüßler, Rudolf. 1997. *Kooperation unter Egoisten: Vier Dilemmata*. 2nd (first:1990) ed. München: R. Oldenbourg Verlag. 7, 10, 11

Trivers, Robert L. 1971. "The Evolution of Reciprocal Altruism." *The Quarterly Review of Biology* 46:35–57. 19

Zacharias, Sebastian. 2013. The DarDarwin Revolution as a knowledge reorganisation. A historical-epistemological analysis and a reception analysis based on a novel model of scientific theories PhD thesis Max-Planck-Institute for the History of Science Berlin: . 3

Zollman, Kevin. 2009. "Review of Eckhart Arnold, Explaining Altruism: A Simulation-Based Approach and its Limits." *Notre Dame Philosophical Reviews* .
URL: *http://ndpr.nd.edu/news/23961-explaining-altruism-a-simulation-based-approach-and-its-limits/* 13, 18

Émile Durkheim. 1977. *Über die Teilung der sozialen Arbeit*. Frankfurt am Main: suhrkamp.