The Dark Side of the Force: When computer simulations lead us astray and "model think" narrows our imagination. (revised version, October 2006)

Eckhart Arnold

October 14th 2006

Abstract

This paper is intended as a critical examination of the question of when and under what conditions the use of computer simulations is beneficial to scientific explanations. This objective is pursued in two steps: First, I try to establish clear criteria that simulations must meet in order to be explanatory. Basically, a simulation has explanatory power only if it includes all causally relevant factors of a given empirical configuration and if the simulation delivers stable results within the measurement inaccuracies of the input parameters.

In the second step, I examine a few examples of Axelrod-style simulations as they have been used to understand the evolution of cooperation (Axelrod, Schüßler). These simulations do not meet the criteria for explanatory validity and it can be shown, as I believe, that they lead us astray from the scientific problems they have been addressed to solve.

Version history:

- October 14th 2006: revised Version
- May 31st 2006: pre-conference draft

Contents

1 Introduction **1**

2 Different aims of computer simulations in science **2**

3 Criteria for *explanatory* simulations **3**

4 Examples of Failure: Axelrod style simulations of the "evolution of cooperation" **8**

4.1 Typical features of Axelrod style simulations 8

4.2 How Axelrod style simulations work 11

4.3 The explanatory irrelevance of Axelrod style simulations in social sciences 13

4.4 Do Axelrod style simulations do any better in biology? 18

5 Conclusion **21**

1 Introduction

Computer simulations have become a popular tool in various branches of science, including even the social sciences. The reasons are easy to understand: Computer simulations provide a simple and yet powerful tool to explore the implications of theoretical assumptions. They are cheaper than experiments and often easier to construct and to handle than mathematical models. At the same time they confine the realm of what can be modeled only to what can be described algorithmically, which gives them a very broad scope. With this tool at hand it should be possible to bring into the reach of exact treatment even such questions that have traditionally seemed to defy the use of formal methods.

However, upon closer inspection it becomes apparent that computer simulations do not always deliver what they promise. Often they remain in the state of purely theoretical "toy simulations" and never get to the ground of empirical testability. In the following, I will first try to put forward a few straightforward criteria for proper explanatory computer simulations. After that I will analyze some examples of computer simulations that fail to meet these criteria and I will try to point out the consequences this failure has.

Much of what will be discussed in the following concerns the limits of formal modeling in general and not just computer simulations. Yet it seems that the question is more urgent in the case of computer simulations. For, in the case of the relatively new technique of computer simulations the sensitivity of the scientific community for the need of empirical justification or, more general, the feel for what are good practises and what are bad practises when employing this new tool has not yet grown quite mature. I have a faint hope that the following discussion of the limits of computer simulations will help to develop this sort of sensitivity, even if in some places it may turn out to be wrong or overly critical.

2 Different aims of computer simulations in science

Computer simulations can be employed in science not only for generating explanations but for various different purposes. They can, for example, be used to merely express certain theoretical assumptions or concepts. In this sense they provide a sometimes weaker and sometimes stronger but usually simpler and more flexible alternative to mathematical modeling. Or they can be used to prove the "logical possibility" of certain general assumptions such as the assumption that cooperation is possible among egoists. Or they can be used to explore the possible consequences or implications of certain assumptions. All of these previously mentioned uses of computer simulations can be subsumed under the general title of *exploratory simulations* or, as these are sometimes also called, *speculative simulations*. It is the distinctive mark of this type of simulations that the simulations do not need to resemble empirical reality. If there exists any resemblance at all then it is typically vague and consists in the plausibility of the assumptions.

Another – potentially more important – class of computer simulations are *predictive simulations*. The purpose of predictive simulations is to generate accurate predictions for some empirical process. An example might be simulations in meteorology that predict how the wheather is going to be in the future. The assumptions that enter into predictive simulations do not need to be in any way realistic. As long as the predictions prove to be reliable, it is permissible to use strongly simplified assumptions about the modeled process or even assumptions which are known to be false. This shows that just because a simulation produces successful predictions it does not necessarily also provide an explanation for the predicted phenomena, even though successful predictions may be one among several indicators for a simulation to be explanatorily valid.

The most desired case, however, would be that of an *explanatory simulation* that is a type of computer simulation that actually allows us to explain the empirical phenomena that are modeled in the simulation. It is this class of simulations that I am concerned with in this paper.

3 Criteria for *explanatory* simulations

But in what sense can a computer simulation be explanatory? And what are the criteria a computer simulation must meet in order to be explanatory?

A computer simulation can be called *explanatory* if it adequately models some empirical situation and if the results of the computer simulation (the *simulation results*) coincide with the outcome of the modeled empirical process (the *empirical results*). If this is the case, we can conclude that the empirical results have been caused by the very factors (or, more precisely, by the empirical correspondents of those factors) that have brought about the simulation results in the computer simulation.

To take an example, let us say we have a game theoretic computer simulation of the repeated prisoner's dilemma where under certain specified conditions the strategy *Tit For Tat* emerges as the clear winner. Now, assume further that we know of an empirical situation that closely resembles the repeated prisoner's dilemma with exactly the same conditions as in our simulations. And let us finally assume that also in the empirical situation the *Tit For Tat* strategy emerges as the most successful strategy. Then we are entitled to conclude that *Tit For Tat* was successful in the empirical case, because the situation was a repeated prisoner's dilemma with such and such boundary conditions and because – as the computer simulation shows – *Tit For Tat* is a winning strategy in repeated prisoner's dilemma situations under the respective conditions.

Now that we have seen how explanations by computer simulations work in principle, let us ask what are the criteria a computer simulation must fullfill in order to deserve the title of an *explanatory simulation*. The criteria should be such as to allow us to check whether the explanation is valid that is whether the coincidence of the results is due to the congruence of the operating factors (in the empirical situation and in the computer simulation) or whether it is merely accidental.

As criteria that a computer simulation must meet in order to be an explanatory model of an empirical process, I propose the following:

1. *Adequacy Requirement*: All (or at least all known[1]) causally relevant factors of the modeled empirical process must be represented in the computer simulation.

2. *Robustness or Stability Requirement*: The input parameters of the simulation must be measurable with such accuracy that the simulation results are consistent within the range of inaccuracy of measurement.

3. *Descriptive Appropriateness or Non-Triviality Requirement*: The results of the computer simulation must reflect all or at least some important features (that is features the explanation of which is desired) of the results of the modeled empirical process.

If all of these criteria are met, we can say that there exists a *close fit* between model and modeled reality. The claim I wish to hold is that only if there is a close fit between model and reality we are entitled to say that the model explains anything. Even though these criteria are very straightforward, a little discussion will be helpful for a better understanding.

Regarding the first criterion, it should be obvious that if not all causally relevant factors are included then any congruence of simulation results and empirical results can at best be accidental. Two objections might be raised at this point: 1) If there really is a congruence of simulation results and empirical results should that not allow us to draw the conclusion that the very factors implemented in the computer simulation are indeed all factors that are causally relevant? 2) If we use computer simulations as a research tool to find out what causes a certain empirical phenomenon, how are we to know beforehand what the causally relevant factors are, and how are we ever to find out, if drawing reverse conclusions from the compliance of the results to the relevant causes is not allowed?

To these objections the following can be answered: If the simulation is used to generate empirical predictions and if the predictions come true then this can indeed be taken as a hint to its capturing all relevant causes of

[1]The restriction to all *known* causes was suggested by Claus Beisbart to avoid an epistemic impassé when simulations are employed as a tool to find out just what the causally relevant factors of a given empirical process are.

the empirical process in question. With certain reservations we are then entitled to draw reverse conclusions from the compliance of the results to the exclusive causal relevance of the incorporated factors or mechanisms. The reservations concern the problem that even if a simulation has predictive success it can still have been based on unrealistic assumptions. Sometimes the predictive success of a simulation can even be increased by sacrificing realism. Therefore, in order to find out whether the factors incorporated in the computer simulation are the causally relevant factors we should not rely on predictive success alone, but we should consult other sources as well, such as our scientific background knowledge about the process in question. Also, if we already know (for whatever reason) that a certain factor is causally relevant for the outcome of the empirical process under investigation and if this factor is not included in the simulation of this process then even if the simulation predicts correctly, it cannot be said that it explains correctly.

Furthermore, drawing conclusions from the predictive success of a simulation to its explanatory validity is impermissible in the case of ex-post predictions. For, if we only try long enough, we are almost sure to find some computer simulation and some set of input parameters that match a previously fixed set of output data. The task of finding such a simulation amounts to nothing more than finding any arbitrary algorithm that produces a given pattern. But then we will only accidently have hit on the true causes that were responsible for the results in the empirical process.

Therefore, only if we make sure that at least all factors that are known to be causally relevant are included in the simulation, we can take it as an explanation. And usually we cannot assure this by relying on the conformance of the simulation results and the empirical results alone without any further considerations. Summarizing we can say: *If the first criterion is not fullfilled, then the computer simulation does not explain.*

The second criterion is even more straightforward. If the model is unstable, then we will not be able to check whether the simulation model is adequate. For, if it is not stable within the inevitable inaccuracies of measurement, this means that the model delivers different results within the range of inaccuracy of the measured input parameters. But then we can

neither be sure that the model is right, when the model results match the empirical results, nor that it is wrong, when they don't (unless the empirical results fall even outside the range of possible simulation results for the range of inaccuracy of the input parameters). Let's for example imagine we had a game theoretic model that tells us whether some actors will cooperate or not cooperate. Now assume, we had some empirical process at hand where we know that the actors cooperate and we would like to know whether they do so for the very reasons the model suggests or, in other words, we would like to know whether our model can explain why they cooperate. If the model is unstable then – due to measurement inaccuracy – we do not know whether the empirical process falls within the range of input parameters for which the model predicts cooperation or not. Then there is no way to tell whether the actors in the empirical process cooperated, because of the reasons the model suggests or, quite the contrary, inspite of what the model predicts.

A special case of this problem of model instability and measurement inaccuracies occurs when we can only determine the ordinal relations of greater than and smaller than of some empirical quantity but not its cardinal value (perhaps, because it does not have a cardinal value by its very nature such as the quantity of utility in economics[2]). In this case the empirical validation of any simulation that crucially depends on the cardinal values of the respective input parameters will be impossible. Briefly put, the morale of the second criterion is: *If condition two is not met, we cannot know whether the computer simulation explains.*

In connection with the first criteria the requirement of model stability (in relation to measurement inaccuracy) gives rise to a kind of dilemma. In many cases an obvious way to make a model more adequate is by including further parameters. Unfortunately, the more parameters are included in the model the harder it becomes to handle. Often, though not necessarily, a model looses stability by including additional parameters. Therefore, in order to assure that the model is adequate (first criterion), we may have to lower the degree of abstraction by including more and more parameters. But then the

[2]This is a well known restriction that affects a large part of the modeling done in economics.

danger increases that our model loses stability (second criterion).

There exists no general strategy to avoid this dilemma. In many cases it may not be possible at all. But this should not come as a surprise. It merely reflects the fact that the powers of computer simulations are – as one should certainly expect – at some point limited. With the tool of computer simulations many scientific problems that would be hard to handle with pure mathematics alone get within the reach of formal treatment. Still, many scientific problems remain outside the realm of what can be described with formal methods, either because of their complexity or because of the nature of the problem. This remains especially true for many areas of the social sciences.

The third criteria requires that the output of the computer simulation should reflect the empirical results with all the details that are regarded as scientifically important and not just – as it sometimes happens – merely a much sparser substructure of them. For example, we may want to use game theoretic models like the prisoner's dilemma to study the strategic interaction of states in politics. The game theoretic model will tell us whether the states will cooperate or not, but most probably it will say nothing about the concrete form of cooperation (diplomatic contacts, trade agreements, international contracts etc.) or non cooperation (embargos, military action, war etc.). Therefore, even if the model or simulation really was predictively accurate, it does at best provide us with a partial explanation, because it does not explain all aspects of the empirical outcome that interest us. In the worst case it's explanatory or, as the case may be, it's predictive power is almost as poor as that of a horoscope. The prediction of a horoscope that tomorrow "something of importance" will happen easily becomes true, because of its vagueness. Similarly, if a game theoretic simulation predicts that the parties of a political conflict will stop cooperating at some stage but does not tell us whether this implies, say, the outbreak of war or just the breakup of diplomatic relations then it only offers us comparatively unimportant information. We could also say that if the simulation results fail to capture any important features of the empirical outcome then the computer simulation "misses the point".

Summing it up: Only if a computer simulation closely fits the simulated reality – that is if it adequately models the causal factors involved, if it is stable and if it is descriptively rich enough to "hit the point" – it can claim to be explanatory.

4 Examples of Failure: Axelrod style simulations of the "evolution of cooperation"

In what follows I will to discuss a few examples of computer simulations that were designed by its authors to explain certain empirical phenomena but ultimately fail to do so. What I want to show is that these failures result from the violation of one or more the three criteria for explanatory simulations explained before.

The examples that I have chosen to discuss are computer simulations of the "evolution of cooperation" as they have become popular after the publication of Robert Axelrod's famous book with the same title. Admittedly, these examples are examples of bad simulations. But this makes them good examples. Because the failures are just the more obvious in these examples they help us understand what to avoid.

4.1 Typical features of Axelrod style simulations

Robert Axelrod's book on "The Evolution of Cooperation" Axelrod (1984) is a surprising phenomenon for two reasons: First of all, because of the extraordinary success it had as far as its impact on the scientific community is concerned. It spawned virtually myriads of subsequent studies on the repeated prisoner's dilemma (the model Axelrod used) and the "evolution of cooperation" that went more or less along the same lines and employed similar methods as Axelrod. An annotated biography from ten years after the first publication of "The Evolution of Cooperation" (Axelrod und D'Ambrosio, 1994) lists more than 200 articles that directly relate to Axelrod's study.[3]

[3]A brief overview of some of the models and simulations of the repeated prisoner's dilemma can also be found in Dugatkin's book "Cooperation among Animals" (Dugatkin, 1997, p. 24ff.)

But Axelrod's approach is also surprising for a second reason: The almost complete uselessness his and his follower's computer simulations of the reiterated prisoner's dilemma proved to have for the empirical research in the field.

How did Axelrod arrive at his results about cooperation and why did it prove so difficult to support them empirically? In order to find out, if and how cooperation can emerge among egoistic agents, Axelrod started off with a game theoretical model of a certain type of cooperation dilemma, the well known prisoner's dilemma. Since the one shot prisoner's dilemma does not offer many strategic opportunities (no rational player will ever cooperate in the one shot prisoner's dilemma, and any (non-rational) player who does fares worse than if he or she did not), Axelrod built a simulation based on the repeated prisoner's dilemma. He conducted his famous computer tournaments of the repeated two player prisoner's dilemma with strategies that he had got from many different participants. On top of the computer tournament he built an "evolutionary simulation" simulating a population dynamical process among these strategies by using the payoffs they gained in the tournament to calculate their fitness values.[4] Already at this point we may notice that the setup of Axelrod's simulation does not resemble any empirical situation whatsoever. The prisoner's dilemma itself provides a concise abstract description of the essential features of many dilemma situations that occur in reality, but nowhere in this world do we find an arrangement that really corresponds to Axelrod's computer tournament that is based on it. How are we then to draw conclusions from the computer tournament with respect to empirical cooperation dilemmas?

The way Axelrod proceeded was to examine the simulation results and to draw generalizing conclusions from them. This is how Axelrod arrived at such conclusions like: The strategy *Tit For Tat* is generally a very good strategy in the repeated prisoner's dilemma, a strategy should be friendly in the sense that it should not start to defect, a strategy should punish defection

[4]The details are not important here. There exist many descriptions of Axelrod's procedure the best of which is probably still Axelrod's own book Axelrod (1984). Simulations of the repeated prisoner's dilemma similar to Axelrod's computer tournament can easily be found on the web. For example: www.eckhartarnold.de/apppages/coopsim.html

but not be too unforgiving, the evolution of cooperation depends crucially on the continuation of interaction and the like (Axelrod, 1984, ch. 2,3). Unfortunately, subsequent research[5] showed that none of these conclusions was generally true. It suffices to change the simulation setup but a little bit and it pays to be a cheater, or to be unforgiving (Binmore, 1994, p. 194ff.). And, of course, *Tit For Tat* does not always win the race. The general finding that cooperative strategies can be successful in the repeated prisoner's dilemma as such is just a trivial consequence of the game theoretical folk theorem (Binmore, 1998, p. 313ff.). And all other generalizing conclusions Axelrod drew simply were not warranted.

Nonetheless, Axelrod's pioneering work triggered off a multitude of similar computer simulations of the prisoner's dilemma or other games. Few of their authors dared to draw such sweeping conclusions as Axelrod did. Still, regarding the design and the kind of reasoning they rely on, many of these simulations follow the pattern that was set by Axelrod's role model. In order to classify this type of simulation, we may speak of *Axelrod style simulations*.

Generally speaking, *Axelrod style simulations* are computer simulations that share the following typical features:

1. They are constructed from a set of plausible assumptions or on top of a common mathematical model. In many cases they are derived from existing Axelrod style simulations by adding new parameters or changing other boundary conditions. The concrete shape of the model remains largely arbitrary and at the discretion of the scientist who builds it.

2. They are not related to any particular empirical situation. (And most certainly there exists no *close fit* to empirical reality in the sense explained before.) Thus they remain a primarily theoretical endeavor.

3. If any conclusions are drawn from the simulation, they are usually drawn by means of inductive generalizations from the simulation re-

[5]See (Binmore, 1994), (Binmore, 1998) or (Schüßler, 1990) for a discussion of some of the subsequent research.

sults. The simulation is thus used to establish very general points or rules of thumb about its subject matter.

4.2 How Axelrod style simulations work

Let us look in more detail at a typical exponent of this tradition of simulation based research to see how Axelrod-style simulations work in practice. An in many respects good example for this tradition is provided by Rudolf Schüßler's "Kooperation unter Egoisten" (Schüßler, 1990). Schüßler called into question Axelrod's assumption that continued interaction is a necessary precondition for the evolution of cooperation. Quite the contrary to Axelrod's thesis, Schüßler wanted to show that cooperation can even emerge on "anonymous markets". In order to do so he set up is own Axelrod-style simulation where agents are free to break up the cycles of interaction whenever they want. This encourages a kind of hit and run tactic where agents do not cooperate in the last round of the interaction on their behalf and take away the benefit of single-sided non-cooperation without being punished. With the help of his computer simulation Schüßler could demonstrate that even in this case cooperative strategies could – under certain specific simulation conditions – outcompete the cheaters (Schüßler, 1990, p. 78ff.). The reason for this astonishing phenomenon is quite easy to understand: When the interaction is broken up, the previous partners of interaction are forced to pick their new partner from the pool of free players. As the cooperative players tend to be bound in partnerships by other cooperative players, the pool is made up mainly of cheaters. Therefore a cheater has only a small chance to find a new partner that can be exploited.

As can be seen, Schüßler started off with some arbitrary and at best plausible assumptions about an "anonymous market" that are in no way related to any specific empirical situation (points one and two in the above list of features of Axelrod-style simulations). But Schüßler also had a deeper motivation for his simulation experiments, which brings us to the third point: the general conclusions that are derived from the simulation results. With his simulation that showed that cooperation could even emerge on "anonymous

markets" Schüßler wanted to provide arguments against sociological normativism. Sociological normativism is by Schüßler understood as the thesis that social order cannot be upheld without social cohesion and the appeal to common norms. The classical proponents of sociological normativism are – among others – Ferdinand Tönnies with his distinction of "Gesellschaft" (society) and "Gemeinschaft" (community) and Emile Durkheim, who greatly emphasized the importance of social bonds. Schüßler's simulation is linked with the problem of sociological normativism in so far as it proves the "logical possibility" (Schüßler) of norm conformant behavior (if cooperation is taken as normatively desired in this case) even under the absence of authority or other previously fixed coordination mechanisms such as social cohesion. But does the proof of this "logical possibility" really establish a strong point against sociological normativism? This is not at all the case. The fact that something is logically possible does not even remotely imply that it is possible in reality. When sociological normativists speak for the importance of social bonds they usually do not mean to assert that it is by logical necessity that the social order requires some level of cohesion to function properly. Rather, they draw on the social character of human nature. Therefore, in order to refute them, one has to show why their conception of human nature is wrong or that the empirical support for their claims is inconclusive and could be interpreted otherwise. Claims about mere logical possibilities as they appear in the highly stylized and artificial setting of agent based simulations are notoriously weak arguments in sociological discussions. Not the least so because it would most probably be easy to draw up Axelrod-style computer simulations where under different but equally plausible boundary conditions cooperation is bound to break down when social ties are weakend.

To do Schüßler justice it must be mentioned that he is fully aware of the just mentioned explanatory limits of his computer simulations and that he discusses them frankly and with great intellectual honesty (Schüßler, 1990, p. 91f.). It is only that doing so he makes the reader wonder why he filled a whole book with computer simulations that demonstrate so little. The same questions could be asked for many of the simulations that have been done on the topic of the "evolution of cooperation". Most authors were,

like Schüßler, more careful than Axelrod in drawing sweeping conclusions from their computer simulations. But if no conclusions can be drawn from them, the question inevitably arises what these computer simulations are good for after all. It is this question that has become crucial in the case of Axelrod style simulations. In order to answer it, let us see how Axelrod style simulations fare when it is attempted to employ them in the context of an explanation of some real world phenomenon.

4.3 The explanatory irrelevance of Axelrod style simulations in social sciences

The probably most dramatic example for Axelrod's theory of the "evolution of cooperation" is given in his chapter on the trench war on the western front in the First World War (Axelrod, 1984, ch. 4). During the long phases when no great battle took place, a rather surprising phenomenon occurred on many parts of the front in this war: Hostilities lost in intensity and the number of casualties was reduced to a surprisingly small figure given the fact that the soldiers virtually eyeballed their opponents on the other side. The phenomenon has been examined in great detail by the sociologist Tony Ashworth Ashworth (1980), who found out that it was due to a kind of "live and let live" system that emerged on many (roughly one third) of the quieter parts of the front line: The soldiers hoped that if they weren't taking too hard on their enemies then the enemies would exercise the same diffidence on them. Thus, contrary to standing military orders, a kind of cooperation between the opposing front soldiers emerged on the basis of an unspoken "live and let live" agreement. Axelrod draws heavily on the description of Tony Ashworth as a source and he fully acknowledges Ashworth's achievements. Doing so he treats the "live and let live" system in the trench war as a kind of *Tit For Tat* strategy and thus regards it as an excellent confirmation case for his own theory. But would his theory really be able to explain the "live and let live" system? In order to find this out, let us see, whether Axelrod's computer simulations can add anything to the explanation of the "live and let live" system that goes beyond the explanation that is already given in

Ashworth's historical narrative. To do so we first have to briefly reconstruct the explanation that is given by Ashworth and then check whether there exist aspects of the phenomenon that Axelrod can explain better.

Ashworth, in his historical treatment, identifies the following causes for the "live and let live" system:

1. The strategic deadlock. It was virtually impossible to move the front-line for either side.

2. The natural desire of most soldiers to survive the war.

3. The unpersonal, "bureaucratic structure of aggression" (Ashworth, 1980, p. 76ff.).

4. Empathy with the soldiers on the other side of the front line.

5. Whether elite troops or non elite troops were fighting on either side. "Live and let live" was much less frequent where elite troops were involved. (According to Ashworth this was the most decisive factor of all.)

6. The "esprit de corps" that can, however, become either conductive or, in the case of elite troops, impedimental to the emergence of the "live and let live" system.

7. The branch of service. Infantry soldiers had to face a much greater danger and consequently had a greater interest in "live and let live" than artillery soldiers.

8. The limited means of the military leadership to suppress "live and let live". (Only later they found an effective way to do so by organizing raids on the enemy trenches.)

9. Initial causes such as Christmas truces, bad wheather periods when fighting was impossible, coincidental temporary ceasefire due to similar daily routines on both sides (mealtimes).

In order to apply his theory of the "evolution of cooperation" to the trench war, Axelrod first examines how the various combinations of the two alternatives of fighting wholeheartedly or fighting lackluster on either side should be estimated in terms of the assumable preference of the soldieres to survive. Doing so he comes to the conclusion that the soldiers are in some kind or prisoner's dilemma, because fighting lackluster on both sides ("live and let live") they certainly enjoy a higher chance of survival than when both sides fought wholeheartedly, although they would surely prefer to overrun the enemy if they only had a chance to do so. By furthermore interpreting the historical description of Ashworth, Axelrod goes then on to show that the prisoner's dilemma the soldiers were caught in, was indeed a repeated prisoner's dilemma. Since – according to his computer simulations – *Tit For Tat* emerges as the most successful strategy in evolutionary simulations of the repeated prisoner's dilemma, Axelrod thus arrives at his explanation of the "live and let live" system as a kind of evolved *Tit For Tat* strategy (Axelrod, 1984, ch. 4). He is aware of the fact that there is more to Ashworth's rich description than can be captured in his model. For example, Axelrod notices the evolution of an "ethics" of cooperation (due to point 6 above, the "esprit de coprs") side by side with the evolution of cooperation in the trench war. But he treats this as just another phenomenon brought about by the repeated prisoner's dilemma, not so much as another cause.

When trying to assess whether Axelrod's theory of the "evolution of cooperation" does a good job in explaining the "live and let live" system, we have to ask how many of the causes identified by Ashworth Axelrod's theory captures and how well it captures them. At first sight it would seem that Axelrod's computer model hardly captures any of these causes. If at all then only the first cause, the strategic deadlock situation the soldiers were caught in, could roughly be interpreted as a repeated prisoner's dilemma. But then, this is only one in a long list of causes, which means that Axelrod's model is far from fullfilling the adequacy requirement. I presuppose here that Ashworth has given in his book sufficient reasons to assume that all of the above listed factors do indeed play a causal role in bringing about the "live and let live" system. (In this respect Ashworth's book seems to me to be a very

solid piece of historical research, although I do not have the space here to justify my high esteem.) Axelrod could not be blamed for leaving out factors that do not really play a causal role, but a model is too be blamed if it leaves out factors we already know to be causally relevant by our background knowledge about the process in question (adequacy requirement). And the background knowledge presupposed by Axelrod is to be found in Ashworth's treatment. It would be a strong distortion of the historical situation if we were to maintain nonetheless that the soldiers cooperated in the "live and let live" type fashion, mainly because they were caught in a repeated prisoner's dilemma situation and because – as computer simulations demonstrate – "tit for tat" often is a good strategy in such situations.

However, if the model helps to give us a deeper or more precise understanding of one of the different factors that contributed to the "live and let live"-system, Axelrod's model would still have some explanatory value, if only as a partial explanation. Also, we could still try to link some of the other causes to Axelrod's model by assuming that they determine the preferences of the soldiers and thereby affect the payoff parameters of the repeated prisoner's dilemma that – according to Axelrod's interpretation – the soldiers play with their enemies. For example, it is plausible to assume that the status of the troop (elite troop or non elite troop) had a bearance on how the soldiers valuated the situation they were in. While a non elite soldier would prefer to be a coward and live, an elite soldier might prefer to fight and risk death. Consequently, elite soldiers might not even face a prisoner's dilemma. Quite in harmony with Axelrod's model, which suggests that cooperation is the rule and non cooperation the exception, this could help to explain why "live and let live" appeared only in one third of all cases.

But this still does not really rescue the Axelrod's simulation as a partial explanation of the "live and let live"-system. For, the way Axelrod proceeded when determining the payoff parameters was to asses by plausible reasoning the ordinal relations between the different alternatives for soldiers according to their assumed preferences. Unfortunately this is not enough, because the outcome of Axelrod's simulation is sensitive to the cardinal values of

the payoff parameters.[6] This violates the stability requirement. Therefore we cannot really know whether the soldiers followed the "live and let live" strategy because of what Axelrod's model suggests.

More generally, the difficulty of applying Axelrod style simulations to political or historical science results from the problem that the values of the required input parameters cannot be found ready made in the historical records. They must be reconstructed through a complicated and error-prone interpretation process. It is therefore hard to see, how the stability requirement can be fullfilled at all for simulations that are not extremely robust against deviations of the input parameters right from the beginning. As we shall see later, a similar problem applies for the application of Axelrod-style simulations in biology. Only that there we have more reason to hope that it can be overcome by simulations that are more closely knit to the measurable quantities of the empirical processes.

What then are we left with? Since Axelrod's simulation as applied to the "live and let live" system of the First World War violates both the adequacy requirement and the stability requirement, it cannot claim to be explanatory. At best it delivers us an alternative metaphorical description for the strategic situation the soldiers found themselves in in terms of game theoretical concepts. Offering no more than that it has hardly anything to add to the detailed explanations Ashworth offers within his historical narrative.

The example shows how difficult it is to make any good use of Axelrod style simulations in the social sciences. Partly, this has to do with typical difficulties that all formal approaches face in the social sciences outside economics. There are two main reasons for the limited success of formal methods in social sciences: First of all, social processes do often result from an intricate set of interwoven causes (see the example above), for only some of which we have a formal description ready at hand. But if we cannot single out the causes that can be described formally then any accuracy that is gained by

[6]To verify this, try Axelrod's evolutionary simulation with the strategies *Dove, Grim, Hawk, Joss, Random, Tat For Tit, Tit For Tat, Tranquillizer* and then change the payoff parameter R from 3 to 3.5 . In the first case (R=3) *Tit For Tat* wins, in the latter case *Dove* plays best. (The simulation software can be downloaded from: www.eckhartarnold.de/apppages/coopsim.html)

the formal description inevitably gets lost when we reintegrate the formally described causes with the other causes in a comprehensive explanation. The second reason is that measurement is difficult in social sciences and that only few quantities can be measured with accuracy. (In the above example, how would one measure the empathy the soldiers felt for the likes of them on the other side of the fontline?) It is also true for computer simulations that our formal modeling is just as good as our measurement capabilities. However, part of the reason why Axelrod style simulations fare so badly is due to the fact that it is just a very incautious type of modeling.

4.4 Do Axelrod style simulations do any better in biology?

The sceptical conclusion about Axelrod-style simulations the last section closes with becomes even more pronounced when we look at examples from biology, a field where the obstacles against formal modeling are much smaller than in social sciences. Not being a biologist myself, it would of course be difficult for me to estimate the usefulness of Axelrod style simulations for the explanation of cooperative behavior in biology. Luckily, there exists a comprehensive survey by the biologist Lee Allen Dugatkin on "Cooperation among Animals" (Dugatkin, 1997) that pays some particular attention to the manifold of game theoretical models and computer simulations that have come up in the aftermath of Axelrod's "Evolution of Cooperation". In the beginning of his book Dugatkin lists various game theoretic computer simulations and their results (Dugatkin, 1997, p. 24ff.), which – being the results of computer simulations alone – are purely theoretical of course. The major part of his book consists of a survey of the empirical research on the various instances of cooperative behavior that can be found in the animal kingdom. Interestingly, there exists not a single instance of cooperative behavior in the animal kingdom to which any of these computer simulations could be applied in a strict sense.

This is not to say that biologists did not try to do so. The attempt has been made, for example, to apply Robert Axelrod's and William D.

Hamilton's theory of the evolution of cooperation to the behavior of predator inspection that is found among various types of shoal fishes. In an early paper by Manfred Milinski on the topic (Milinski, 1987), Milinski tries to find out – with the help of an inventive experimental setup – whether pairs of inspecting fishes play *Tit for Tat* like Axelrod and Hamilton had postulated it for the repeated prisoner's dilemma. In order to do so Milinski also assesses (or rather estimates) the payoff parameters of Axelrod's model as applied to this particular case. Like Axelrod in the case of the "live and let live" system in the trench war of the First World War, he confines himself to an assessment of the ordinal relations between the payoff parameters, which, as we have seen, is unfortunately not sufficient, since the simulation is sensitive to the cardinal values of the payoff parameters. In later studies on the topic of predator inspection the attempt to explain this type of behavior with Axelrod's theory of the "evolution of cooperation" seems to have been completely dropped. In a paper that appeared ten years after the first study, Milinski and Parker, even leave the question open, whether pairwise predator inspection is an instance of cooperative behavior at all (Milinski und Parker, 1997).[7] A major methodological problem is that – despite some very ingenious experiments – it is extremely difficult to measure or to estimate reliably both the risk a fish runs when inspecting a predator and the fitness relevant payoff a fish receives from inspecting. (The former has to some degree been achieved by Milinski and Parker, but the latter remains an open riddle).

As Dugatkin summarizes the situation in the concluding chapter of his book, there exists, with one exception, no case of cooperative animal behavior where the payoff parameters required as input for the game theoretical computer models could be measured. Therefore, it is no surprise that none of the many Axelrod style simulations of the evolution of cooperation could be applied strictly to any of the empirical instances of cooperation in biology. And it is very doubtful whether this type of simulations (which remains remote from concrete empirical research and rests purely on "plausible" assumptions)

[7]For a summary of the heated debate that took place about the *Tit For Tat* strategy in biology in the meantime see Dugatkin's book on animal cooperation (Dugatkin, 1997, p. 67-70).

is of any use for biologists at all. Another leading exponent of the game theoretic approach in biology puts it the following way: "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, p. 83) And after a long discussion of problems that the study of cooperative behavior of animals faces the same expert concludes: "Most certainly, if we invested the same amount of energy in the resolution of all problems raised in this discourse, as we do in publishing of toy models with limited applicability, we would be further along in our understanding of cooperation." (Hammerstein, 2003, p. 92)

One might object that maybe some of the models can be further developed so that they actually fit some of the empirical examples of reciprocity. And there would indeed be some truth in this objection: It does not matter whether one starts constructing a model with a certain empirical application case in mind and builds it around measurable quantities (*bottom up approach*) or whether one starts with arbitrary plausible assumptions and only later on tries to adjust the model to specific empirical situations (*top down approach*). But one way or the other our models and the empirical processes they are meant to explain should be brought together. For, just because we have a model that shows us that for this or that reason cooperation evolves or breaks down, we cannot conclude in any empirical case of the evolution or breakdown of cooperation that it did so by virtue of the very same causes for which it did in the model. It could also have been the effect of quite different causes. Unless there is a "close fit" between model and reality we will never know.

The downside of computer simulations that do not achieve a "close fit" between model and reality is not only that they do not work, but that they tend to give us a wrong picture of the subject matter at hand. The "skew towards reciprocity in theoretical literature" on altruism (Dugatkin, 1997, p. 167f.) is most probably also due to the fact that the simulation business lost contact to the empirical research in this field. Instead of seeking to

achieve a "close fit" between model and reality, the tradition of Axelrod-style modeling of the "evolution of cooperation" largely proceeded a different course. Computer simulation followed after computer simulation, each of them changing the basic configuration in some way or other or trying the addition of new and different parameters. But most of these simulations never got to the ground of empirical testability. This way, however, computer simulations only lead us away from the real scientific problems.

5 Conclusion

Quite a few lessons can be learned from the previous examples of failures of Axelrod style computer models. Some of them are truisms, but as they are often neglected they are important nontheless.

First of all, if our models are to be explanatory then the establishment of a *close fit* between model and reality is at least as important as the construction of the model itself. The biological examples such as Milinski's and Parker's studies on predator inspection suggest that establishing this fit may even be much harder and more time consuming than constructing the model itself.

Secondly, when there is no *close fit* between model and reality then the model has hardly more epistemological strength than a mere metaphor. Therefore, one must be very careful when drawing conclusions from them. *Computer generated metaphors* are no better than ordinary metaphors. At best one can regard these conclusions drawn from them as mere hypotheses that still require an independent empirical confirmation. Without this empirical confirmation explanations based on computer simulations amount to nothing more than *model based story telling*. Such computer simulations are in a way comparable to non falsifyiable theories, because there is no way to test whether they simulate correctly the empirical process they are meant to simulate.

Finally, we should be aware of the fact that although the ease and power of formal modeling has been greatly increased with the advent of the computer, there still remain scientific areas where the advantages of formal modeling are doubtful or where it is not possible at all. Computer simulations

are just one scientific tool among others. It is helpful in some situations but useless in others. Where computer simulations cannot not go beyond a merely metaphorical resemblance of empirical reality their use is probably not worthwhile.

References

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

[Axelrod 1984] Axelrod, Robert: *Die Evolution der Kooperation.* deutsche Übersetzung, 5. Auflage (2000). R. Oldenbourg Verlag, 1984

[Axelrod und D'Ambrosio 1994] Axelrod, Robert ; D'Ambrosio, Lisa: Annotated Bibliography on the Evolution of Cooperation. Center for the Study of Complex Systems; University of Michigan, 1994. – URL `http://www.cscs.umich.edu/RESEARCH/Evol_of_Coop_Bibliography.html`

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

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

[Dugatkin 1997] Dugatkin, Lee A.: *Cooperation among Animals.* Oxford University Press, 1997

[Hammerstein 2003] Hammerstein, Peter (Hrsg.): *Genetic and Cultural Evolution.* Cambridge, Massachusetts / London, England : MIT Press in cooperation with Dahlem University Press, 2003

[Milinski 1987] Milinski, Manfred: TIT FOR TAT in sticklebacks and the evolution of cooperation. In: *nature* 325, January (1987), S. 433–435

[Milinski und Parker 1997] Milinski, Manfred ; Parker, Geoffrey A.: Cooperation under predation risk: a data-based ESS analysis. In: *Proceedings of the Royal Society* 264 (1997), S. 1239–1247

[Schüßler 1990] Schüssler, Rudolf: *Kooperation unter Egoisten: Vier Dilemmata.* 2. Auflage (1997). München : R. Oldenbourg Verlag, 1990