Chapter 6

Learning from failure

Ihr Instrumente freilich, spottet mein,
Mit Rad und Kämmen, Walz' und Bügel.
Ich stand am Tor, ihr solltet Schlüssel sein;
Zwar euer Bart ist kraus, doch hebt ihr nicht die Riegel.

Goethe, Faust I

We have so far been looking at several computer simulations that sought to help us to explain reciprocal altruism. We have furthermore looked at a number of empirical example cases that confirmed some of the general ideas suggested by the outcome of the computer simulations but which – at the same time – raised very strong doubts concerning the explanatory power of the computer simulations described. As any theory is only as good as its confirmation and as we certainly want to know, how good a theory of reciprocal altruism based on game theoretical computer simulations *can* be, we need to enter into some general considerations concerning the epistemology or, if preferred, the theory of science of computer simulations. The question here is a question of *can*, because as we have seen previously when looking at the concrete examples, it is a fact that so far explanations of reciprocal altruism based on computer simulations have not been successful.

6.1 Epistemological requirements for computer simulations

As has to be expected for a comparatively new scientific tool like computer simulations, the field of the epistemology of computer simulations is not very far developed. The most important epistemological question concerning any computer simulation is: How do we know that what happens in the simulation represents what happens in reality? (Of course, a simulation does not need to represent exactly what happens empirically, but it should represent what happens empirically well enough, so that we can draw conclusions from the simulation with respect to reality. So, how do we know that this is the case?) In the more technically orientated textbook literature on computer simulations (Gilbert and Troitzsch, 2005) there is little to find that could answer this question. This type of literature centers around how to program a simulation, how to visualize the data and how to debug the program, that is, it tells us how to proceed once we have decided to use the tool of computer simulations, but it does tell us little about whence and where computer simulations are an appropriate tool for investigating a certain scientific question. And astonishingly little thought is usually dedicated to the question what requirements a simulation must meet so that we can say it is a *good* simulation, i.e. a simulation that fulfills its purpose.[1]

A philosophical literature on the epistemology of computer simulations that could fill in the gap which is left open by the technical literature is only beginning to emerge. And often, unfortunately, it amounts to little more than stocktaking of what goes on the field of computer simulations, while only the surface is scratched of the epistemological questions (Hegselmann et al., 1996) concernd. A more recent example, where this is different, is Paul Humphreys' "Extending Ourselves" (Humphreys, 2004), which discusses at length the impact of computer simulations on today's scientific methodology. Regarding agent-based simulations (which is the broader category under which the simulations of the evolution of altruism presented earlier fall) Humphreys' conclusions are somewhat sceptical, as the following quotations may demonstrate:

> One of the more important questions that arise about agent-based modeling is the degree of understanding which is produced by the models. [...]
>
> In fact ... because the goal of many agent-based procedures is to find a set of conditions that is sufficient to reproduce behavior, rather than to isolate conditions which are necessary to achieve that result, a misplaced sense of understanding is always a danger. (Humphreys, 2004, p. 132)
>
> As we have seen, it has been claimed for agent-based models that one of their primary uses is exploratory, in the sense that it is of interest to show that simple rules can reproduce complex behavior. But this cannot be good advice without imposing extra conditions. [...] Because it is often possible to

[1] Troitzsch and Gilbert reserve only three pages for topic of "validation" of computer simulations (Gilbert and Troitzsch, 2005, p. 23-25).

> recapture observed structural patterns by using simple models that have nothing to do with the underlying reality, any inference from a successful representation of the observed structure to the underlying mechanisms is fraught with danger and can potentially lock us into a model that is, below the level of data, quite false. (Humphreys, 2004, p. 134)

Actually, as we have seen in the previous chapter (chapter 5), already on "the level of data" the computer simulations of the evolution of altruism hardly represented the "observed structure", let alone on the level of the underlying causal mechanisms. What is important here are the "extra conditions", which according to Humphreys must be imposed so that we do not fall prey to the "misplaced sense of understanding" that computer simulations all too easily convey. In the following I make a proposal concerning the conditions which computer simulations ought to fulfill in order to allow us a real understanding of the simulated phenomena. For this purpose, I first distinguish different types of computer simulations (section 6.1.1). Then I present and discuss a set of criteria for the most important of these types, *explanatory simulations* (section 6.1.2).

6.1.1 Different aims of computer simulations in science

Computer simulations can be employed in science not only for generating explanations but for various different purposes. In order to distinguish different types of computer simulations according to their purpose, we draw on our earlier distinction between a "conceptual level" and an "application level" of the employment of computer simulations (see page 152) and develop it by two further distinctions into a more fine-grained typology of four basic types. The two types that fall under the "conceptual level" are *proof-of-possibility simulations* and *exploratory simulations*. For the application level *predictive simulations* and *explanatory simulations* will be distinguished.[2]

[2]The broader distinction between what I have termed a "conceptual level" and an "application level" of simulations is more or less common in the simulation literature, although there is no established terminology. Kliemt, for example, distinguishes between "thin" and "thick" simulations (Kliemt, 1996, p. 15), where thin simulations correspond more or less to what I have termed the "conceptual level" and thick simulation to the "application level" in my terminology. Troitzsch and Gilbert speak of simulations that merely serve the goal of understanding a certain kind of process (Gilbert and Troitzsch, 2005, p. 15ff.) in the cases that I would describe as the "conceptual level". Just as Humphreys, I believe that this kind of "understanding" can be ever so misleading, wherefore I prefer to avoid this terminology. Also the precept – on which I draw in the recipe section (see section 6.3) – to design "conceptual level" simulations as simple as possible and "application level" simulations as acurate (i.e. as complex) as necessary is common knowledge.

The most basic type, *proof-of-possibility simulations*, are computer simulations that are merely used to demonstrate the theoretical possibility of certain assumptions or to disprove the theoretical necessity of certain commonly held beliefs. An example would be computer simulations of the evolution of altruism through group selection, which show that group selection can promote the evolution of altruism in the long run, even if altruism is always selected against within the group (see chapter 4.3.1). Typically, proof-of-possibility simulations are simple, small and not necessarily very "realistic" simulations. Such simulations are quite commonly also referred to as "toy simulations" or "toy models", which is not always meant in a pejorative sense.

Instead of proving theoretical possibilities the scientist already had in mind when constructing a simulation, computer simulations can also be employed to explore the possible consequences or implications of certain assumptions or to search for phenomena which can occur under certain theoretical conditions but which are yet unknown. Simulations that serve this purpose will be called *exploratory simulations*. Typically, this kind of simulation takes the form of large series of simulations, or, as it is sometimes called, "massive" simulations. (It should be understood that the adjective "massive" only refers to the technical complexity and does not say anything about the scientific quality of the simulation or the credibility of its results.) An example for such a "massive" simulation is the simulation series on reciprocal altruism presented in chapter 4.1.4. Just as proof-of-possibility simulations, exploratory simulations are of theoretical nature and 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 underlying assumptions.

The next class of computer simulations are *predictive simulations*. The purpose of predictive simulations is to generate true predictions of some empirical process. An example might be simulations in meteorology that predict how the weather 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. As an explanation we would accept a predictive simulation only if the assumptions built into the simulation are consistent with our background knowledge (consisting of the accepted scientific theories) about the modeled process.[3]

The most desired case, however, would be that of an *explanatory simulation*, which is a type of computer simulation that actually allows us to explain the empirical phenomena that are modeled in the simulation. From an explanatory simulation we expect that it does capture the real causes in virtue of which the modeled empirical phenomena happen. In this sense explanatory simulations are epistemologically stronger than predictive simulations. But in another sense they are not, because we do not demand from an explanatory simulation that it generates predictions. Thus a simulation may be explanatory even if it offers only ex-post explanations.[4] Explanatory simulations therefore do not form a subclass of predictive simulations.

Because the simulations of the evolution of altruism largely failed to provide substantial (i.e. not just metaphorical) explanations for the empirical instances of altruism, we will now discuss the criteria that proper explanatory simulations should meet. This will help us to understand the reasons for this failure.

6.1.2 Criteria for "explanatory" simulations

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

---

[3]It has to be admitted that this requirement rests on specific epistemological commitments concerning the generality of scope of scientific theories. I assume that if a scientific theory is well confirmed then it tells us something about anything that falls within its scope, even in cases where we have to deal with a configuration that is too complicated to analyze it in terms of the theory. If, in contrast, one follows Nancy Cartwrights "Dappled World" (Cartwright, 1999) and assumes that the validity of scientific theories is always locally restricted to its successful application cases then no conflict between predictive simulations and background theories can arise, because a successful predictive simulation that rests on assumptions that break with the background theories would then merely resemble another limit of the scope of these theories. We would then lose any ground on which we could deny the title of an "explanation" to our simulation.

[4]The motivation for allowing ex-post simulations is founded in the fact that many scientific explanations, especially in the social sciences, only work ex-post. For example, there exists a number of good explanations for the wave of democratization of the former communist states of Eastern Europe in the late 80s and early to mid 90s of the 20th century. But who could have predicted it? It would be unfair to demand from explanations that are based on computer simulations to offer more than can be accomplished by conventional science in the respective field. My criticism of Axelrod-style simulations in the context of social sciences (see chapter 5.2.2) does not rest on the charge that they provide mere ex-post interpretations but that they are far too simplistic.

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. (Probably, the easiest way to bring this about would be by conducting a game theoretic experiment, where the conditions can be closely monitored.) 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 fulfill 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 known[5] causally relevant factors of the modeled empirical process must be represented in the computer simulation.

   (This requirement is roughly equivalent to demanding that the theoretical assumptions built into the simulations should not break with or ignore our background knowledge about the modeled process, because it is only in virtue of this background knowledge that we know about the causally relevant factors of the modeled empirical process.)

   In the case of predictive simulations this first requirement would have to be replaced by the requirement of *predictive success*. A predictive simulation does not need to model the causes of a process

[5]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.

realistically. But if it does not then at least its predictions must come true.

2. *Robustness or Stability Requirement*: The input parameters of the simulation must be measurable with such accuracy that the simulation results are stable within the range of inaccuracy of measurement.[6]

3. *Descriptive Appropriateness or Non-Triviality Requirement*: The results of the computer simulation should reflect at least some important features (that is features the explanation of which is desired) of the results of the modeled empirical process. In particular, the results should not already be deducible without any model or simulation from the empirical description of the process.

If all of these criteria are met, we can say that there exists a *close fit* between model and modeled reality. What I wish to claim is that only if there is a close fit between model and reality are we entitled to say that the model explains anything. Even though these criteria are very straightforward, a little discussion will be helpful for 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 the causes of a certain empirical phenomenon are, how are we to know beforehand what the causally relevant factors are, and how are we ever to find it 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 strong hint to its capturing all relevant causes of the empirical process in question. With certain reservations we are then entitled to draw reverse conclusions from the compliance of

[6]The importance of stability is often emphasized in the simulation literature. Especially so, because there are certain types of systems (chaotic systems) for which stability cannot be achieved in principle. Often, however, stability is merely treated as a kind of internal property of simulations (Gilbert and Troitzsch, 2005, p. 23) and not, as it should be done, as a relational property between simulation and measurement capabilities which bears consequences for the epistemological strength that can be ascribed to a simulation.

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 indeed 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, we are bound to conclude that it does so only accidentally.

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 hard enough, we are almost sure to find some computer simulation and some set of input parameters that matches 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 accidentally have hit on the true causes that were responsible for the results of 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 fulfilled, 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 are 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, in spite of what the model predicts.

A special case of this problem of model stability and measurement inaccuracies occurs when we can only determine the ordinal relations of greater and smaller of some empirical quantity but not its cardinal value (perhaps, because it does not have a cardinal value by its very nature, which is the case for the quantity of utility in many contexts). In this case the empirical validation of any simulation that crucially depends on the cardinal value 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 loses 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 danger increases that our model will not be sufficiently stable any more to fulfill the 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 – limited at some point. With the tool of computer simulations many scientific problems that would be hard to handle with pure mathematics alone come within the reach of a 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 most areas of the social sciences.

The third criterion 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 (embargoes, 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 its explanatory (or, as the case may be, its 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 all (or at least the most) 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" – can it claim to be explanatory.

6.2 Reasons for failure

The establishment of criteria for explanatory simulations allows pinpointing the reasons why computer simulations of the evolution of altruism failed to explain the evolution of altruism:

1) For hardly any of the empirical instances of altruism a computer simulation existed which could be called *empirically adequate*. It is very difficult to find an empirical study of the evolution of altruism wherein recourse to a simulation model is taken. In the few instances where this was the case, it ultimately turned out to be a failure (see page 154 and chapter 5.1.3). In the sociological examples the difficulties to capture all causally relevant factors in a computer simulation were even more obvious (see chapter 5.2.2). In neither biology nor sociology, however, do the difficulties seem completely insurmountable in principle. If the right empirical example cases were picked and if the simulation models were built to fit the respective empirical instances of altruism, they might one day indeed contribute to the explanation of the evolution of altruism.

Presumably, one of the main reasons for the explanatory failure of computer simulations consists in a misconception about there being some such thing as an "in principle explanation" by a computer simulation. Robert Axelrod, one of the pioneers of the method, believed that by analyzing how and why cooperation evolves in a computer simulation that is based on sufficiently plausible model assumptions, he could devise an in principle explanation for the evolution of altruism. This explanation, he believed, could then be applied to any empirical instance of cooperation that somehow exposed a pattern of interaction that resembled his winning strategy *Tit for Tat*. It should be obvious by now that the implicit epistemological conception of explanatory computer simulations behind this belief is severely mistaken. Of course, most other authors of simulation models are far more modest about the explanatory claims they derive from their models. Rudolf Schüßler, for example, admits at one point quite frankly that his simulation models, which are similar to Axelrod's, hardly provide any decisive argument in the debate about sociological normativism to which they are related (Schüßler, 1990, p. 91).[7] But then he leaves the reader with the question what his simulations are good for, if they cannot prove any point at all.

2) Just as the requirement of empirical adequacy, the second requirement, stability, was hardly anywhere fulfilled. It should be understood that stability is a relational property between the model and its empirical application case. Except for the special case of chaotic processes, stability issues can therefore be resolved either by redesigning the model so that it reacts less sensitively to changes in parameter values or by devising more precise measurement procedures. Regarding the latter, however, it seems that in biology the problem of measuring the payoff parameters for game theoretical models poses an extremely obstinate problem (see page 154). In the social sciences this problem can to some degree be remedied if the payoff is understood in monetary terms. This is especially true for experimental economics, where the experimenter simply can pay the participants a certain amount of money depending

[7]The passage from Schüßler's book reads: "Game theoretical arguments can usually explain little empirically, but they can help to correct unfounded judgements, point out possibilities, and reduce fears of the ever looming decline of values und the stability of modern societies. How much or little that is, is a question of perspective and aptitude to make do with the art of the possible (Kunst des Möglichen)". It seems that for Schüßler game theoretical arguments do more to serve a therapeutical purpose or one of political propaganda for that matter, than a scientific one. But then it would be more logical to conclude that game theory may just not be the right tool to tackle the sort of questions that Schüßler deals with and that one should rather give other methods a try instead of confining oneself to the "art of the possible" within the narrow limits of game theoretical arguments.

on the outcome of the games played. However, as far as evolutionary models are concerned, there would still remain the problem of linking the monetary payoff to the replicator dynamics.

In some cases a model seems to be appropriate even if the parameters cannot be measured and just on behalf of the fact that the empirical process exposes a strong similarity to the modeled process on the phenomenological level. For example, grooming behavior in impala (see page 146) seems to resemble very closely the kind of interaction that takes place in the repeated Prisoner's Dilemma. Yet, because the model is sensitive to variations of the numerical values of the payoff parameters and because we cannot measure the parameter values, we cannot strictly check the validity of the model. Therefore, the model can at best be granted the epistemological status of a good metaphor in such a case.

The problem of model instability due to the use of immeasurable input parameters in the simulation models suggests that one should first consider what kinds of parameters can be measured in a given empirical situation and then try to construct the simulations around the measurable quantities. This principle could be called the *build to order principle*, because it means that the models should be build according to the restrictions and demands of empirical research just as a customer configurable product should be built according to the order of the customer. Of course, there exists a possibility of conflict between this principle and the empirical adequacy requirement in the case where certain factors which are known to be causally relevant depend on quantities which are not measurable. But then we should also consider that the underlying theory which makes use of immeasurable (hidden) factors may not be a very suitable one. (Example: Game theory which relies on payoff parameters when applied in situations where the concept of utility appears questionable.)

3) While the first requirement, empirical adequacy, is related to the input parameters of simulation models, the third criterion, *descriptive appropriateness or non triviality*, is related to the output parameters. In the case of repeated game models of the evolution of altruism the output is some kind of altruistic or non altruistic strategy. This is just what the scientist asks for when investigating altruistic behavior so that it can be granted that at least the third criteria is fulfilled for repeated game simulations of the evolution of altruism.

There are, of course, borderline cases, where even this might be disputed. In the case of the "live and let live"-system in World War One, the output of the model certainly does not capture all the nuances of the strategies that the soldiers employed to keep alive the "live and let live"-system. Most notably it does not capture the means of signaling and clandestine communication that the soldiers invented as part of their strategy. Still, as the information whether the front soldier's actions will converge to a cooperative or non cooperative equilibrium is far from trivial, it is not the non-triviality requirement because of which the simulation largely fails to explain the "live and let live"-system, but the fact that it misses many of the causes that were decisive for the evolution of this system (see chapter 5.2.2).

Summing it up, the reason why the computer simulations of the evolution of altruism failed to explain the evolution of altruism in reality, can now precisely be stated as the result of the violation of – in almost all cases – the stability criteria and additionally – in many cases – the empirical adequacy criteria.

6.3 How to do it better

If the common brand of computer simulations of the evolution of cooperation or altruism has been largely a failure, the question naturally arises how such computer simulations can possibly be done better. Turning from diagnosis to therapy, I am therefore going to to make a few proposals on what precautions must be taken when devising computer simulations so that they do not remain mere toys but become useful and valuable tools of science. For the sake of simplicity, these proposals will be cast in the form of four simple recipes, each of which covers one of the above distinguished types of simulations. Doing so, my aim is not so much to give technical advice on how to design and program computer simulations, but to give recommendations that may help to get the epistemological issues right, so that in the end the computer simulations really yield some substantial scientific results and do not remain mere toys.

6.3.1 Recipe 1: Proof-of-possibility simulations

The object of a proof-of-possibility simulation is to demonstrate theoretical possibilities. In order to assure that the proof of a theoretical possibility via a computer simulation is scientifically valuable the following steps should be taken:

1. *Does the proof of the theoretical possibility in question really contribute to answering the scientific question by which it was motivated? If not, a computer simulation may not be the tool of choice.*

Often, what is needed to be known in order to decide a certain question are not theoretical possibilities but real possibilities. But then the proof of a mere theoretical possibility bears no significance at all for the original question.

Examples of the violation of this principle:

(a) Rudolf Schüßler demonstrated with the help of a computer simulation that cooperation can evolve on "anonymous markets" without norms or enforced repetition of interaction as in the common reiterated games models (see appendix 8.5). This was meant as a contribution to the discussion about sociological normativism, i.e. the position that social order (cooperation) crucially depends on the norms of the society and the social bonds between its members. Since sociological normativists are not at all forced to deny that there exists a theoretical possibility of cooperation without norms in some arbitrary game theoretical setting, Schüßler's demonstration remains without much relevance for the original question.

(b) Michael Taylor somewhat famously demonstrated the theoretical possibility of an anarchic political order by game theoretical reasoning. Since among the many historical precedents of anarchy there exists hardly a single one where the state of anarchy was a state of order, his possibility-proof remains extremely question-begging (Taylor, 1997).[8]

(c) Somewhat similar to Taylor, Brian Skyrms employs computer simulations of the stag-hunt-game allegedly to investigate the evolution of political order (Skyrms, 2004). Again, as these abstract game theoretical models bear hardly any resemblance to any historical instances of the genesis of political order, they remain very question-begging. In contrast, the just-so-stories of 17th century social contract theorists like Thomas Hobbes draw their plausibility from the historical and political experiences they are related to, which makes them far more convincing than any of the game theoretical models.

2. *Can the same results non-trivially be derived from the background theories, anyway? If yes, there is not really a need to build a computer simulation.*

[8]The only examples that come close to Taylor's vision concern highly decentralized federal state systems which, however, are not anarchic in the sense of a more or less equal distribution of power on the level of individuals (or at least small families) or the non existence of any centers of power whatsoever.

Of course a computer simulation can in this case still serve as an illustration. Also, there may be cases, where it is not obvious how a result could be derived from the theory, so that a computer simulation may be a faster way to obtain the result.

3. *Design the simulation as simple as possible.*

   As for proof-of-possibility simulations only extremely weak empirical adequacy requirements ("plausibility") must be fulfilled, the simulation does not need to be overly complex. It should only demonstrate the possibility in question in the simplemost way and not more.

4. *Massive simulations should be avoided when only a possibility proof is needed.*

   Massive simulations may be useful to search for unknown theoretical possibilities (see recipe 2). But to merely demonstrate a theoretical possibility, running a whole series of simulation is superfluous.

5. *Don't tell stories and avoid jumping to conclusions by drawing empirical analogies.*

   If a computer simulation proves a certain theoretical possibility, say, for example, the possibility that *Tit for Tat* can be evolutionary successful in the repeated two person Prisoner's Dilemma, then it proves just that, nothing more and nothing less. It should not be pretended that the computer simulation demonstrates how Palestinians and Israelis can live in peace together or the like. To relate proven theoretical possibilities to empirical questions in a meaningful way is a matter of careful and cautious interpretation.

6.3.2 Recipe 2: Exploratory simulations

The object of exploratory simulations is to detect new theoretical phenomena or possibilities within a certain artificial setting. The epistemological and pragmatic questions involved are very similar to those involved in proof-of-possibility simulations.

1. *Is it to be expected that any theoretical phenomena will be discovered that are of scientific relevance? If not, simulations might be beside the point.*

   This is very much the same point as in the first recipe. The rationale behind this precept is that one should have some strategic goal in mind regarding what shall be achieved with the simulation. Merely toying with computer simulations is just not sufficient. It might be objected that playful behavior should have its place in science and that some of the most brilliant discoveries of science have been found by accident. But then, one can hardly base a research strategy on the hope for accidental discoveries.

2. *Use "massive" simulations and "Monte-Carlo" simulations for exploring.*

   Unlike the case of merely demonstrating a theoretical possibility, increased complexity of the simulation may pay in the case of exploratory simulations. If one has a certain idea in mind what kind of phenomena could appear, one might also employ systematic search algorithms instead of random searching ("Monte-Carlo simulation") or even evolutionary algorithms to look for the presumed phenomena.

3. *If new phenomena have been discovered, try to isolate them and demonstrate them in a simpler setting.*

   In order to understand the phenomenon, it needs to be isolated. For example, the simulation series on reciprocal altruism presented earlier (chapter 4.1.4) uncovered two "surprising" phenomena: A strong success of the strategy *Hawk*, and a more than marginal success of the strategy *Dove*. Both phenomena could then be explained by isolating them (see pages 98 and 103). In order to demonstate that *Dove* can be more successful than *Tit For Tat* even in the presence of exploiting strategies, the phenomenon was isolated in a single simpler proof-of-possibility simulation (see figure 4.16).

4. *Don't tell stories and avoid jumping to conclusions by drawing empirical analogies.*

   "Massive simulation" or "Monte-Carlo simulation" sound awfully impressive, but as long as they are not grounded empirically, they remain completely theoretical and, as has been shown at length in chapter 5, there is a certain danger that the thereby obtained results may ultimately turn out to be highly irrelevant for empirical science.

6.3.3 Recipe 3: Predictive simulations

Predictive simulations are simulations that are meant to predict empirical(!) phenomena of a certain class. Predictive simulations do not need to be realistic, as long as the predictions are successful. Because they are intended for empirical application, building predictive simulations is a much more demanding process.

1. *Clearly determine the empirical process(es) which the simulation is supposed to simulate and give an empirical specification of the input and output parameters.*

   This implies that the input parameters must be measurable (or at least determinable) quantities and not hidden factors. For example, in many empirical situations, the utility payoff assumed in game theoretical models is a hidden quantity. Often it is not even clear whether this quantity has a direct empirical counterpart at all. To avoid stability issues, the simulation should therefore be constructed around empirically interpretable and measurable input parameters that is, it should be "built to order" (see above).

2. *Assure that the stability and descriptive appropriateness requirement are met.*

   The simulation model must deliver stable results within the measurement inaccuracies of the input parameters (stability) and its output must be informative within the measurement inaccuracies of the output parameters.

3. *Calibration of the simulation:*

   In order to calibrate the simulation properly, proceed by the following steps:

   (a) Pick an empirical sample case, measure the input parameters, let the simulation generate a prediction and compare it with the empirical data.

   (b) If the simulation predicted the data correctly, it is calibrated and the calibration process is finished.

   (c) If not, revamp the simulation so that it fits (i.e. correctly predicts) the sample case. Pick a new sample case and proceed with step one. Repeat, until the simulation fits a sample case right away. When revamping, make sure that the simulation continues to fit all previous sample cases.

   Calibration can also take place ex post, as long as there are enough sample cases and the sample cases are not "used up" before calibrating is finished.

4. *Only when a simulation has been calibrated properly, which is testified by its having made at least one successful prediction, can we say that it simulates the process.*

   It is a mistake to assume that merely by revamping and tweaking a computer simulation until it fits the data of some empirical process, we get a simulation of that process. At best what we obtain is an arbitrary (and probably unnecessary complicated) algorithm to produce a certain pattern of output data. But if the simulation predicts correctly then it would be a "miracle", had it not hit upon some underlying causal structure of the simulated empirical process.

   The requirement of proper calibration may turn out to be frustrating, because in many cases we may – following the above procedure – fail to reach a calibrated simulation. But then this just means that devising a proper computer simulation is a much more demanding process than it is often thought to be. Merely fitting a simulation ex-post on some set of data is simply not enough. *Only a calibrated simulation simulates.*

6.3.4 Recipe 4: Explanatory simulations

Differently from purely predictive simulations, we demand from an explanatory simulation that it models the real causes of the simulated process. While it is desirable that an explanatory simulation should also be predictive, this is not a requirement. But if it is not predictive, its empirical adequacy must be secured by other means. To devise a truly explanatory simulation, I recommend the following steps.

1. *Check whether really all causally relevant factors of the simulated process can be rendered in a formal simulation model. If the simulation models only a substructure of the process then it must be assured that this substructure can be causally isolated.*

   Often it is only a substructure of a more complicated process that can be rendered in formal terms. For example, the strategic component of the diplomatic, economic, or – as the case may be – military interaction of nation states can in many cases be rendered in game theoretical terms. However, as the outcome of the respective interaction processes is also determined by other factors (psychological, ideological, cultural factors etc.) that cannot be rendered in formal terms, constructing too elaborate game theoretical models is probably not worth the effort.

2. *Clearly determine the empirical process(es) which the simulation is supposed to simulate and give an empirical specification of the input and output parameters.*

   Same as above.

3. *Assure that the stability and non triviality requirement are met.*

   Again, same as above.

4. *Finally, check whether the simulation results really match the empirical data.*

   If changes in the simulation are necessary to make it match the data, the question should be clarified whether these changes are consistent with the background knowledge (or, respectively, the known causal factors) about the simulated process.

6.4 Closing Words

The general morale of this chapter can be summarized as follows: Computer simulations are not an end in themselves but a scientific tool the use of which ought to depend on the scientific purpose. This means that computer simulations should be designed in view of the purpose that they are to serve and in such a way that in the end we can check whether the simulations were an appropriate means to their designated end. There may be cases where this is impossible to achieve. But then it is also doubtful whether employing computer simulations in these cases is worthwhile. The most important purpose that computer simulations can serve is that of finding scientific explanations for phenomena that appear in the real world. In order to assess whether computer simulations will serve the purpose of providing an explanation for some empirical phenomenon, I have proposed the three criteria of *empirical adequacy*, *robustness* and *non-triviality*. Having analyzed with the help of these criteria the reasons why computer simulations of the evolution of altruism largely fail to provide an explanation for why altruism evolves in nature and society, it is difficult to avoid the conclusion that the tool of computer simulations is only of limited use in this context.

However, the value of a scientific tool should not only be judged by its present usefulness, but also by its future potential. If the epistemological justification requirements are raised too high, there is a certain danger of discouraging a new approach with good prospects or rejecting a promising new scientific tool just because it does not live up to all expectations in its premature stages. Regarding this aspect, the tool of computer simulations may still become useful for the explanation of phenomena as empirical research progresses and as new experiments and measurement techniques are developed. But in order to ever become a useful tool of science it is important to have an idea of the direction into which the development of computer simulations must go. The wrong direction would certainly be to continue, as it has been done before, by basing computer simulations on plausible assumptions or on existing computer simulations through adding or changing a few parameters. Such aimless simulating just leads astray from the "real" questions of the evolution of altruism and gives a false impression of knowledge about empirical processes that in reality we do not possess. In the fashion that computer simulations have been used to study the evolution of altruism until now, they have mostly been more of a toy than a useful scientific tool.

Excerpt from:

Eckhart Arnold:

Explaining Altruism. A Simulation-Based Approach and its Limits,

ontos Verlag Heusenstamm 2008.

Abstract:

Employing computer simulations for the study of the evolution of altruism has been popular since Axelrod's book "The Evolution of Cooperation". But have the myriads of simulation studies that followed in Axelrod's footsteps really increased our knowledge about the evolution of altruism or cooperation? This book examines in detail the working mechanisms of simulation based evolutionary explanations of altruism. It shows that the "theoretical insights" that can be derived from simulation studies are often quite arbitrary and of little use for the empirical research. In the final chapter of the book, therefore, a set of epistemological requirements for computer simulations is proposed and recommendations for the proper research design of simulation studies are made.