Eckhart Arnold

Explaining Altruism

A Simulation-Based Approach and its Limits

www.eckhartarnold.de

Contents

1 Introduction **7**
1.1 The explanation of altruism as a scientific problem . . . 8
1.2 Method and central theses 9
1.3 On the structure of this book 14

2 The riddle of altruism **19**
2.1 Altruism in a hostile world 19
2.2 The definition of altruism 21

3 The generalized theory of evolution as theoretical framework **27**
3.1 The concept of Darwinian evolution 28
3.2 Biological evolution 30
3.3 Evolutionary theories of culture 33
3.3.1 Genetic theories of human behavior 35
3.3.2 Cultural evolution as a Darwinian process 40
3.4 Theory and models 65

4 Modeling the evolution of altruism **69**
4.1 Reciprocal altruism 70
4.1.1 A simple model of reciprocal altruism 73
4.1.2 Discussion of the simulation 83
4.1.3 Reciprocal altruism in cultural evolution 85
4.1.4 A more refined model of reciprocal altruism . . . 87
4.1.5 A quick look at other models and simulations of the same class 117
4.1.6 Summary and conclusions about modeling reciprocal altruism 122
4.2 Kin selection 124
4.2.1 The fundamental inequation of kin selection . . . 124
4.2.2 Transferring the concept of kin selection to cultural evolution 126
4.3 Group selection 128

4.3.1 A toy model of group selection 129
4.3.2 Extending the model? 139
4.3.3 Group selection in cultural evolution 139
4.4 Summary and conclusions 140

5 Empirical research on the evolution of altruism 145
5.1 The empirical discussion in biology 147
5.1.1 Altruism among animals 147
5.1.2 A more recent example: Image scoring cleaner fish 162
5.1.3 An in-depth example: Do sticklebacks play the repeated Prisoner's Dilemma? 165
5.2 Empirical findings in the social sciences 170
5.2.1 Laboratory experiments 171
5.2.2 A real world example: Altruism among enemies? . 180
5.3 Conclusions 189

6 Learning from failure 191
6.1 Epistemological requirements for computer simulations . 191
6.1.1 Different aims of computer simulations in science 193
6.1.2 Criteria for "explanatory" simulations 195
6.2 Reasons for failure 200
6.3 How to do it better 203
6.3.1 Recipe 1: Proof-of-possibility simulations 203
6.3.2 Recipe 2: Exploratory simulations 205
6.3.3 Recipe 3: Predictive simulations 207
6.3.4 Recipe 4: Explanatory simulations 208
6.4 Closing Words 209

7 Summary and final reflections 211

8 Appendices 217
8.1 Strategies for the reiterated Prisoner's Dilemma 217
8.1.1 Ordinary strategies 217
8.1.2 Parameterized *Tit for Tat*-strategies 221
8.1.3 Two state automata and their implementation . . 222
8.1.4 The family of *Signaling Cheater* strategies 224
8.2 Implementation details of the population dynamics . . . 225
8.3 Comprehensive results of the simualtion series 230
8.3.1 "Big series" overall results 231
8.3.2 The influence of correlation 235
8.3.3 The influence of game noise 243
8.3.4 The influence of evolutionary noise 251

8.3.5 The influence of degenerative mutations 261
8.3.6 The influence of different payoffs 269
8.3.7 "Monte Carlo series" results 279
8.4 Implementation details of the group selection model . . . 283
8.4.1 Listing 1: The deme class 287
8.4.2 Listing 2: The super deme class 289
8.4.3 Listing 3: A deme class for Prisoner's Dilemma players 290
8.5 Cooperation on anonymous markets: A simplified version of Schüßler's model 291
8.5.1 Listing: Beispiel_Schuessler_1.py 294
8.6 Backward induction as an evolutionary process 296
8.7 The simulation software and the full simulation results on DVD. 300
8.7.1 The simulation programs 300
8.7.2 Browsing the results of the simulation series . . . 301

Chapter 1

Introduction

In this book I examine evolutionary explanations of altruism that are based on computer simulations. When speaking of *explanations* of altruism, this means that this book is not primarily a study that tries to explain altruism itself, but a critical examination of how these explanations work. Its aim is twofold: On the one hand, it will expound this type of explanations of altruism, describe its working mechanisms and the results that can be obtained. In this respect this book strongly draws on the simulation based approach to the evolution of altruism that was pioneered by Robert Axelrod and William D. Hamilton (Axelrod and Hamilton, 1981) and made popular through Axelrod's book on "The Evolution of Cooperation" (Axelrod, 1984). However, after the more than twenty years that have passed since the publication of this book, the fact can hardly be ignored that the simulation-based approach to the explanation of altruism did not quite live up to the very expectations and aspirations that it once gave rise to and to the "simulation hype" it caused. Therefore, this book will on the other hand broadly discuss the limits of this approach. My aim is to give a clear diagnosis of this failure, to explain why this approach remained largely unsuccessful and also to point out what lessons regarding the research design of computer simulations can be learned in order to allow a more purposeful employment of computer simulations for scientific explanations in the future.

In this introduction, I first say a few words about the topic and theoretical background that is, about why the evolution of altruism is a topic that interests us, why an evolutionary approach may be suitable to tackle the question of altruism and, finally, how computer simulations come into play here. Then, I briefly explain my method for examining the simulation-based evolutionary explanations of altruism and its alleged failure. Basically, my method consists in conducting some simulations in the common fashion of this approach myself and looking at

the corresponding empirical research both in biology and in the social sciences. I also give in this introduction a very brief overview of the main results of my inquiry. Finally, I present the structuring of this book and, in this context, further describe the methodological decisions I have taken.

1.1 The explanation of altruism as a scientific problem

The explanation of altruism poses an intriguing riddle both in biology and in the social sciences. In biology the question is how, if "survival of the fittest" is the rule, altruistic behavioral traits can evolve when altruism means by definition the giving-up of some of an organism's own fitness in order to increase the fitness of another organism. Yet, as ants, honeybees or the behavior of brood care in almost any species testify, altruism does exist in nature. How then, did it arise?

Similarly, while we all believe that humans are moral creatures that can by proper education and appropriate incentives learn to behave as altruists, the question still remains why, if – as we observe in many areas of life – egoism is the road to success, altruistic norms should continue to enjoy a high and general esteem. Should not a lack of secular success of the adherents of altruistic norms mark such norms as unrealistic if not foolish?

Moreover, altruism raises not only important questions in the empirical sciences, but also for moral philosophy and metaphysics. For, when we postulate altruistic moral norms we surely want to know (if we are not pure *Gesinnungsethiker*) whether and to what degree we can realistically expect obedience to these norms. From a metaphysical perspective the question of the viability of altruism links to the old question of whether the world as a whole is good or bad and, if bad, whether it can be made any better or if we will have to cope with the fact that "the realm of virtue is not from this earth" (Schopenhauer, 1977).

Thus, the existence of altruism demands an explanation and the desirability of altruism calls for an understanding of the circumstances under which altruism can flourish. In this book an examination will be made as to what an evolutionary simulation-based approach can contribute to the understanding of altruism.

1.2 Method and central theses

Why use an evolutionary approach for the explanation of altruism? In biology the answer to this question is obvious: Any phenotypic trait of any organism must – according to Darwin's theory of evolution – have evolved through natural selection. If a certain organism or species exposes an altruistic behavioral trait then there must be an evolutionary explanation for it. The situation is different in the social sciences. As is usual in the social sciences there exist many competing paradigms upon which a scientist could draw in order to explain the genesis of social norms, including norms that prescribe altruistic conduct. The evolutionary theory of culture which seeks to apply the principles of the Darwinian theory of evolution (reproduction, variation, selection) to the evolution of cultural traits is a comparatively young contender. Its practical value for the social sciences is still disputed[1] and, due to the fact that there exist many good alternative explanations for cultural developments, it would be too much to expect that the evolutionary theory of culture could repeat in the social sciences the very success that Darwinism had in biology. Yet, there are some good points in favor of it. First of all, the evolutionary theory of culture may prove able to explain things that other theories of cultural developments cannot explain.[2] Then, where it proves able to explain cultural developments, it most probably can provide general patterns of explanation that can be applied both in biology and in social sciences. If the evolutionary theory of culture should prove to be successful then it could be regarded as a great advance in terms of the economy of knowledge. Finally, explanatory patterns that cover different areas of research may profit from synergistic effects, which means that an advancement of modeling or empirical research in one of the fields may carry over to the other fields.

However, there are also downsides to such a generalizing approach. Most notably there is the danger of overlooking peculiarities of the respective areas of research and, as always with generalizing, there is the danger of oversimplification. Ultimately, the choice to use an evolutionary approach to study altruism is – as far as the social sciences are concerned – to some degree a matter of preference and motivated by the desire to find an explanation for altruism as broad as possible.

Given that it has been decided to use an evolutionary approach to study altruism the next question would be why computer simulations should be employed to furnish the evolutionary research on altruism. In

[1] See Bryant (Bryant, 2004) for a fundamental criticism of the evolutionary theory of culture.

[2] See Arnold (Arnold, 2002) for some speculations on this topic.

principle, there would be four different alternatives: 1) One could rely on purely verbal reasoning to explain the evolution of altruism. But then, verbal evolutionary explanations tend to be notoriously weak. It is almost always possible to construct some sort of evolutionary story of why some certain trait had to evolve and often it is just as easy to explain on the same level why its opposite should have evolved (even if in fact it did not) if only because it is usually easy to feign some plausible selective conditions under which the trait in question would be advantageous. 2) Another alternative is mathematical modeling. It allows – as one should presume – for a very precise expression of the concepts in question, but it can easily become extremely complicated and tedious, once it rises above the mere expression of the concept of, say, reciprocal altruism to models that can halfway realistically depict a situation in the real world where altruism evolved.[3] 3) The latter problem can potentially be addressed by numerical models, which class includes also the computer simulations of altruism. Computer simulations are an extremely flexible, easy to use and powerful tool. Of course all computer simulations rely on mathematical background theories such as, for example, evolutionary game theory. In this sense there does not really exist an opposition between computer simulations and mathematics but rather a dependency. But with computer simulations it is easily possible to go beyond what can be modeled in purely mathematical terms. Because of their ease and power, computer simulations seem to have been regarded by many as the tool of choice for the study of the evolution of altruism. 4) Last but not least, there is the empirical approach to altruism, which roughly means looking at empirical instances of potentially altruistic behavior and drawing inferences about these by means of common reasoning.

In principle, the empirical approach should not be regarded as an alternative to the theoretical approaches described above. For, any systematic empirical research must be guided by theories or at least theoretical preconceptions about the subject matter. In turn, the models and theories should of course be tested against empirical data. However, in practice there really exist two approaches with quite a different style and flavor to each of them. The empirical approach is a "bottom up" approach, where scientists start with empirical observations and gradually develop more and more complex models to account for them. The the-

[3]See Boorman and Levitt (Boorman and Levitt, 1980) for a comprehensive treatment of the mathematical modeling on the genetics of altruism. It seems that Boorman and Levitt received comparatively little attention in the philosophical literature on the evolution of altruism. This may be due the difficulties for most readers to understand the mathematical presentation or to the fact that computer simulations of altruism have become so popular in the meantime.

oretical approach (as opposed to the empirical approach) is what could be called a "top down" approach, where scientists start with theoretical considerations and models and then (hopefully) adjust them to the empirical instances that these are to be applied to. Unfortunately, in the case of the research on altruism there exists a wide gap between the theoretical and the empirical research[4]. From the vast amount of computer simulations on altruism produced, hardly any has been successfully applied in empirical research. Partly, this gap is due to the division of labor in science, where one group of scientists develops the models and another group does the empirical research. But this alone cannot explain why there is such a lasting discrepancy between the computer simulation based theories and the empirical research.[5] The discussion of this problem, the understanding of its causes and the consequences that should be drawn form the central topic of this book.

In the course of this book, I look at both computer simulations and empirical research in order to examine this question. Purely mathematical models of altruism will not be discussed. The reasons for leaving them out are primarily of pragmatic nature. The epistemological questions concerning mathematical models are not exactly the same as those concerning computer simulations, although presumably many of the results about the epistemology of computer simulations arrived at in this book will also hold true for purely mathematical models. Also, the just mentioned problem of a strong discrepancy between theoretical modeling and empirical research in the study of the evolution of altruism seems to be even more glaring in the case of computer simulations if only because the use of computer simulations makes modeling much easier and more powerful so that the mere popularity of this tool has exposed dangers that are already imminent in purely mathematical modeling.

In order to better understand how computer simulations of the evolution of altruism work, several simulations and simulation series in the Axelrod-fashion will be carried through to simulate different kinds of altruism. There are basically three different kinds of altruism: Reciprocal altruism, kin selection and group selection. Most simulations will be done on reciprocal altruism and some on group selection. For the sake of completeness, kin selection will also briefly be discussed but not simulated. Although they are intended to illustrate the use of a certain method rather than to be particularly original, the simulations presented here are new in the sense that they are not merely repetitions of computer simulations that have already been carried out and described in the scientific literature on the subject. It is, however, one

[4]See Dugatkin (Dugatkin, 1998) for a discussion of this problem.

[5]See Hammerstein (Hammerstein, 2003a) for a vivid depiction of this discrepancy.

of the main points to be established in this book that *the results of such purely theoretical simulations (be they as new or unique as they may) are typically not of great scientific relevance.*[6]

Just how irrelevant very many of the models of reciprocal altruism are becomes obvious when they are held against the empirical research on altruism. No empirical research has been done specifically for this book. Instead I review some of the empirical research that has been done in biology and in the social sciences, especially in behavioral economics. Not being a specialist in either biology or economics I am quite aware of the dangers involved with reviewing the results of branches of science that one can at best claim to have a laymans knowledge of.[7] The dangers include misunderstanding, misrepresenting, mistaking the inessential for the essential etc. But these are problems that any kind of interdisciplinary research faces. The only secure way to avoid these dangers would be to refrain from interdisciplinary research altogether or to ignore scientific results in philosophy, neither of which can seriously be considered an option. To the extent to which the more recent scientific research in the two above mentioned fields has found its way into textbooks it is still fairly easy to access. Therefore, I have tried as far as possible to rely on this kind of scientific literature. However, the latest research can only be found in articles in scientific journals. As far as these are concerned, I can only say that I have tried to report the content of the articles that I have quoted as faithfully and accurately as I could as a layman.

Having shown by examining the empirical research that computer models of the evolution of cooperation or altruism can tell us only very little about how altruism evolves, this naturally raises the question why they failed to do so. My answer to this question, which is at the same time my central thesis, generalizes from the simulations of the evolution of altruism and states that *the main reason why computer simulations often fail to fulfill their expectations in science is that the epistemological conditions under which they can possibly explain or prove something are not yet well understood.* Computer simulations are still a relatively new tool in science so that "best practices" for their design or employ-

[6]The reason why I do not think they are is explained in chapter 4.1.6.

[7]My field of specialization is political science. Regarding political science, however, I seriously doubt that computer simulations of the evolution of cooperation can provide us with any important insights beyond mere trivialities. See Arnold (Arnold, 2005a) for an extensive criticism of this approach, which also contains *in nuce* some of the arguments that have been expounded in greater detail in this book. In this scepticism regarding the value of mathematical models for political science I feel strongly confirmed by the criticism of the rational choice approach as applied to the political sciences by Ian Shapiro and Donald Green (Green and Shapiro, 1994; Shapiro, 2005), which unfortunately I had not been acquainted with at the time of writing this book.

ment are only beginning to emerge. There still seem to exist quite a few insecurities as to how computer simulations can be used properly in the context of scientific explanations. At any rate, the "tradition" of Axelrod-style simulations of the evolution of cooperation seems to have gone astray if the aim really was to explain how cooperation or altruism evolves. That a whole school or "tradition", if I may call it so, of science is going amiss may be due to the fact that the very business of science sometimes proceeds in an astonishingly naive if not narrow-minded way. In this case, Axelrod had set with his computer simulations a seemingly successful new role model for the study of the evolution of cooperation. What could have been more advisory for aspiring scientists in this field than to pick up Axelrod's model, change it here and there a little bit or even challenge it by designing a similar model that would lead to divergent conclusions and thus produce fascinating new results about the evolution of altruism? And it was so easy: One only needed to know a little bit about computer programming and one could do research on "the evolution of cooperation". (Even philosophers could do that!) Now, the naivety with which science sometimes proceeds – and it certainly proceeded too naively in this case – is to some degree to be excused because if one wants to examine some subject matter one cannot for (economical reasons) at the same time occupy oneself too much with the examination of the method of the examination of the same subject matter. But if this is true then it surely is a philosopher's job to make up leeway and to reflect on what science does and whether it does right what it does. Therefore, the final and most important part of this book is dedicated to the discussion of the epistemological conditions under which computer simulations can be used in the context of scientific explanations. Just as we demand from ordinary scientific theories that they be empirically testable before we grant them the honorable status of a "scientific" theory (that is a theory that can potentially explain certain empirical phenomena), *we need criteria for computer simulations that allow us to classify computer simulations into those for which it can (empirically) be decided if they simulate correctly and those for which this cannot be done.* The criteria I am going to propose in this book are those of *empirical adequacy*, *robustness* and *non triviality*. "Empirical adequacy" means that all causal factors that have a significant impact on the outcome of the simulated process are somewhere represented in the simulation. "Robustness" requires that the output of the simulation is stable within the range of measurement inaccuracy of the input parameters. And "non triviality" simply requires that the output of the simulation gives us some important information about the outcome of the simulated empirical process. (The last criteria may seem trivial or

self-evident itself, but unfortunately experience has shown that this is not the case.[8]) These criteria raise the bar for "explanatory simulations" quite high and it will be discussed at some length if such strict criteria are really necessary. But if they are more or less accepted then it follows that the sort of example simulations that have been presented in this book to demonstrate the principle of Axelrod-style simulations, and with them very many of the simulations published in the literature on the evolution of altruism must be rated as insufficient if any explanatory claim would be based on them. This is quite in accordance with the lack of empirical success of the simulation-based approach to altruism mentioned earlier. But with the above mentioned criteria at hand we can better understand just why most of the computer models of the evolution of altruism had to fail.

Once the epistemological conditions for the proper application of computer simulations in an explanatory context are well understood, it is not only possible to soundly criticize the misguided use of computer simulations. It is just as well possible to derive guidelines of how to design and use computer simulations properly. In order to supplement the critical discussion of what I consider to be a failure of computer simulations with a positive outlook for the future, I offer my own proposal for such guidelines in form of a few simple recipes that scientists can follow if they want to be assured that their simulations are epistemically valid.

1.3 On the structure of this book

The book is organized into four parts. In the first part (chapter 2 and chapter 3) I explain why the existence of altruism, which is a fact of the natural as well as the social world, poses a scientific and philosophical problem. Furthermore, I give a definition of altruism that is broad enough for both biology and the social sciences and I justify this definition at some length. The first part closes with an exposition of the "generalized theory of evolution" (Schurz, 2001), which constitutes the greater theoretical context into which the following models of the

[8]To me it seems that the sort of computer simulations that Brian Skyrms devised for the study of the "social contract" (Skyrms, 1996) or "social structure" Skyrms (2004) are trivial to a point where they must be regarded as mere toys. It would be very difficult to draw from his simulations any tenable conclusions with regard to the subject matter of political order (social contract) or social structure that they are allegedly related to. (For a criticism of Skyrms see Arnold (Arnold, 2005a).) A similar objection holds for Schüßler's simulations of cooperation on "anonymous markets", only that Schüßler is at least aware of the problem and honest enough to discuss it (Schüßler, 1990, p. 91f.).

evolution of altruism can be integrated.[9] Because the application of evolutionary theory outside the field of biology is a controversial issue, the different flavors of theories of cultural evolution will be discussed at some length.

In the second part (chapter 4) the three basic evolutionary explanations of altruism will be explained and the modeling on the evolution of altruism will be discussed. The presentation of a whole field or branch of science always raises a certain methodological question: Should one rather give an extensive but in its details necessarily sketchy overview over the whole field or should one present and discuss a few select examples "pars pro toto" in all detail. I have taken the second approach and will present a few self-made computer simulations in order to demonstrate how this type of modeling works in detail. Of course, I could also have taken models that were described in articles in scientific journals. But usually the description in journal articles does not present all the details of a simulation, hardly ever is the source code of the simulation software given and often the information is too sketchy to reconstruct the simulation in an unambiguous way. Also, programming simulations on one's own is quite an instructive exercise. It allows one to notice how many ad hoc decisions enter into the construction of a simulation. By presenting the computer simulations and their results in detail it will be possible to point out both the usual working mechanisms of such simulations as well as the common traps and pitfalls of simulations. The description of these (as I hope) paradigmatic example simulations will be supplemented by a review of a selection of the simulations of the evolution of altruism published in the respective literature. The discussion will cover all forms of evolutionary altruism that is, reciprocal altruism, kin selection and group selection. The greatest emphasis is laid on reciprocal altruism as this is the type of altruism for which the method of computer simulations has been used the most excessively. As will become apparent from the discussion of the simulations conducted by myself as well as those published in the literature on the subject, there is an arbitrary large space of logical possibilities that could be explored by simulations while at the same time hardly any generalizable results can be derived from simulations alone. The reason why all three forms of altruism are covered even though reciprocal altruism would arguably have sufficed to prove the point against the method computer simula-

[9]Of course the models of the evolution altruism do not necessarily need to be understood in the context of a *generalized* theory of evolution. For example, as long as we only talk about altruism among animals it would suffice to interpret them against the background of the theory of evolution in biology. But as evolutionary explanations of altruism can be given both in biology and in the social sciences a generalized theory of evolution that does not confine itself to genetic evolution alone provides a very suitable paradigmatic background.

tions is that these different forms of altruism do often not appear strictly separated in the empirical literature on the subject (if only because it is often very difficult to tell apart the different forms of altruism in an empirical context) and it would otherwise be difficult to compare the simulation studies with the empirical research.

In the third part (chapter 5) of this book the results of the computer simulations will be contrasted with the empirical research on the evolution of altruism. It is here where it becomes most obvious that a wide gap exists between the simulation research and the assumptions about the evolution of altruism based on it on the one hand and the empirical research on the other hand. Again, when presenting the results of the empirical research on the evolution of altruism, a similar methodological issue as in the case of the presentation of the simulation research arises. Should one rather give a broad overview of the research or should one discuss only a few exemplary studies in detail. I have tried to combine both approaches and therefore give a broad – though for the sheer size of the topic necessarily incomplete – overview of the empirical research (in biology) first. This way the fact can be assessed that cases where empirical researchers could make good use of the results of simulation studies on the evolution of altruism are extremely rare. In order to understand just why they are so rare, I pick out some examples (both from biology and from social sciences) and discuss them in detail. Since I am going to make a case against the simulation based approach, I was careful to pick out examples that could (at the time of their publication) be considered as showcases for the application of the results of simulation based research to empirical problems of the evolution of altruism. If these fail then the simulation based approach in its present form is confronted with a serious problem. And they do fail, as I hope to be able to demonstrate.

Turning from the diagnosis of failure in the third (and partly already the second part) of this book to the explanation of the failure in the fourth part (chapter 6), I propose and discuss the above mentioned criteria for "explanatory simulations". It can easily be seen that hardly any of the simulations on the "evolution of cooperation" meets these criteria. It is more difficult to show that the fulfillment of these criteria is both necessary and sufficient for a computer simulation to claim explanatory power in a scientific context. Since the epistemology of computer simulations is a relatively young field in the philosophy of science with many open questions, I can hardly maintain to have found the definite answer to the question of potentially explanatory qualities of computer simulations. The fourth part therefore has more or less the character of a philosophical discussion that is, I try to defend these criteria as good

as possible against conceivable objections. Given that the proposed criteria provide at least a reasonable guidance, I finally turn to practical considerations and try to devise some "recipes" for the proper use of computer simulations in a scientific context.

In a short concluding chapter the results of this book will be summed up. The main results are that the simulation based approach to the study of the evolution of altruism was largely a failure. This failure resulted from a lack of understanding of the epistemological conditions and requirements of the employment of computer simulations in the context of scientific explanations. Yet, if carefully applied, computer simulations can be a very valuable tool of scientific research. Regarding the requirements of "good" computer simulations, I have made a few proposals in the last part of my book. These may or may not prove sufficient and practical in the future, but if I was able to convey a sense for the necessity to take epistemological considerations into account for a proper research design of simulation based research, then my attempts have not been wholly futile.