Chapter 5

Empirical research on the evolution of altruism

The last chapter closed with the conclusion that substantial scientific results about the evolution of altruism cannot be obtained by looking at computer simulations alone. The situation would be different if there were only one right way to model altruism. But because there are so many plausible ways to do it only a look at the empirical examples can tell which one is the right one. In the following we will therefore examine some of the empirical research on altruism. We will first look at biology and then at the social sciences. When surveying the research in these fields, there are two questions that are important for us: First of all, we do of course want to find out whether, how and why altruism evolves in nature and among humans. Theoretical models and computer simulations demonstrate how it *could* evolve. Empirical research, hopefully, can tell us something about how, why and where it *does* evolve. The second question concerns the method and research strategy. Already in the previous chapter there has been opportunity to raise some doubts concerning the usefulness of the tool of computer simulations for the understanding of reciprocal altruism. Now we want to know how these simulation models live up to the empirical research, that is whether they are helpful for conducting such research and whether they prove valuable for the explanation of the results of the empirical research.

A survey of empirical research on the evolution of altruism raises certain methodological issues by itself, which shall briefly be discussed, before entering into the discussion of the empirical material. First of all, there is the question of the selection of the material. As the research on altruistic behavior is a wide and varied field both in biology and in the social sciences and as the focus of empirical scientists and the categories they employ are often not the same as those the theoreticians develop, a selection of materials is unavoidable. In the following, I have tried to

choose examples that are most closely linked to the theoretical models and to the concepts of reciprocal altruism, kin selection and group selection described earlier. This criterion of selection also has advantages for addressing our second question, the question of the usefulness of simulations as a method. For, if this method fails in those cases that we would assume it is best suited to deal with, then we have good reason to assume that it is a bad method (at least in the way it is applied today) without worrying that we might have been unfair. Still, it must be admitted that the following selection of empirical example cases is quite eclectic. This is unavoidable given the sheer extent of this field of research, but – as should frankly be admitted – it is also partly due to the fact that I am neither an expert in biology nor in experimental game theory.

Another methodological issue when surveying research, concerns the question as to whether one should give a broad overview covering as much of the research as possible or whether one should rather pick out a few examples and discuss them in depth in order to demonstrate how the respective kind of research works and what degree of credibility can be attributed to it. Regarding the biological examples, I have tried to combine both approaches. First, an overview of a larger number of empirical studies on reciprocal altruism will be given to convey an idea of where this research stands. Then, one example will be picked out and discussed in depth to see how reliable the results of this research are and especially how well the theoretical models do when submitted to the "on-road test". For the social sciences I confine myself to the discussion of a few select examples. The reason for this is that while there exists a lot of empirical research on cooperation dilemmas of one kind or other, there are hardly any empirical studies that are closely attuned to the kind of models that have been discussed before.[1] It would be spurious to present a summary of research on behavioral economics that mostly falls outside the narrower topic of this book.[2] But just as in the case of biology, one of the examples from the social sciences will be discussed in depth. For the in depth discussion I have in both cases picked examples that were by their authors intended as show cases for the application of reiterated Prisoner's Dilemma models. Therefore, these examples should be best suited to assess the possible merits and defects of this

[1]This is even true for Axelrod's popular model of reciprocal altruism, which has spurred myriads of further model studies (Dugatkin, 1997, p. 24ff.), but remained quite infertile for the empirical research.

[2]A fairly recent overview of the research on altruism in experimental economics can be found in (Fehr and Fischbacher, 2003). The bulk of this research is concerned with the question how altruism works among humans. While this has some bearing on which kind of evolutionary explanations are more plausible than others, only few evolutionary models seem to be have been put to the empirical test directly.

type of modeling.

5.1 The empirical discussion in biology

5.1.1 Altruism among animals

As in any other field of science the specialist literature on altruism in biology comes in two different brands. First of all, there are articles in different biological journals. Then, there are books on the topic written by specialists that usually present the results of the research published in articles in a condensed and simplified form. For a non-specialist it is advisable to stick to the latter kind of literature, for otherwise there exists a considerable danger of misunderstanding and of giving too much weight to unimportant details and too little weight to important ones. Luckily, there exists a treatment of the subject in book-form by an author who is strongly committed to a game theoretical approach to the study of altruism. This treatment is Lee Allan Dugatkin's already afore mentioned "Cooperation Among Animals" (Dugatkin, 1997). In what follows I therefore present mostly examples from Dugatkin's book. Unfortunately, the book was issued in 1997 and therefore does not cover the latest research. For this reason, later on I also discuss an example of a study that has been published on the topic since.

The empirical research which Dugatkin reviews, cannot always be sorted neatly into different categories of altruism like reciprocal altruism, kin selection or group selection. The reason for this is that when scientists set out to research altruistic behavior in certain animal species they usually are not sure beforehand what kind of altruism is concerned. And quite often the data they are able to obtain does not allow making the distinction afterwards. Often it is not even clear whether the behavioral trait in question is altruistic at all or merely some kind of byproduct mutualism.[3] In the following, different examples of cooperative and potentially altruistic animal behavior that are described in Dugatkin's book will be presented. The main aim is to clarify whether the theoretical categories for altruistic behavior (reciprocal altruism, kin selection and group selection) can be identified empirically and to what degree assumptions about the type of altruism can be ascertained. Also,

[3]The difference between altruism and byproduct mutualism is that while both entail benefits for some other individual, it must in the case of altruism at least be possible to cheat, while in the case of byproduct mutualism cheating is impossible in principle that is, an exchange of benefits still may or may not take place, but if it takes place cheating is not an option. An example to illustrate this might be two people warming each other in winter by moving closer together. None can enjoy the warmth of the other without giving warmth him- or herself, which means that there is no way to cheat.

it will be asked in how far models such as those presented in the previous chapter can be validated empirically and whether and in how far these types of models have been useful to empirical research.

Cooperative behavior as it occurs in nature

Egg Trading An often quoted example of reciprocal altruism in particular is that of egg trading among hermaphroditic fish. According to Dugatkin it is best documented for sea bass (Wolfsbarsch) (Dugatkin, 1997, p. 46). Sea bass (as well as many other egg trading fish species) parcel their eggs into small packages. When mating, one fish starts by releasing a parcel of its eggs, which typically consists of only a small fraction of the eggs it has. At the same time the partner releases sperm. Then they switch roles and regularly alternate the release of eggs subsequently. These cycles of alternating egg spawning suggest an interpretation of this process as a repeated game. But is the game a Prisoner's Dilemma and do the sea basses use a reciprocal strategy, i.e. would they retaliate if being cheated? Dugatkin's answer is that it can loosely be interpreted as a repeated Prisoner's Dilemma if the release of one parcel of eggs by one partner and the following release or failure of release by the other partner is interpreted as one round of the repeated game and if it is assumed that producing eggs is more expensive than producing sperm. Although it is difficult to quantify the costs, the latter assumption is almost certain to be true (Dugatkin, 1997, p. 48). A problem is that due to the lack of quantitative data (and – as of now – the lack of measurement techniques to obtain such data), it is impossible to fill in the payoff matrix of the game other than by rough estimates. But then it is not even sure whether *Tit for Tat* is a suitable equilibrium strategy. Regarding the question whether fish engaged in egg trading do in fact play *Tit for Tat*, there exists, according to Dugatkin, some anecdotal evidence (i.e. non-systematic evidence from incidental observations) for certain types of fish that they do in fact play some deviant version of *Tit for Tat*. It is reported that black hamlets and chalk basses retaliate by waiting much longer to parcel out eggs if a partner failed to reciprocate before. But sometimes they omit retaliation, which suggests that they are really using a *Generous Tit for Tat* strategy (Dugatkin, 1997, p. 48).

The repeated Prisoner's Dilemma model of Axelrod and Hamilton (Axelrod, 1984) which assumes a fixed number of rounds or at least a fixed termination probability is not the only model that can potentially be applied to the egg trading behavior among fish. Dugatkin also describes another interpretation of the egg trading behavior by R.C. Conner that is related to a species of plycheate worms and according

to which there is no fixed termination probability but each partner decides continuously whether to continue or to break off the interaction. For Connor this is simply a matter of whether the benefit of staying[4] exceeds the benefit of leaving and, given his interpretation is right, he justly speaks of "pseudo-reciprocity" instead of reciprocity (Dugatkin, 1997, p. 49). However, without more precise quantitative data it is not possible to decide this question.

Alloparenting Another type of potentially altruistic behavior is that of *alloparenting*, which according to Dugatkin means "the dispensing of 'parental' behavior to young that are not one's own" (Dugatkin, 1997, p. 101). "Alloparenting" concerns sexually mature individuals that *could* also produce offspring of their own. From an evolutionary point of view such a behavior demands explanation because animals that want to spread their genes should primarily be interested in raising their own children not those of others. Nonetheless *alloparenting* is quite widespread and found among various kinds of mammals, birds and fish. *Alloparenting* among fish has been studied for *Lamprologus brichardi*, a type of perch (Barsch) found in the Lake Tanganyika in East Africa. For this species it is typical that the young stay at the nest for a while even after they have grown sexually mature and help cleaning eggs and maintaining and defending the territory. That this kind of helping activity is costly is illustrated by the fact that the young that stay at the nest have a slower growth in comparison with young that do not stay at the nest. The benefits that mature young derive from staying and helping at the nest include relative safety from predators and rearing kin that is at least closely related even if it is not their own. (Other suggested benefits were not confirmed or at least not measurable by experimental research.) This suggests that both byproduct mutualism (safety from predators) and kin selection are involved in the *alloparenting* behavior of *Lamprologus brichardi*. But according to Dugatkin there is also a reciprocal element present because when the mature young start to reproduce themselves they are expelled from the nest by their parents.(Dugatkin, 1997, p. 50) The only factor promoting altruism that could strictly be measured was that of kin selection, which of course is relatively easy to measure. The assumption that byproduct mutualism and reciprocal altruism are involved as well can, according to Dugatkin, be confirmed by observation but it is not possible to actually measure the payoff parameters of the game matrix and apply any of the game

[4]Although Dugatkin does not say anything about this in his report of Connor, one should assume here that what is meant is the *expected* benefit of staying, as the possible future benefit also varies according to when the other partner decides to break up the interaction.

theoretic models, let alone computer simulations in any strict sense.

In other species the *alloparenting* behavior naturally takes a different form. A type of *alloparenting* common among many mammals is *allonursing* by giving milk to unrelated conspecifics. It has been researched in some detail for the evening bat *Nycticeius humeralis*, where "approximately 20% of nursing bouts involved females feeding unrelated pups" (Dugatkin, 1997, p. 109). Among the discussed benefits are the decrease of weight during foraging bouts following the nursing and the decrease of chances of infection as a consequence of not storing surplus milk in the mammary glands. Both of these advantages would fall under the category of byproduct mutualism (which is according to our definition of altruism in chapter 2.2 not altruistic). But there could be more to it. According to Dugatkin, who relates to a study by G.S. Wilkinson, females are more likely to nurse unrelated female pups than unrelated male pups (Dugatkin, 1997, p. 109), which may be due to the fact that the males disperse. If this is true then this means that some degree of reciprocity is also involved. Another variant of *alloparenting* which has been described for Rodriques fruit bats consists in the provision of assistance in the birth process by unrelated females ("midwives") (Dugatkin, 1997, p. 109). Though it has not been determined how the altruistic behavior has evolved in this case, it is reasonable to assume that it is somehow connected with the extremely social nature of the long-lived individuals of this bat species. Again, if this is true, bat-"midwives" would at best be described as reciprocal altruists (Dugatkin, 1997, p. 109). Given the social nature of this species, one might – by drawing a somewhat risky comparison – speculate if these altruistic acts may not somehow resemble the sort of friendship altruism among humans that goes beyond the "bookkeeping kind of altruism" that reciprocal altruism is often assumed to be (Silk, 2003). But this is of course just a speculation.

Staying with the bats, one of the classical examples of animal altruism is that of blood sharing among vampire bats (Dugatkin, 1997, p. 113/114). Empirical research indicates that it is a mixture of both kin selection and reciprocal altruism. Again, the precise conditions (i.e. payoffs) cannot be measured, but several indications make the assumption highly plausible that reciprocal altruism is involved: 1) A high probability of future interaction, 2) the relatively cheap cost of providing a meal in comparison to the benefit of receiving one (the latter can be a question of life and death), which means that the threshold to offering an altruistic benefit is low, and 3) the ability of the vampire bats to recognize one another (Dugatkin, 1997, p. 114). *Alloparenting* behavior is also documented for many primate species, though here it typically

does not include the provision of food by the allomothers and usually the allomothers are immature animals (Dugatkin, 1997, p. 138) so that they do not fall under the strict definition of *alloparenting* any more.

Alarm Signals Yet another type of potentially altruistic behavior that has attracted the interest of researchers is that of giving alarm calls or alarm signals. As in many of the other instances of possibly altruistic behavior the empirical data is often too scarce to decide in any specific case whether giving an alarm call really constitutes an instance of altruistic behavior or not. In willow tits the giving of alarm calls seems to be related to the place in the dominance hierarchy and thus probably falls into the category of byproduct mutualism as the benefits derived by the survival of group members as a consequence of giving a call depend on the position of the group member. However, reciprocity has also been suggested in this context (Dugatkin, 1997, p. 86). In other bird species, downy woodpeckers and black-capped chickadees, alarm calls mainly serve the purpose of mate protection, which is demonstrated by the fact that alarm calls are not given in same sexed flocks. Then alarm calls do not provide an example of altruism but of byproduct mutualism. Still, byproduct mutualism sometimes is the first step in an evolutionary history that may eventually lead to altruism. As Dugatkin imparts, byproduct mutualism typically evolves in harsh environments. In this case the "harshness" consists in "the decreased probability of acquiring new mates" (Dugatkin, 1997, p. 86). In terms of chances of reproduction it may pay off to risk one's own survival (by giving an alarm call) in order to increase the probability of survival of a mate. Regarding the different explanations for the same type of behavior in willow tits, chickadees and woodpeckers, it should be borne in mind that it is not necessarily the case that the same type of behavior has the same evolutionary causes if it occurs in different species.

Another species for which alarm calls have been studied quite extensively are Belding's ground squirrels. Here it is quite well assessed that kinship based altruism is the decisive factor for giving alarm calls. For, typically alarm calls are given by females, and in this species females are sedentary and breed near their natal sites, while males leave their natal sites (Dugatkin, 1997, p. 97/98). The hypothesis is further strengthened by the observation "that 'invading' (non-native) females gave alarm calls less frequently than native females." (Dugatkin, 1997, p. 98). A fairly well known example of alarm calls is that of alarm calls in vervets provided by Cheney and Seyfarth in their book "How monkeys see the world". Among other things Cheney and Seyfarth found out

that the vervets' alarm calls vary depending on whether the approaching predator is a leopard or an eagle or a snake, with a different reaction elicited by the respective alarm call in each case. With respect to altruism the important question is whether the alarm call is really given with the intention to warn other conspecifics as opposed to the possible intention to signal to the predating animal that it does not need to bother because it has been detected (Dugatkin, 1997, p. 136/137.). But the former is obviously the case as different alarm calls elicit different escape reactions. As alarm calls are given with a higher probability either if offspring is present or if mates are present (in the latter case there exists again a further dependency on the dominance hierarchy), kinship and byproduct mutualism provide the most plausible explanations.

That giving alarm signals does not necessarily need to be an instance of altruistic behavior and not even a form of byproduct mutualism is illustrated by the stotting behavior that occurs in Thomson's gazelles (and also in some other less well studied species), a curious kind of behavior "wherein individuals take all four legs off the ground simultaneously and hold them straight and stiff in the air" (Dugatkin, 1997, p. 94). From numerous hypotheses that have been put forth to explain stotting only two could be confirmed according to Dugatkin, namely that stotting is meant to inform the predator of the health of the stotting animal (which means that the predator will know that the stotting animal will be difficult to catch and will rather "lock on" some other individual) and that young animals stott to attract the attention of their mother in dangerous situations (Dugatkin, 1997, p. 95). In both cases altruism or cooperation is not involved.

Grooming Most of the examples of cooperative or altruistic behavior among animals so far have been examples of kin selection or byproduct mutualism, but in spite of the fact that there is a strong "skew towards reciprocity in the theoretical literature" (Dugatkin, 1997, p. 167) there have been very few clearcut cases of reciprocal altruism, let alone of group selection. One kind of behavior that from its very appearance seems to fit the conception of reciprocal altruism quite well and is often mentioned as a kind of role model in this context is that of grooming. Dugatkin relates several studies about grooming in primates as well as other mammal species. One non-primate species where grooming has been studied are impala, an antilope species. It is at the same time one of the rare examples that really fits the model of a repeated game – at least on a qualitative level. According to Dugatkin who refers to two studies from Hart and Hart and Mooring and Hart, impala exchange

bouts of grooming, each bout consisting of a repeated "upward sweep of the tongue or the lower incisors along the neck of the partner" (Dugatkin, 1997, p. 91). These exchanges of grooming bouts expose several striking features which strongly suggest that grooming in impala is an instance of pure reciprocal altruism: 1) There is an almost perfect match between bouts of grooming received and bouts delivered; 2) the exchange of bouts ends after one partner stops allogrooming. This rules out the possibility of byproduct mutualism, which could otherwise offer an explanation if it is assumed that ticks provide some extra nutrition for the impala; 3) there is no correlation with the rank in the dominance hierarchy (Dugatkin, 1997, p. 91-94). All in all, this finally seems to be a clearcut example for the kind of reciprocal altruism that is described by the repeated Prisoner's Dilemma model. However, even in this case the match between model and empirical reality can be ascertained only on the basis of qualitative similarity because a quantitative measurement of the payoff parameters has not been done.

Grooming is also one of the most salient behavioral features of our closest relatives in the animal world, the primates, and therefore has caught a lot of attention by researchers. The patterns of grooming exchanges among primates are much more complex than among the impala just described. In primates, grooming can serve many different functions next to the purpose of removing ectoparasites. Among these are the reduction of tension (which could otherwise result in conflicts), coalition formation, where grooming serves as a means to "bribe" others to become allies, and, more general, grooming as an "exchange currency" to gain other favors in return. While all these describe possible benefits of grooming, Dugatkin notices that in most studies very little is said about the costs of grooming (Dugatkin, 1997, p. 117). But certainly there are costs. Apart from the time and energy spent, it has been recorded that the lowered attention of mothers engaged in grooming activities results in their unattended offspring being significantly more often being harassed by other animals (Dugatkin, 1997, p. 117/118). There is good evidence that grooming is to a certain degree reciprocal in chimpanzees, though the reciprocal nature of grooming is not as clear cut as in the case of impala. In vervets (Meerkatzen) the relation of grooming and coalition forming has been studied. Here grooming does increase the probability of responding to solicitation calls for unrelated animals but not for related animals (where the probability of responding is high, anyway). These results are not completely undisputed (Dugatkin, 1997, p. 120), but if they are true, then it appears to be a case of reciprocal altruism because kinship can be ruled out and, as there exists an opportunity for cheating (groomed animals could fail to respond to solicitation

calls), byproduct mutualism can be ruled out as well. Further kinds of grooming in exchange for "goods and services" have been documented in chimpanzees and macaques. In chimpanzees grooming sometimes is related to food exchange (Dugatkin, 1997, p. 123). In an experiment conducted by Stammbach, a single subordinate member of a group of macaques was trained to operate a complex lever mechanism for food release (from which all group members could eat). While the subordinate "specialist" did not rise in rank, it received significantly more grooming than before by other group members. The acts of grooming did, however, not take place in strict connection with acts of operating the mechanism (Dugatkin, 1997, p. 124). So, if any kind of reciprocity is involved here, it is not the strict type of "bookkeeping reciprocity" that the repeated Prisoner's Dilemma model suggests. Quite a lot of studies on primates emphasize the factor of kinship in grooming (Dugatkin, 1997, p. 124).

Eusociality The most astonishing example of cooperation in the animal kingdom is that which is found in bee hives or ant hills, where a large state of insects operates in what appears to be an extremely cooperative and coordinated manner. Biologists call these kinds of insects *eusocial insects*, where *eusociality* is defined by three criteria: 1) Reproductive division of labour, 2) communal care for the young and 3) overlapping generations of workers in the colony. Eusociality is not only found in insect species like bees, wasps, ants, termites but also in certain vertebrates like naked mole rats and Darmland mole rats. When one compares the forms of cooperation that take place in eusocial animals with the other instances of cooperative behavior that have been described in this chapter one cannot help but notice the extraordinary qualitative difference that eusociality makes for cooperation and altruism. Eusocial animals do not just cooperate with respect to a single function (like grooming in mammals) but they seem to cooperate in any possible form and manner. Of the many possible examples of cooperative behavior among eusocial insects, Dugatkin describes in more detail the cooperative behavior of honey bees in foraging, hive thermoregulation and anti-predator behavior. When foraging, honey bees cooperate in different ways. They inform each other about the location of food resources via the famous "waggle dance" and they coordinate their foraging activity with regard to the level of food supply in the hive in a complex manner (Dugatkin, 1997, p. 152/153). Hive thermoregulation is achieved by the bees behaving in such a way as to keep the temperature inside the bee hive at an ideal 35 degrees Celsius. As the

temperature of the whole hive only marginally depends on the activity of a single bee, this raises a typical collective goods problem, where one would expect that the individual bees are encouraged to cheat. But in fact they do not (Dugatkin, 1997, p. 154/155). Even more admirable is the self sacrificial behavior of honey bees for the defense of their colony. Because honey bees die when stinging, this behavior appears to be an extreme case of altruism to the advantage of the colony.

How is the astonishing variety of forms of cooperative behavior as well as the intensity that altruistic behavior reaches in eusocial animals to be explained? The best known explanation is that by inclusive fitness. It has been found out that eusocial insects are haplodiploid species, where the males carry only a single (haploid) set of chromosomes while the females have a double (diploid) set of chromosomes. The female descendants of the queen all share the same genes from their father and on average 50% of their mother's genes. In consequence, the worker sisters are on average 75% related to each other. Thus cooperation in eusocial insects is easily explained by kinship, one should think. But there are problems with applying the inclusive-fitness-theory to eusocial animals. One problem is that there exist eusocial species where the queen has multiple matings and others where there are several queens in one colony (Dugatkin, 1997, p. 144). Therefore, kinship cannot be the only explanation for eusociality. Dugatkin discusses in this context a number of alternative hypotheses on eusociality (Dugatkin, 1997, p. 144-149). But rather than entering into the complex debate about these hypotheses, which for a layman would be difficult to present accurately anyway, I confine myself to a few general reflections on eusociality as an example for the evolution of cooperation.

In order to do so, I distinguish between two different questions: 1) Why do the workers in the colonies not reproduce? Or in other words, why did centralized reproduction evolve and how is it maintained? 2) Given that the workers cannot reproduce, why do they cooperate? I am going to answer the second question first because it seems to be an almost trivial question. If, for whatever concrete reason, the workers really cannot reproduce individually, then it follows that the best thing they can do to spread their genes is to cooperate as well and as completely as possible with the rest of the colony. For, imagine that due to a mutation some of the worker ants hatching in an anthill were lazy ants that did nothing to contribute to the colony. Then although the lazy ants would greatly profit from letting the others do all the work, they would not be able transform this advantage into greater reproductive success within the hive simply because they cannot reproduce themselves. At the same time the anthill as a whole would suffer increased

selection pressure from other anthills without lazy ants. One could say that the scenario that explains the cooperation within eusocial species is that of group selection, only that the within-group selection that counteracts the evolution of altruism in group selection models is inhibited. Therefore, in order to produce altruism, evolution only has to solve the technical problem of coordinating the behavior of the eusocial insects as well as possible but evolution does not have to resolve a conflict of reproductive interests any more, which in non-eusocial species acts against the emergence of altruism. This explains both the extraordinary intensity of altruistic behavior (up to self-sacrifice!) as well as the great variety of cooperative behavior in eusocial species. Strictly speaking, however, our definition of altruism in chapter 2.2 would preclude calling the cooperative behavior of eusocial insects altruistic if the "benefits" in the definition are understood in terms of reproductive fitness. Because the workers in a colony do not reproduce, no fitness costs are incurred by them by acting altruistically.

Given that the altruistic behavior of eusocial animals is easily explained by (uninhibited) group selection, the remaining question is, how did the workers ever become so altruistic as to stop reproducing individually and why do they remain so? It is in answer to this question that other mechanisms like inclusive fitness or byproduct mutualism come into play. In mole rats, Dugatkin maintains, it was byproduct mutualism forwarded by harsh environmental conditions such as successive prolonged droughts in the evolutionary history of certain mole rat species that caused the evolution of eusociality:

> ... at the evolutionary onset of cooperation in naked mole rats, when reproductive division of labor was likely minimal, a "harsh environment" central to byproduct mutualism, rather than kinship per se, may have been the predominant selective agent. (Dugatkin, 1997, p. 106)

Differently from typical eusocial insect species, mole rats have a diploid set of chromosomes, which once more shows that eusociality does not by necessity depend on the genetics of a haplodiploid set of chromosomes. Still, it is plausible to assume that the close kinship ties in haplodiploid species facilitate the evolutionary transition to a reproductive division of labor because the fitness cost of giving up individual reproduction in favor of centralized reproduction in a colony is much lower if the relatedness is close. The mechanisms by which the reproductive division of labor is maintained do – as one should expect – also vary from species to species. For honeybees, for example, a mechanism

called "worker 'policing' " has been described, where the males that hatch from worker laid eggs[5] are killed by other workers. The behavior is probably best explained by kinship. (If the queen has multiple matings, workers are more related to their brothers than to their nephews (Dugatkin, 1997, p. 150).) But Dugatkin also suggests that group selection may play a role "in that without policing a much greater degree of within-colony aggression would exist, and this, in turn, could decrease group productivity" (Dugatkin, 1997, p. 151). Another obvious way to ensure the monopoly of reproduction is aggression on part of the queen by which the workers are coerced into their role. This has been reported for the previously mentioned mole rats (Dugatkin, 1997, p. 106).

If the alleged altruism of eusocial species is easily explained by the reproductive division of labor, then the cooperation of several queens in one colony must still be explained by the other mechanisms of the evolution of altruism. And indeed, here we can find some striking cases of reciprocal altruism and even group selection. One such case is the "social contract" that is found in paper wasps (*polistes fuscatus*) (Dugatkin, 1997, p. 157/158). In paper wasps dominant queens tolerate other, subordinate queens in their nest. Both dominant and subordinate queens lay queen-destined as well as worker-destined eggs. But subordinate queens disappear by the time the workers emerge. Cooperation between dominant and subordinate queens requires that they leave each other's eggs unharmed. Experimental research has shown that subordinate queens reacted aggressively to simulated oophagy on queen destined eggs, but not on worker destined eggs, while the dominant queen did not show such a reaction. This strongly hints to reciprocal altruism on part of the subordinate queens. The suggested reason why dominant queens do not react to simulated oophagy at all is that they can still produce queen-destined eggs after the subordinates are gone, while the subordinates themselves do not get a second chance. For the dominant queen it is a different deal, so to speak.

An example of cooperation between colony founding queens that is probably due to group selection can be found in desert seed harvester ants (*Messor pergandei*) (Dugatkin, 1997, p. 159). For some populations of this species it has been observed that the queens jointly produce workers when founding a colony. Once the workers have emerged, the queens fight to the death until only one queen is left. Another feature of the desert seed harvester ant is that different colonies are engaged in brood raiding against each other. According to Dugatkin's account, the

[5]In honeybees workers lay eggs, but these are unfertilized and only develop into males, whereas the queen can control which of her eggs are fertilized and thus develop into females and which are not fertilized and develop into males.

following holds:

> In the case of *M. pergandei*, the trait of interest is the production of workers, which, although selected against within groups (via the cheater problem), may be selected for as groups with many cooperators survive brood raiding (i.e. differential productivity of groups). (Dugatkin, 1997, p. 160)

As the relative isolation of groups is a vital requirement for group selection to operate towards the evolution of cooperation, it is no surprise that the cooperative behavior only occurs in populations of *M. pergandei* "where environmental factors aggregate starting colonies, which occur only in the sandy ravine bottoms where soil moisture is available" (Dugatkin, 1997, p. 160). Other populations of the same species that live in different habitats do not display cooperative behavior in the founding phase of a colony, but here queens react aggressively to any rival right from the beginning (Dugatkin, 1997, p. 160/161). The conclusion that cooperation in *M. perganei* is a result of group selection has not gone completely undisputed however. As in this case – just as in any other of the empirical instances of the evolution of cooperation in biology described so far – no quantitative measurement of payoffs could be made, it is of course difficult to assess these findings beyond what can be deduced from the mere phenomenology of this instance of cooperation. Still, similar results have also been obtained for another ant species, *Acromymes versicolor* (Dugatkin, 1997, p. 161), which bestows the explanation by group selection in this case with some additional credibility.

Discussion: Do the computer models of altruism live up to the empirical research in biology?

The list of examples of cooperative and altruistic behavior among animals that has just been given is, of course, far from being complete. Still, it shows how far reaching and varied the forms of cooperative behavior that exist in nature are. But apart from this scientific fact, which is certainly interesting in its own right, our main concern here is to find out in how far the kind of modeling of altruism that has been demonstrated in the previous chapter proves to be helpful for the understanding of the empirical instances of altruism and, if not, what are the causes for this failure. In order to tackle these questions we must distinguish different levels of the application of formal models and in particular of computer simulations to the empirical problem:

1. *Conceptual Level*: On this level the model is merely meant to demonstrate how a certain mechanism works in principle. For this purpose it is not necessary that the model is empirically very adequate or that the parameter values used in the model are based on more than plausible assumptions. Still, the model cannot be arbitrary. It must at least give us some indication of how the empirical phenomenon can be identified as one that falls within the class of phenomena which the model describes. For example, repeated Prisoner's Dilemma models of reciprocal altruism indicate that there must be repeated interaction and that the situation should be a (repeated) dilemma situation, not just one where the participants profit from their interaction anyway, as in byproduct mutualism. This alone – as the previous brief survey of empirical examples has shown – can already be difficult to determine.

2. *Application Level*: At this level we require that there is a close concordance between the model itself and the empirical phenomenon or class of phenomena that the model describes (or "models"). The concordance must be close enough so that we can empirically determine 1) whether the model applies to the empirical phenomena in question and 2) whether it describes them correctly. If the model contains quantitative magnitudes as input or output values then this implies that we must be able to measure these magnitudes in some way or other.

We will elaborate on these two categories of models a little more in chapter 6. Here the distinction is made mainly to preclude a certain defense strategy that is often used to excuse spurious modeling. This defense strategy consists in replying, whenever somebody calls into question that the model fits empirical reality, that it is just a model and that from a model, being by definition a strongly simplified representation of reality, one cannot expect a representation of the modeled empirical situation that is accurate in every possible respect. However, as not every model can be a model for anything, there must be a limit up to which this excuse is acceptable. And this limit certainly depends on what claim is connected with the model. If the claim is that the model can actually be applied, the requirements are certainly higher than when it is just meant to give expression to a certain idea or concept.

Regarding the empirical examples from biology that have been presented so far, it can safely be concluded that *not a single one* of the simulation models of the kind that have been presented in chapter 4 proved to be applicable in a strict sense. In the beginning of his book on "Cooperation among Animals" (Dugatkin, 1997) Dugatkin lists a

whole array of such models. But even though he is extremely sympathetic towards this approach, he almost nowhere in his book refers to any of these models. There is no instance – except one which ultimately turned out to be a failure (see chapter 5.1.3) – where the empirical research he presents is related or can be related to any of the theoretical simulation models. The reasons for this are hinted at by Dugatkin himself in the last chapter of his book: Save for one exception, Dugatkin was not able to present a single empirical study where the payoff parameters, which are crucial for the application of any game theoretical model, have been or could be measured. The one exception concerns an experimental study on blue jays, where blue jays could trigger a "cooperate" or a "defect" button (Dugatkin, 1997, p. 80/81) and thereby release food according to a Prisoner's Dilemma game matrix or – in a second experiment – according to a stag hunt game matrix (which is one way to circumscribe byproduct mutualism in game theoretical terms). The result was that blue jays never cooperated in the Prisoner's Dilemma, even though it was repeated, and always cooperated in the stag hunt game. The authors of the experiment concluded that no strategies for interaction in the repeated Prisoner's Dilemma have evolved in blue jays, which leads them to doubt the "general significance of the Prisoner's Dilemma as a model of non-kin cooperation." (Dugatkin, 1997, quoted p. 80). Notwithstanding this skeptical conclusion about the Prisoner's Dilemma as a proper model for non-kin cooperation, Dugatkin regards it at least as a serious attempt to address the issue of quantifying the payoff matrix (Dugatkin, 1997, p. 165). This can surely be granted, but it is still a long way until a satisfatory mode of quantification will be reached. For, in order to quantify the payoff matrix we would need to know the payoff values in terms of reproductive fitness and not merely in terms of food release, which does most probably not transform proportionally into relative numbers of offspring.

If this was the only example where the empirical research was approaching the measurement of payoff paramaters and if – as we have seen in chapter 4 – the computer models of altruism crucially depend on the values of the payoff parameters then this means that the level of empirical applicability of these models has not yet been reached – at least not at the time when Dugatkin compiled his surveying study on "Cooperation among Animals" (1997).[6]

But what about the conceptual level? If the computer models are not (yet) really applicable, do they perhaps help us to form sound concepts and provide us with categories of analysis? Even on the conceptual

[6]This still seems to be true today (see the following section).

level, it has in many cases been difficult to decide which type of altruism is at work in a specific case and whether it is altruism at all and not merely byproduct mutualism. At the same time, game theoretical models (though not game theoretical models alone) allow for a relatively sharp conceptualization of different types of altruism, which is helpful even if these types do in many instances not appear in a pure form in nature (grooming among impala being one of the few exceptions). One could say that on this level they serve a similar function as the "ideal types" do in the social sciences according to Max Weber: Even though they contain very strong abstractions they can help to get a better grip on empirical reality. The heuristic benefits of game theoretical thinking for the understanding of altruism become apparent in the case of grooming among primates. Here, as Dugatkin notices (Dugatkin, 1997, p. 117), behavioral ecologists have mostly focused on the benefits of grooming but not often asked the question of the costs of this type of behavior. This is quite understandable from the point of view of behavioral ecologists because from its very appearance the grooming behavior does more strongly suggest to ask the question of what it is good for than the question of its costs (which even might seem quite negligible at first sight). But from the theoretical perspective it is clear that the question of why this kind of potentially altruistic behavior evolved is a question of benefits *and* costs. Thus, theoretical reflection on models of altruism, even if they are toy models, may help to direct the empirical research in a useful manner.

This said, there is of course an important caveat that has to be mentioned right away. The benefits just described of modeling on the conceptual level (clarifying and sharpening our concepts, directing empirical research) only hold for the most elementary and simple models, but not for complicated models, massive simulations and in general the whole baroque richesse of theoretical models and simulations that can be derived from any simple model by changing parameters, adding further "plausible" conditions etc. Judged against the background of the empirical findings that are summarized in Dugatkin's book (which, after all, is the book of an author who is very sympathetic towards the modeling approach), simulations in the fashion of those of which a small sample has been discussed in chapter 4.1.5 and of which a role model has been presented in detail in chapter 4.1.4 have turned out to be as good as completely useless. Neither did they provide us with important insights on the conceptual level that went beyond what can already be demonstrated by much simpler toy models, nor was any simulation of this type empirically applicable in the sense described above.

Given that the simulation models turned out to be largely useless for the explanation of the evolution of altruism in nature, the question is, of course, what are the reasons for this deplorable state of affairs. One possible explanation could be that most of the empirical research surveyed in Dugatkin's book was not designed to put any particular models of altruism or cooperation to the test, but that the behavioral ecologists conducting such research had other research interests. This might be especially true for the field research on cooperative behavior as opposed to the experimental research. Usually, there exists a time lag with which newly invented concepts and methods pervade a whole science. If this was true then maybe the only problem was that Dugatkin wrote his book too early, at a time when only a small part of the empirical research was informed by the latest models of the evolution of altruism? But then we should expect to find more usage of simulation models in the empirical research on altruism that has been published since. In order to check whether this is the case we will briefly examine a more recent example of the empirical research on altruism in the following section (section 5.1.2). It will turn out that just as little use could be made of simulation models as in any of Dugatkin's examples. In order to further pinpoint the difficulties that prevent the application of simulation models, or, more precisely, the brand of simulation models that has dominated the modeling of the "evolution of altruism" for a long time, I finally discuss in depth one of the few examples where biologists set out with Axelrod's and Hamilton's concept of reciprocal altruism but soon became aware of the limits of this theoretical background (see chapter 5.1.3).

5.1.2 A more recent example: Image scoring cleaner fish

The discussion of Dugatkin's survey on "Cooperation among Animals" has shown that there is a wide gap between the modeling of altruism and cooperation on the one hand and the empirical research on cooperative behavior among animals on the other hand. While the theoretical models did allow formulating certain concepts of altruism, it was not possible to relate the simulation models of altruism to the empirical instances of cooperative behavior in any more than a metaphorical sense. But is this limitation due to systematic difficulties of applying abstract simulation models or is it, maybe, just an interim problem that can ultimately be overcome by more refined empirical research methods? Since Dugatkin's survey dates from 1997 it is reasonable to ask whether the situation has changed till then. Therefore, we will look at one recent example of empirical research on altruism in biology. Again, the pur-

pose of the discussion of this example is primarily epistemological. No claim is made that the examples discussed in the following, concern very important or representative types of altruism in nature (although they fit well into the overview of animal altruism given previously). We want to find out, how much use is made of theoretical models of altruism in typical empirical research studies.

The study concerns "Image scoring and cooperation in a cleaner fish mutualism" (Bshary and Grutter, 2006). *Image scoring* is variant of reciprocal altruism, where cooperation depends on whether the partner has been seen to cooperate with others. Image scoring is thus a type of indirect reciprocity because it is the altruistic act that has been bestowed unto someone else that is being reciprocated. The rationale behind indirect reciprocity is that someone who has behaved cooperatively towards someone else may also behave cooperatively to oneself. Another type of indirect reciprocity that does only occur among humans is reputation based cooperation, where one gains reputation by cooperating with people that have a high reputation. Differently from mere image scoring, reputation can be passed on by telling about it. Image scoring only requires that the partner's behavior is observed in a similar situation. In contrast to reputation based cooperation the cognitive requirements for image scoring are therefore only comparatively low. In fact they may be even lower than the cognitive requirements for the evolution of altruism in repeated Prisoner's Dilemma situations because for image scoring no bookkeeping or partner recognition is required so that it does not come as a surprise that image scoring behavior can be found even among relatively "primitive" animals.

In the cleaner fish *Labroides dimidiatus* (also known as Striped Cleaner Wrasse, or in German: "Putzerlippfisch") that Bshary and Grutter experimented with, the clients "invite" the cleaner fish for inspection. The cleaner fish then usually feed upon the ectoparasites of the client. But they could also feed on the mucus of the client and there is evidence that the mucus is actually their preferred nourishment. Thus, the cleaner fish can either cooperate by removing the ectoparasites or cheat by munching the client's mucus. The client on the other hand cannot cheat the cleaners. Due to the asymmetry of the situation, cooperation could not have been evolved via direct reciprocity. That image scoring is a potential candidate for the explanation of cleaner fish cooperation is suggested by field research on cleaner fish according to which: "Client fish almost always invite a cleaner's inspection if they witnessed that the cleaner's last interaction ended without conflict, invite less if they do not have such knowledge, and invite the least if the last interaction ended with conflict." (Bshary and Grutter, 2006, p. 975).

In order to test the image scoring hypothesis Bshary and Grutter conducted two experiments, one on the client behavior and one on the behavior of the cleaner fish. In the first of these experiments a client was placed in the middle of an aquarium divided by one-way mirrors into three basins. In one of the side basins a group of cleaner fish fed on prawns attached to a model client fish. In the other side basin a group of cleaners was placed without a model. The result of this experiment was that the client spent significantly more time near the group of cleaners that was engaged in cleaning activity. This result suggests the conclusion that clients prefer cleaners that can be observed to be cooperative over cleaners with an unknown cooperation level.

The second experiment was more complicated. This time the cleaners were placed in either an image scoring or a non image scoring scenario. In both scenarios the client fish was simulated by plates to each of which two different types of food items, fish flakes and prawn items, were attached. *Labroides dimidiatus* prefers prawns to fish flakes just like it prefers mucus to ectoparasites. The question that the experiment was intended to answer was whether the cleaner fish would cooperate by feeding against their preferences in the image scoring scenario. In both the image scoring and the non image scoring scenario the cleaner fish could feed from two identical plates. In the image scoring scenario both plates would be removed immediately after one prawn item was eaten from one of the plates, while in the non image scoring scenario only the plate from which the prawn item was eaten was removed. To make sure that the cooperative or non cooperative behavior did not merely depend on the sheer amount of nourishment available a third scenario was tested, where the cleaner fish could feed only on one plate which was also removed immediately after a prawn item was eaten. The result was that in the image-scoring scenario the cleaner fish fed significantly more often against their preference when feeding on the first plate than when feeding on the second plate or a single plate or when feeding on the first plate in the non-image-scoring scenario.

The experimental results thus strengthen the assumption that cooperation in cleaner fish is due to image scoring. It is noteworthy that the cleaner fish do not merely react to the presence of another client, a condition which was fulfilled in the image scoring and the non image scoring scenario, but to the reaction of the other client that is present. This means that the cleaner fish do only cooperate if the clients actually engage in image scoring.

Now the crucial question for our purpose, the assessment of the value of theoretical models for the empirical research, is whether and to what level Bshary and Grutter could draw upon theoretical models of the

evolution of cooperation. Bshary and Grutter do not make more than passing mention of the mathematical models and computer simulations on image scoring (Bshary and Grutter, 2006, p. 975). Not to enter upon a discussion of these models is quite reasonable for them as the specific features of these models remain completely irrelevant for their empirical research. It is only the basic concept of indirect reciprocity that Bshary and Grutter draw upon for their empirical research. The concept of simple indirect reciprocity requires "image scoring by clients and an increased level of cooperation by cleaners in the presence of image-scoring clients" (Bshary and Grutter, 2006, p. 796). Both these requirements have been tested experimentally by Bshary and Grutter. Again, we find a concordance of theoretical modeling and empirical research only on a basic conceptual level.

5.1.3 An in-depth example: Do sticklebacks play the repeated Prisoner's Dilemma?

In order to show what difficulties the attempt to apply the models of reciprocal altruism meets in practice, I discuss in the following an example where biologists tried to apply the theory of the "evolution of cooperation" of Axelrod and Hamilton (Axelrod, 1984) (which is based on computer simulations that have been a role model to the ones presented above) to a case of altruistic behavior in nature. The example concerns a behavioral trait called "predator inspection" that is found in certain types of shoal fish like sticklebacks. The behavior of "predator inspection" has among others been examined in two empirical studies by Manfred Milinski and Milinski and Geoffrey Parker. The earlier of these two studies (Milinski, 1987) still draws heavily on Axelrod's and Hamilton's model of the repeated Prisoner's Dilemma. The other study that has been described in a paper that appeared ten years later (Milinski and Parker, 1997) and employs a totally different theoretical interpretation of the results. As I try to demonstrate in the following, both studies taken together show that the choice of an appropriate formal description of reciprocal altruism (or cooperation) raises very difficult and often by no means unambiguous questions of interpretation and measurement. Against this background any game theoretical model research that is not closely linked to empirical questions must appear like a pure "Glasperlenspiel".[7]

"Predator inspection" is a behavior that is found (among other

[7]That this has nothing to do with the usual gap between theory and practice or between theoretical and empirical research but reflects a specific impasse of the modeling approaches in evolutionary game theory will have become clear at the end of this section and will be discussed again in chapter 6.

species) in sticklebacks. Sticklebacks are small fish living in shoals. If a predator (a pike for example) comes within a certain range of the shoal, it can be observed that either a single stickleback or a pair of sticklebacks leaves the shoal and carefully approaches the predator. The sticklebacks do so in order to inspect the predator, presumably to gain information about the type, size, location and movement of the predator. Typically, a pair of sticklebacks gets much closer to the predator than a single stickleback. If the sticklebacks approach as a pair, it can be observed that they advance with characteristic jerky movements in such a way that one stickleback swims a short distance ahead and then "waits" for the other, who follows in a similar jerky movement (Milinski, 1987, p. 433). This suggests interpreting the sequence of jerky movements as a repeated Prisoner's Dilemma, where the sticklebacks play *Tit for Tat*. In his earlier paper Milinski tried to confirm this assumption by simulating the partner stickleback with different types of mirrors so that the mirrored fish either appeared at the same distance from the predator (simulating a cooperative partner) or a little bit further behind (simulating a non cooperative partner). The result was that the sticklebacks advanced much closer to the predator when they were accompanied by a cooperative partner. Milinski interpreted this result as an empirical confirmation of Axelrod's and Hamilton's theory of cooperation. By and large this seems correct if we ignore for a moment the fact that the results of Axelrod's and Hamilton's simulations were more contingent than was known at that time. But there exists a problem in so far that Milinski confines himself to assessing that the two inequalities $T > R > P > S$ and $2R > T + S$ hold. Now, as the simulation results above show, the simulation is sensitive to changes in the concrete values of the payoff parameters, and unfortunately these would be very hard to measure in the case of the sticklebacks.

After much further experimental research on sticklebacks in the later paper, Milinski and Parker offer quite a different formal description of the same behavioral trait of "predator inspection". There is not much talk about the repeated Prisoner's Dilemma any more. While it is still true that the situation of two sticklebacks approaching a predator can (at a certain distance range) be interpreted as a Prisoner's Dilemma, this assertion alone does not shed much light on the problem. Instead of meddling with the Prisoner's Dilemma, Milinski and Parker therefore examined the possible utility calculus that controls the behavior of the sticklebacks.[8] According to Milinski and Parker, even a single stickle-

[8]In the following Milinski's and Parker's construction will only be described in general terms. For the mathematical details see Milinski and Parker (1997). A major problem of this construction, which is also the reason why Milinski and Parker only reach an ambiguous conclusion, is that the fitness benefits of

back will approach a predator up to the point where the advantages (of gaining information about the predator) are balanced by the risk of being eaten (Milinski and Parker, 1997, p. 1241/1242). For the case when two sticklebacks jointly approach the predator, Milinski and Parker offer two alternative descriptions one that assumes cooperation (Milinski and Parker, 1997, p. 1242) and another one that does not necessarily presuppose cooperation (Milinski and Parker, 1997, p. 1242-1245). Milinski and Parker do not ultimately reach a decision which of these descriptions is the right one. For, even if one does not assume cooperation, two fish will – according to their model description – move closer to the predator than a single fish. The reason is this: The distance to the predator can be divided into three zones, the "far zone", the "match zone" and the "near zone". In the "far zone" that is, when the distance to the predator is still very great, each of the two fish gets an advantage from moving closer to the predator, even if the other fish stays back. In the "match zone" (medium distance to the predator) a partner that has fallen behind will try to catch up with its forerunner, although neither of the two partners gets an advantage from taking the lead (from which it follows that both fish can only advance synchronously if one does not assume at least a minimum of reciprocal altruism). Finally, in the "near zone" the "best reply" of each fish is to stay back behind the other one.

If there are two different theoretical descriptions of the behavior of a pair of "inspecting" sticklebacks, one that assumes cooperation between the sticklebacks and one that does not, then this raises the question which of these is true or whether the sticklebacks in reality cooperate or do not cooperate when jointly inspecting a predator. At the time of writing the second paper Milinski and Parker come to the conclusion that the current state of research does not allow to decide this question: "However, it is not yet possible to analyze quantitatively whether pairs are conforming to the cooperative or non-cooperative ESS [**E**volutionary **S**table **S**trategy, E.A.]." (Milinski and Parker, 1997, p. 1245) How can this result be reconciled with their earlier study that seemed to confirm Axelrod's and Hamilton's theory of the "evolution of cooperation"?

inspection can only be guessed. While it is plausible to assume that the benefits decrease with decreasing distance from the predator, there exist no exact measurement procedures for the benefits. Therefore, both the type of the function (Milinski and Parker present two alternatives, an exponentially decreasing and a linearly decreasing function) and its parameter values can only be guessed. – In response to a criticism that appeared slightly earlier, Dugatkin, who worked theoretically and empirically on the same topic as Milinski, still defends the notion that predator inspection behavior is best understood as a *Tit for Tat* strategy (Dugatkin, 1996). But he misses out the problem that the respective Prisoner's Dilemma models are notoriously unstable and he seems to assume that there exist only the two alternatives to explain the behavior of predator inspection either as the outcome of a reiterated Prisoner's Dilemma or as byproduct mutualism. But as the later paper from Milinski and Parker (Milinski and Parker, 1997) suggests, these are not the only alternatives to conceive of predator inspection (see the main text below).

The answer is that obviously the earlier conclusions have been drawn too rashly, probably due to a subtle misconception in the earlier experiment's setup: An uncooperative fitness maximizing fish would never have behaved as the uncooperative fish simulated by the mirror did. Therefore, the reaction of the real fish that stopped at a specific distance from the predator does not necessarily need to be interpreted as a "punishment" which is part of a *Tit for Tat* like strategy. The distance at which the real fish stopped may just have been its optimal distance (from a purely "egoistic" point of view) given the presence and distance of the simulated partner fish.

The result shows how difficult it is, even in a biological context, to apply simulation models of reciprocal altruism such as those described above. The repeated Prisoner's Dilemma does not seem to be an appropriate model for the sort of behavior Milinski examined. As has been shown previously, other examples for reciprocal altruism from biology meet the same difficulties. The same conclusion is confirmed by other biologists that work in the field of evolutionary game theory. An expert in this field, Peter Hammerstein, writes: "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, 2003a, p. 83) In face of the vast multitude of models of reciprocal altruism and the "evolution of cooperation" this is a rather sobering conclusion. Yet, it must be taken seriously. And if it is taken seriously, it strongly confirms the skepticism towards purely theoretical simulations that has already been expressed earlier. As it appears, "blind modeling" (that is modeling that is not informed by empirical research but relies only on plausible assumptions alone) is not a proper research tool that allows us to find anything out about reciprocal altruism beyond the merest truisms.

Is there really nothing that can be done about it? In a critical appraisal of the game theoretical computer simulations in biology, Dugatkin described the situation roughly as follows: In order for the models to contribute to scientific progress, models and empirical research must be part of a feedback loop that is, theoretical models may help to direct empirical research but then the insights and results of the empirical research must be "fed back" into the construction and refinement of models (Dugatkin, 1998, p. 54ff.). Obviously, the feedback loop was not closed, insofar as the bulk of simulations on the evolution of cooperation did never really take into account the restrictions and conditions of the empirical research on the subject. The question of

the relation between empirical research and theoretical modeling will be elaborated a little more in chapter 6, where a *build to order principle* of computer modeling will be proposed, according to which models that aim to go beyond a merely conceptual level should always be constructed around empirically measurable quantities. That the burden of accommodation is thus laid on the theoreticians finds its justification in the fact that much stronger restrictions apply when devising measurement procedures (including the restriction that only certain quantities can be measured at all) than for the design of models which has become comparatively simple with the advent of computers.

5.2 Empirical findings in the social sciences

Empirical research on cooperation and altruism in the social sciences falls roughly into two different categories. One part of the empirical research consists of laboratory experiments, where the predictions of game theory are tested by letting subjects play different types of cooperation games. For this type of research subjects are placed in highly stylized and artificial laboratory situations. This allows creating situations which are somewhat similar to the highly abstract settings presupposed by mathematical models or computer simulations of cooperation. Although the laboratory experiments are usually not designed to match a particular model[9], they allow for some degree of comparison between theoretical results and empirical reality. Following as before the pars pro toto approach, I discuss two selected examples of this type of research and highlight the epistemological issues involved.

The other and more important part of empirical research on cooperation would be real world examples that potentially expose the patterns of cooperation predicted by the theory. In spite of the extreme popularity of Axelrod's book on the "Evolution of Cooperation" (Axelrod, 1984), there exist only relatively few empirical field studies that make use of the theory of the evolution of cooperation which is based on the repeated Prisoner's Dilemma.[10] Usually, such studies rather draw a sort of general inspiration from the ideas related to the repeated Prisoner's Dilemma model than relate to any simulation models in particular. But then – as has already become apparent in the biological case – the respective computer simulations are not really suitable for empirical application. For, it is often close to impossible to measure the relevant parameters, to exclude interferences of coefficients not captured by the theory or just to ascertain which kind of game is played in a given situation. In order to show what difficulties are involved when one tries to apply the results of computer simulations to real world problems, I discuss the application of the reiterated Prisoner's Dilemma model of the evolution of altruism to the "live and let live" system that evolved in the First World War among soldiers of opposing armies. This example is particularly well suited for demonstrating the epistemological

[9]They are rather designed with certain research questions in mind, taking into account the pragmatic restrictions of the laboratory and not always strictly relating to theoretical results.

[10]For an overview of the literature that relates to Axelrod's theory see (Axelrod and D'Ambrosio, 1994), (N.M.Gotts et al., 2003) or (Hoffmann, 2000). It is characteric that the the only empirical application scenario that the latter quotes is the ultimately failed attempt of Milinski to interpret the predator inspection behavior of sticklebacks in terms of the repeated Prisoner's Dilemma (see also chapter 5.1.3). All three surveys strengthen the impression that the modeling business is mostly self-contained and quite detached from the empirical research.

issues involved in the use of simulation models in the context of empirical research because it has originally been advanced as a showcase to demonstrate the power of the simulation based approach to the study of the evolution of altruism (Axelrod, 1984, p. 67-79). The example will be discussed in depth and it will be demonstrated that far from being a showcase for the use of simulation models it exposes some severe limitations of this method. The criticism will be elaborated thoroughly in the final chapter (chapter 6).

5.2.1 Laboratory experiments

The evolution of institutions

Laboratory experiments usually center around simple "standard" dilemma situations like the Prisoner's Dilemma or a public goods problem. One particular question that has been examined experimentally is that of how punishing institutions can evolve. The evolution of punishing institutions is a riddle because in those situations where a punishing institution would be needed to solve a dilemma, a new dilemma arises that precludes the evolution of punishing institutions. One such constellation has been examined experimentally by Gürerk, Irlenbusch and Rockenbach (Özgür Gürerk et al., 2006). They set up an experiment where subjects interact anonymously in a public goods dilemma for 30 rounds. Each subject can decide which amount of its income to donate for the provision of a public good. The return value was such that each subject profited strongly from the overall contributions, but still had an incentive to let the others pay the public good and not to contribute himself or herself. Typically, the provision of public goods degrades in such a situation after only a few rounds.[11] To make matters more interesting, the subjects could choose to join either of two groups, one group that was provided with a sanctioning institution and one that was sanctioning free. After each round, subjects could choose to change groups. The sanctioning institution worked in the following way: In the group with sanctioning, each participant was allowed to punish or reward other participants within the same group. Both punishment and rewards cost the punishing or rewarding subject one monetary unit. Persons punished would lose three monetary units, while persons rewarded would gain one monetary unit. (Punishments and rewards were issued in the same round after the contributions were made.) The de-

[11] It should be remembered that the provision of a public good is an N-person dilemma. In an N-person dilemma the evolution of reciprocal altruism faces much stronger barriers than in the 2-person dilemma, though it has indeed been demonstrated that there exists a theoretical possibility for conditional cooperation to be stable even in an N-person game (Taylor, 1997, p. 82ff.).

cision to punish or to reward was left entirely at the discretion of the participants. Since punishment was costly, the provision of punishment therefore constituted a second level free rider problem.

Although the authors of the experiment were not primarily concerned with studying evolutionary social processes, their experimental setup does in fact resemble a kind of group selection scenario with two levels of selection. On the within-group level selection takes place between a cooperative and a non cooperative strategy, both of which – as one could say – compete for players adopting them. At the same time a between group selection process takes place between the sanctioning and the non sanctioning group, which compete for group members. (See also chapter 4.3.3.) However, the situation does not exactly reflect the group selection model presented earlier, insofar as the groups do not merely differ in the composition but also by the presence or absence of the sanctioning institution and because – as will be seen – the selection pressure within the sanctioning group does not counteract the between-group selection pressure as it is assumed in the theoretically most interesting case of group selection (see chapter 4.3).

The result of the experiment was that after 30 rounds almost everybody had joined the group with the sanctioning institution and almost everybody cooperated almost entirely (i.e. donated almost all of the income to the provision of the public good). About 3/4 of the subjects in the sanctioning group did exert punishment. (Rewards proved to be far less effective, since they even had a slightly negative influence on cooperation, supposedly, because subjects thus rewarded conclude that they have given too much.) Interestingly, subjects that changed the group quickly adopted the behavior common in their new group, both with regard to cooperation and non cooperation and with regard to punishment and refraining from punishment. The main question that this experiment raises is why punishing behavior did not erode as did cooperation in the non sanctioning group. Given that punishment is a second order public good and that it thus raises a free rider problem that is structurally similar to the original social dilemma situation simulated in the experiment, this appears quite surprising. Several explanations are possible[12]: 1) Human beings are just not so rational as the theory of public goods assumes. Therefore, in some instances (second level problems) they provide goods, even though they would be better off cheating. But then, why didn't a more rational mode of behavior evolve if this would entail greater revenues? 2) The subjects come from a society where certain modes of behavior including punishment, revenge etc. have –

[12]Only some of these are discussed in Gürerk's, Irlenbusch's and Rockenbach's paper.

for whatever reason – already evolved and are now transferred by them to the game. 3) There exists a certain amount of conformism. That is, people imitate other people's behavior and only deviate if they have a strong incentive to do so. As the necessity of punishment decreases over time because people that tried to cheat and were punished have learned their lesson, conformism suffices to uphold punishing behavior before it can deteriorate (in the end the payoff disadvantage of punishers is about 2% compared to non punishing cooperators). In other words, the costs of punishment become small enough to fall under the conformism threshold. 4) There is also a possibility that the second level public goods problem falls into a category where it may even pay for a participant to provide the good all on his or her own even though nobody else is willing to share the cost. In order to find out whether the problem of providing costly punishment falls into this category of public goods problems, it would be necessary to measure the gain in the provision of the first level public good that is effected by the successful betterment of reluctant cooperators.

Having thus briefly discussed the results of a typical study of experimental economics, the question shall now be considered, how this can be related to simulation models of the kind that have been presented previously (chapter 4). There are a few things to say regarding this question: The setup of the experiment does not precisely fit any of the simulation models presented earlier, neither does it closely resemble any other particular simulation study that has been published on the evolution of cooperation. It follows the common pattern of public goods problems as they are also expressed in the respective models that illustrate the theory of public goods. Of course it would be easy to draw up a computer simulation that more or less resembles the experimental setup. But what could the goal and possible benefit of such an endeavor be? As experiments provide *prima facie* stronger evidence than any simulation, why would anything need to be demonstrated by a computer simulation that has already been shown experimentally? One might reason that computer simulations could be helpful for deciding between the four possible explanations given for the punishment cooperation above. But this would only be the case if the decision between these alternative explanations were one that did on one point or other rest on the question of the mere theoretical possibility of any of these and this is not the case, except perhaps for the last alternative for which, however, a simple calculation should suffice. In order to decide between the other three alternative explanations, further experiments or further measurements would be required, but not more models.

Still, experiments of economic behavior provide a type of empirical

research where a close fit between model and empirical reality is comparatively easy to achieve. If indulging in computer simulations of the evolution of altruism appears so little rewarding because it is so far removed from any empirical problem of cooperation or altruism – an impression that was very much strengthened by the earlier discussion of the empirical literature on the evolution of altruism in biology – , experimental economics finally offers a basis where simulation models of the evolution of altruism and empirical research can be linked together in a more than merely metaphorical and story telling way. One might wonder why this should work in economics but not in biology. The probable reason is that for the simulations to be applied in biology it would be necessary to measure the reproduction relevant fitness payoff of certain types of behavior, which obviously is a task that is extremely difficult to accomplish in most cases. The one exceptional example given in Dugatkin's comprehensive empirical meta-study (Dugatkin, 1997) where payoff parameters were actually measured, was an experimental study about blue jays. And even in this case the measured payoff parameters did not resemble a payoff in terms of the reproduction rate (see page 154).

Experimental studies such as the one outlined above can potentially be linked to computer simulations because they take place in an artificial laboratory setting that is streamlined and simple enough to reproduce it in a mathematical model of a computer simulation. But at the same time experimental laboratory studies raise certain epistemological concerns of their own, which are similar to those of computer simulations. Regarding computer simulations of the evolution of cooperation, there exists the problem of transferring the results of the computer simulations to empirical situations. As has been demonstrated in the case of biology (see chapter 5.1) this can be a very difficult problem to solve, especially if the simulations are not designed to fit empirical problems but merely express more or less plausible theoretical assumptions. Now, a similar transfer problem exists for experimental research in economics. For, how are we to know if the behavior of participants in a laboratory experiment is the same as the behavior of people in "real" life? Typically, the laboratory situations are very much simplified compared to the real life situations they are supposed to resemble. Interfering factors such as the psychological factors that drive our behavior in small group interactions are deliberately excluded by putting the participants into small closed boxes, where they sit in front of a computer screen and only receive information about the other participant's choices without ever getting to see their faces or being able to talk to them. Furthermore in many of the experimental studies the participants are university

students and not a representative sample of the population. These few remarks should suffice to indicate that there exists a transfer problem in the case of experimental economical research as well. It seems that when the explanatory gap between models and reality is closed by designing experiments which resemble the models, another gap is opened between experiments and the empirical world outside the experiments.

Trust and cooperation in internet auctions

Can the just mentioned dilemma ever be solved? In fact the dilemma can be solved in certain special cases. It can be narrowed or closed if 1) either, we are lucky and find some empirical setting that is indeed simple enough to be easily compared to laboratory setups, or 2) in cases where economic institutions have deliberately been designed to match a previously tried experimental setup. (For example, in order to exploit a certain experimentally proven effect.) A very prominent example that fits these conditions is provided by the economic research on the behavior of buyers and sellers in internet auctions. Internet auctions provide by their very nature a simple and streamlined setting that strongly resembles that of laboratory experiments. Furthermore, some of the economists that have studied the behavior in internet auctions also work as consultants for internet auction companies like eBay. Therefore, we can also expect that the concrete procedures of such auctions are to some degree designed according to precepts learned from economic experiments.

In the following I describe one series of experiments on the behavior of internet traders that was conducted by Gary E. Bolton, Elena Katok and Axel Ockenfels (Bolton et al., 2004). The problem that their series of experiments is centered around is that of why internet traders trust each other. Described in game theoretical terms an internet auction is an asymmetric one-shot and non zero-sum game. It is asymmetric because it is the rule that first the buyer sends the money and upon receiving the money the seller sends the product to the buyer. This means that the seller can cheat, but not the buyer. If the buyer enters upon the interaction, the buyer must therefore trust the seller. The game is one-shot because typically neither the buyer nor the seller have met before, nor will they be trading partners after the trade has taken place. Finally it is a non zero-sum game because both the buyer and the seller profit from the interaction. If they did not, then either the buyer would not bother to enter upon the interaction or the seller would not offer his product. Bolton, Katok and Ockenfels model these conditions by assuming that both buyer and seller retain a payoff of 35 if no trans-

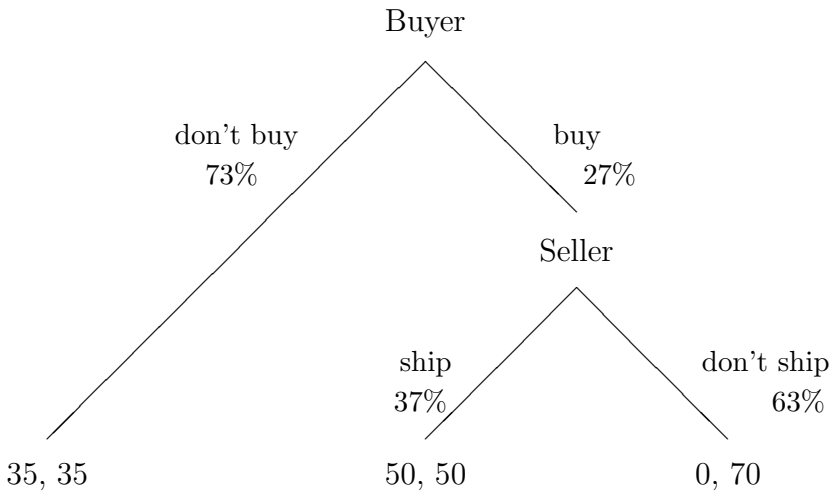

Figure 5.1: The original trust game used in the experiments by Bolton, Katok and Ockenfels. Source: (Bolton et al., 2004). The percentage values indicate how many subjects chose which course of action in the experiment.

action takes place. If the transaction takes place, both buyer and seller receive a payoff of 50. And if the seller cheats that is if the seller takes the money but does not send the product to the buyer, then the seller receives a payoff of 70 while the buyer ends up with a zero payoff. (See figure 5.1.) Except for the asymmetry the situation is thus the same as in the Prisoner's Dilemma game. Theoretically, no interaction should take place. For, if both trading partners were rational egoistic utility maximizers, then the seller would be sure to cheat if an interaction did take place and the buyer, anticipating the seller's cheating, would not even initiate the interaction (Bolton et al., 2004, p. 188).

Now, everyone knows that people in this world (luckily) are not totally rational egoistic utility maximizers, as classical economic theory assumes, but that they are also driven by normative concerns such as fairness considerations. Bolton, Katok and Ockenfels distinguish three different types of such concerns: Fairness in terms of reciprocity, fairness in terms of equal distribution and, finally, collective efficiency concerns. Reciprocity as a fairness concern[13] does in this context mean that the seller might be induced to send the product to the buyer because he or she feels obliged to do so since the buyer has sent the money. Fairness in terms of equal distribution means that the seller cooperates because

[13]This should not be confused with *reciprocal altruism* in evolutionary models of the repeated Prisoner's Dilemma, which does not evolve because of any fairness concern but because it yields the highest payoff in the long run.

otherwise the outcome would result in a very uneven distribution of goods (70 vs. 0 instead of 50 vs. 50). And the seller is driven by efficiency concerns if his reason is that the net result for both players is higher than when cheating (100 vs. 70). The model as it stands does not allow distinguishing between these motives. Therefore, Bolton, Katok and Ockenfels draw up an additional model, where buyer and seller retain 105 and 35 points if no interaction takes place, both end up with an equal payoff of 70 if a trade is made and the seller cheats, and where the buyer earns 120 and the seller 50 points if the seller does not cheat. (See figure 5.2.) From the perspective of rational choice theory this second model is equivalent to the first one: Both trading partners would be better off if the trade took place and the seller did not cheat than if no trade took place at all. At the same time, if the trade is initiated by the buyer, the seller gains most if he cheats, wherefore – anticipating rationality of the seller – the buyer would be best off not to initiate the trade at all. However, with regard to fairness concerns, the buyer would initiate a trade and the seller would cheat if both were driven by a "fairness as equality" ideal, while the seller would not cheat if driven by reciprocity or efficiency concerns (Bolton et al., 2004, p. 191).

In an experiment participants were asked to play one of these two games in either the role of the seller or the role of the buyer (that is no participant played the game twice). While in the first game (where participants receive an equal payoff if the seller cooperates) 37% of the sellers did not cheat, only 7% of the sellers did not cheat in the second game. Interestingly, even though the buyers should expect to be cheated in the second game (just as or even more so than in the first game), they were much more willing to buy in the second game (46%) than in the first game (27%). These results strongly suggest that distributional fairness plays a predominant role in this type of interaction, while efficiency and reciprocity seem to be negligible motives (Bolton et al., 2004, p. 193ff.).

In both games the sellers thus proved to be more trustworthy than their rational self interest would suggest. However, even in the original game the degree of trustworthiness (37%) would not be enough to make the game profitable in monetary terms.[14] Taking the question one step further, Bolton, Katok and Ockenfels proceed to examine how institutional arrangements can influence the development of trust. In the case of online auctions, the primary institution to allow the development of trust is the rating mechanism. To examine the effects of such a mechanism, Bolton, Katok and Ockenfels do, however, start with a setting without such a mechanism. In contrast to the previous experiment the

[14] As can easily be verified, the expected payoff of buying exceeds the payoff of not buying only when the probability of meeting a trustworthy seller is greater than 70%.

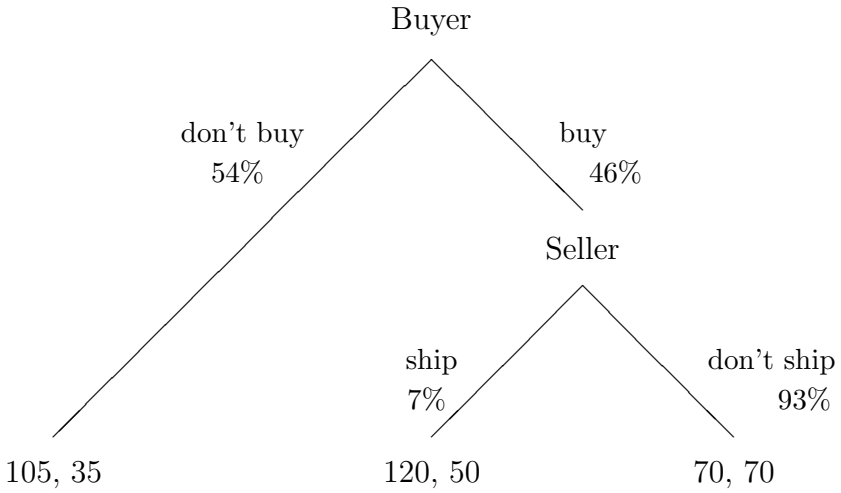

Figure 5.2: A slightly modified variant of the original trust game. Source: (Bolton et al., 2004).

participants play the game repeatedly, but with changing partners and without any information about the previous interactions of the new partner. This setting is called by Bolton, Katok and Ockenfels the *Strangers market* (Bolton et al., 2004, p. 196). The results in the Strangers market are very similar to those in the original experiment (on average 37% of the buyers were willing to buy, while 39% of the sellers actually shipped the product). What the average values conceal is that over time (the participants played the game 30 times) trust collapsed. Obviously, the participants learned that their trust is not sufficiently rewarded in this setting. This was to be expected.

To study the effects of institutional arrangements, Bolton, Katok and Ockenfels contrasted the *Strangers market* with two further settings, the *Reputation market* and the *Partners market*. In the Reputation market, a feedback mechanism was introduced that informed the buyers about all previous interactions of the seller. This is similar to the feedback mechanism in internet auctions such as eBay. Only that in the real internet auctions the feedback consists in a rating by the buyers in previous auctions,[15] while in the experiment the feedback accurately informed about the real behavior of the seller in the experiment. In the Reputation market trust and cooperation did not collapse as in the Strangers market. Instead, 56% of the buyers were willing to enter into

[15]As is well known, the ratings by disappointed buyers are not always fair, which in some cases also leads to lawsuits between buyer and seller.

a trade and 73% of the sellers did not cheat. Interestingly, the rate of cooperation of the sellers is very close to the theoretical borderline of 70% where trade becomes profitable in this game (Bolton et al., 2004, p. 198). In the Partners market, which is distinguished from the Reputation market by the fact that the same partners interact throughout the whole repeated game, the rates of buyer's trust and seller's cooperativeness were yet significantly higher than in the Reputation market (83% and 87%). (Again, this result is unexplainable by normative economic theory based on the rational actor model (Bolton et al., 2004, p. 199).)

The experimental setup that Bolton, Katok and Ockenfels used is still in many respects simpler than the real world situation of internet auctions with a rating system. In internet auctions the seller may not only cheat by not shipping the paid product but also by shipping a product of lower quality than advertised, the information propagated through the rating system may not be completely accurate, both buyers and sellers can still take resort to the legal system if they are unsatisfied, which means that cheaters do not only bear the risk of a bad rating but also that of being sued. Still, the experimental setup comes quite close to what happens in internet auctions. Although it has not been done in this particular study, it is well imaginable to compare the data gathered in this or similar experiments with that gathered from real internet auctions. This would in principle allow checking whether such experiments are realistic.

Conclusions

What can we learn from the experimental research in economics for the explanatory validity of results obtained by computer simulations such as those presented in chapter 4? It has already been noted (chapter 4.1.6) that computer simulations which are not tied to specific empirical constellations can at best prove theoretical possibilities, which as such are often not very informative. One way to link computer simulations to empirical constellations would be to create experimental setups which reflect the simplifying modeling assumptions. (Neither of the previously discussed experiments was of course meant to verify any computer simulations,[16] but given the way these experiments work, one could use similar experiments that match the setup of certain computer simulations.) Of course this requires that the computer simulations use

[16]In fact, it seems that computer simulations do not play a very important role in this branch of research. In the very issue of "Analyse & Kritik" (1/2004) from which Bolton, Katok and Ockenfels' paper (Bolton et al., 2004) was taken and which was as a whole dedicated to the topic of "online cooperation", not a single simulation study appeared among the 17 articles of the issue.

settings that can at least in principle be reproduced experimentally. For population dynamical simulations of tournaments of the 200 times reiterated Prisoner's Dilemma this might turn out to be a bit impractical.

But when one of the restrictions of the method of employing computer simulations is that in the first instance they only allow us to demonstrate theoretical possibilities, then one of the restrictions of the experimental method is that *prima facie* it only allows us to demonstrate *practical possibilities* and that we still do not know how much impact these practical possibilities have outside the laboratory or – to put it simply – how realistic they are. The gap between the demonstration of theoretical or practical possibilities and empirical reality (outside the lab) can under favorable circumstances be closed, either because we are lucky enough to find a constellation in the real world that is simple enough to match our models, or because we examine social institutions that have been designed according to precepts gained by model research and laboratory testing. (Again, these considerations are somewhat tentative and the previously discussed examples of economical experiments do not suffice to fully warrant such conclusions but they should suffice to show their plausibility.)

The question remains, how many of the empirical questions that are of interest to us in the social sciences are of such a kind that they can be tackled with the help simulation models in the way hinted at above.

5.2.2 A real world example: Altruism among enemies?

It has just been argued that there is some hope to link simulation models with empirical reality via laboratory experiments. Usually, however, when it comes to finding real world evidence for models of the evolution of altruism in the social sciences, things start to get difficult. Of course it is easy to think of many situations which more or less resemble a repeated Prisoner's Dilemma (or some other game): the power game of politics for example, or negotiations between opposing political parties when it comes to decisions that need the full consent of all participants. But the problem is that this "more or less" resemblance is simply not enough to explain the situations in question with sparse models such as those described in chapter 4. Rather than enumerating further examples where our models might apply (or might not apply, as the case may be), I am now going to discuss one such example in depth to highlight the (notorious) difficulties that formal modeling faces in the social sciences outside the field of economics.

The example to be discussed is a sort of "classic" of the theory of the evolution of cooperation. It is the "live and let live"-system that

developed at certain stretches of the front line in the trench war of the First World War. The "live and let live" system in the First World War is already discussed in Robert Axelrod's "Evolution of Cooperation" as a prime example for his theory of the "evolution of cooperation" (which is more or less what was here discussed under the heading of "reciprocal altruism"). Because the phenomenon itself is so surprising, it is one of the most stunning examples that have been given for the "evolution of cooperation" in a social science context. Axelrod's exposition of the "live and let live" system has led to much subsequent discussion and criticism most of which centered around the question of whether Axelrod's interpretation of the situation was correct from a game theoretical point of view. Was the situation of the soldiers of the opposing forces really a repeated Prisoner's Dilemma or some other game or, rather, a collective action problem? Were the soldiers of the opposite front lines the players of the reiterated Prisoner's Dilemma or were the soldiers caught in a Prisoner's Dilemma against their own military staff?[17] More important than the problem what kind of game theoretical model can be applied to the "live and let live" system is the question *if* Axelrod's interpretation of the "live and let live" system in terms of evolutionary game theory yields any explanatory power, given that it is by and large correct. Or, to put it more bluntly: Can an explanation in terms of reciprocal altruism give us an explanation of the "live and let live" system that goes beyond what can immediately be inferred from the historical description of the phenomenon alone?

Axelrod's interpretation of the "live and let live" system rests on an extensive historical study of the phenomenon by the sociologist Tony Ashworth (Ashworth, 1980), a debt that Axelrod does, of course, fully acknowledge. Tony Ashworth is neither a game theorist, nor does he try to explain the emergence of the "live and let live" system evolutionarily. Yet, Ashworth does not only describe what happens but also offers an explanation why the "live and let live" system could emerge on a certain front section, how it could be sustained over a considerable period of time and why it eventually broke down again. The crucial question that concerns us here is whether a better explanation for this phenomenon can be given in terms of reciprocal altruism or if at least new light is cast on some of the aspects of the historical events in the First World War that Ashworth has described in his book. In order to answer the question, the explanation that Ashworth offers in his historical treatment must be reconstructed first. For, as it is common in historical literature, description and explanation of the historical events are interwoven

[17]For a summary of the discussion of Axelrod's example in the more game theoretically orientated literature see Schüßler (Schüßler, 1990, p. 33ff.).

in one and the same narrative in Ashworth's book.

Let's first look at the descriptive side and ask the question that all studies in history begin with: What has happend? In our collective memory the First World War is commonly remembered as an unusually brutal and destructive war. It is associated with images of large scale battles, like the battle of Verdun or the battle at the Somme, during which tens of thousands of soldiers died within just a few weeks (James, 2003, p. 52). It is much less known that aside from the scenes of the great battles an astonishing calmness often prevailed over long stretches of the front line. And this calmness prevailed although the soldiers in the trenches virtually eyeballed their opponents on the other side. Moreover, as Ashworth demonstrates in his study, these phases of calmness were not merely the expression of comparatively less intensive fighting but the result of a tacit mutual agreement following a kind of "live and let live" principle. Of course this "live and let live"-system was at no time officially tolerated by the military doctrine and open fraternizing was met with severe disciplinary measures.

But what did the "live and let live" system consist of if open arrangements were impossible? Ashworth identifies several forms that the "live and let live" system could take: The exchange of shells and bullets could be limited to certain times of the day. The shooting could be directed to always the same targets, which the enemy soldiers only needed to avoid getting close to if they wanted to stay alive. Finally, it was possible to miss the opposing soldiers on purpose when ordered to shoot at them. This way the soldiers in the trenches could at the same time report the consumption of ammunition to headquarters and signalize their opponents that they did not really intend to hurt them. All this was of course based on mutuality and the conduct could be changed any minute if the other side did not comply. Ashworth has summarized these aspects of the "live and let live" system under the short formula of the "ritualisation of aggression" (Ashworth, 1980, p. 99ff.). The ritualization of aggression between the opponents was completed by the emergence of a proper ethic among the fellow comrades in arms, according to which "disquieters" or "stirrers" that did not honor the tacit agreement of "live and let live" were hated and disdained (Ashworth, 1980, p. 135ff.).

This was just a very brief outline of the most important aspects of the "live and let live" system. In his book Ashworth discusses many more factors, such as the role of different branches of the armed service and the line of command. But it would lead too far to discuss all these details here, although they are by no means unimportant and it is furthermore by no means unimportant that in the game theoretic analysis all of these subtleties must almost by necessity be left unconsidered.

Now that we have seen what the "live and let live" system consists of, how does Ashworth *explain* it? Because the "live and let live" system was widespread one must expect that it has generic causes (in contradistinction to singular historical causes). According to Ashworth's rough estimate it occurred during one third of the front tours of an average division. This also means that it occurred *only* during one third of the front tours. If one wants to explain why it occurred, one must also explain why in most cases it did not occur. In Ashworth's treatment, the following preconditions and causes for the "live and let live" system can be identified:

1. The strategical 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 impersonal, "bureaucratic structure of aggression" (Ashworth, 1980, p. 76ff.).
4. Empathy with the soldiers on the other side of the front.
5. The "esprit de corps" that can, however, be both either conductive or (in the case of elite troops) impedimental to the emergence of the "live and let live" system.
6. 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.
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 did they find an effective way to do so by organizing raids on the enemy trenches.)
9. Initial causes such as Christmas truces, bad weather periods when fighting was impossible, coincidental temporary ceasefire due to similar daily routines on both sides (for example, same meal times).

But why, then, did not the "live and let live" system occur everywhere and all the time? One could of course think of many plausible answers to this question. Because the "live and let live" system did not comply with the objectives and the very purpose of military warfare it is natural to assume that it was in many cases successfully suppressed by the military

leadership. But as Ashworth is able to demonstrate from the historical sources it was for a long time almost impossible for the military leaders to efficiently suppress what in their eyes must have been a great nuisance to their military mission. It took them quite a while to find the right means to break the "live and let live" system. (But when they finally succeeded in doing so, their success was lasting.) Furthermore, one might assume that the "live and let live" system was quite error prone as no explicit agreements with the other side could be made. But the most decisive factor among the above listed causes for the emergence or non emergence of the "live and let live" system was – according to Ashworth's empirical study – whether the troops involved were elite troops or "regular" troops.[18] Only when non elite troops were facing each other was there a high chance for the "live and let live" system to emerge and to be sustained.

The means by which the military leadership finally managed to break the "live and let live" system was the ordering of raids into the enemy trenches. Raids could not be faked nor could they be ritualized because either the enemy had casualties or the soldiers of one's own side did not come back. And by stirring up emotions of hatred and revenge the raids deprived the "live and let live" system of its emotional foundation in mutual empathy (Ashworth, 1980, p. 176ff.).

So much for Ashworth's historical description of the "live and let live" system and his explanation of these suprising historical events. What can Axelrod's interpretation on the background of the theory of the "Evolution of Cooperation" add to this explanation?

First and foremost Axelrod argues that the situation of the soldiers in the trench warfare can be interpreted as a repeated Prisoner's Dilemma. In order to do so, Axelrod needs to show that the options that were available to the actors in the historical situation correspond to the possible choices of the players in a repeated two person game and are valued by the soldiers in such a way that the game is a Prisoner's Dilemma. That this is indeed the case is demonstrated by Axelrod quite persuasively: In the historical situation single sided defection would mean to fight and meet so little resistance that victory is possible. Clearly, this would be the preferred alternative on any side of the front. Thus, even without assigning particular preference values, we can safely assume that $T > R, P, S$. But if it was not possible to break through the enemy front line then it was certainly better to "keep quiet" as long as the opponents were willing to "keep quiet" because such an arrangement

[18] Among the British troops there was no formal division between elite and non elite, but, as Ashworth points out, military staff as well as the common soldier new fairly well which troop was elite and which was not.

drastically increased the prospects of survival (in Axelrod's formal notation this means that $R > P, S$). Furthermore, mutual abstinence from serious fighting was certainly to be preferred to alternating single sided fighting if that should be considered a viable option at all. Therefore $R > (T + S)/2$ can also be granted. But if the opposing side was not willing to "keep quiet" by ritualizing aggression in the previously described way then it was still better to fight back then to let oneself be overrun ($P > S$).

In order to apply the theory of the "evolution of cooperation" to the situation of the soldiers in the trenches of World War I, some further points need to be clarified such as whether the "game" played really was a *repeated* Prisoner's dilemma, which requires the identity of the players over a longer period of time. Even though the soldiers at the front were periodically exchanged by fresh troops, the predecessors had to familiarize their successors with the situation at their section of the front. Therefore the successors could pick up the "game" exactly at the point where their predecessors had left it. It is a bit less obvious what the evolutionary transmission mechanism that led to the spreading of the "live and let live" system consists of. Axelrod hints to the fact that the system spread over neighboring sections of the front. But, as has been indicated earlier, one may also assume that the "live and let live" system started independently in many different sections of the front. It does not seem to disturb Axelrod that the way the "live and let live" system was initiated and transmitted bears only very little resemblance to the population dynamical transmission mechanism in his simulation model.

Save for this last point it can be granted that Axelrod's analysis is by and large convincing. But in how far does Axelrod's interpretation go beyond Ashworth's study as far as its explanatory power is concerned? If we consider the whole bundle of conditions that Ashworth discusses as causes of the "live and let live" system (see page 177), it becomes obvious that only one of these conditions is captured by Axelrod's game theoretical interpretation. This condition for the "live and let live" system is the strategic situation of the soldiers in the trenches, which Axelrod describes as a repeated Prisoner's Dilemma. It is important to realize that by doing so Axelrod captures only one of many causes for the "live and let live"-system. Therefore, the evolutionary theory of Axelrod cannot reasonably be regarded as an alternative explanation to the one which is offered by Tony Ashworth in his historical narrative. At best, the theory of reciprocal altruism offers a more precise treatment of one

single component of Ashworth's explanation.[19] Whether this is really the case, shall occupy us now.

Is Axelrod at least able to provide a more precise understanding of at least this particular aspect with the help of evolutionary game theory? In order to find out whether such a claim would be warrented it must be examined whether the situation of the soldiers in the trenches can really be described as a repeated Prisoner's Dilemma. Against Axelrod's interpretation the objection has been raised that the front soldiers may have been primarily interested in their own survival after all and that, compared to their survival, being victorious in the battle was much less important to them. Then the soldiers would not really gain any advantage by single sided defection. (The payoff parameter T would be lower or equal the payoff parameter R in Axelrod's notation.) If this interpretation is followed then the problem the soldiers had to solve was a mere coordination problem and not a Prisoner's Dilemma. Independently of how the question is to be answered the objection shows that the assessment of a given situation in terms of game theory is by no means a trivial and unambiguous task. The difficulties become even greater when it comes to estimating concrete values for the different payoff parameters. Axelrod confines himself to establishing the relative proportions of the payoff parameters that are expressed in the two inequalities $T > R > P > S$ and $2R > T + S$, although his model is in fact sensitive to changes in numerical values of the parameters – as has been demonstrated by the simulations in section 4.1.4.

But there exists an even more serious objection to Axelrod's interpretation: The described strategical stalemate was (save for the great battles) more or less the same at all sections of the front line. Nonetheless, the longitudinal analysis showed that the "live and let live" system occurred on average only during roughly one third of the front tours (Ashworth, 1980, p. 171-175). This empirical fact poses a real problem for Axelrod's theory because his theory postulates that in the reiterated Prisoner's Dilemma cooperative strategies will *usually* prevail. However, as the more extensive series of simulations that has been presented earlier (see section 4.1.4) has shown in accordance with earlier criticisms of Axelrod's approach by mathematical game theorists (Binmore, 1998, p. 313ff.), the theoretical foundation for Axelrod's generalizing claim that cooperative strategies like *Tit for Tat* enjoy a high advantage in the repeated Prisoner's Dilemma was lacking. As the results of the simulation series suggest, it is not generally true that cooperative strategies

[19]This is a point that Axelrod seems to be aware of as he mentions that some of the insights of Ashworth's study, such as the emergence of an ethics of cooperation, might be used to extend his theory of the evolution of cooperation.

are the best strategies in the reiterated Prisoner's Dilemma. Depending on the particular circumstances, uncooperative strategies like *Hawk* may be much more successful. It might seem tempting to draw the conclusion that Axelrod's computer model was too crude after all and that our more refined simulation series which suggests an only limited evolutionary success of cooperative strategies is in better accordance with the empirical findings of Ashworth. Thus, while Axelrod's theory in its original form failed it only needed to be refined a little bit on its technical side to make it succeed.

Unfortunately, the epistemological situation is not as simple as that. According to Ashworth, the major factor which determined the occurrence of the "live and let live" was whether the troops involved were elite troops or merely regular soldiers. Whenever elite troops were involved, the "live and let live" system was very unlikely to occur. How can this factor (elite soldiers or non elite soldiers) be reflected in our model? It can be done by assuming that for elite troops a different set of payoff parameters holds because elite soldiers value the viable options (fight hard or "live and let live") according to a set of preferences that differs from that of ordinary soldiers. For example, it is not implausible to assume that elite soldiers might consider it dishonorable to avoid fighting just to save one's own life. But while such an assumption might save our theory it remains doubtful whether much is gained in terms of explanatory power. For, instead of reverting to simple standard assumptions about the payoff parameters in a given strategical situation, it would be necessary to conduct an extensive historical inquiry in order find out how different groups of soldiers may valuate one and the same situation. (In fact, without such an inquiry we might not even be aware that there is such an important difference between elite soldiers and non elite soldiers.) But with the historical inquiry at hand, we would not need a game theoretical model any more to tell us what happend. Or, to put it in another way, almost all of the explanatory work would be done by the theories and historical inquiries needed to determine the payoff parameters, while the game theoretical model making use of this work would be little more than a trivial and illustrating addition. Also, once it is accepted as a fact that it depended on the elite status of the troops whether they would fight or attempt to engage into "live and let live" with their enemies, this fact can be explained more simply than by any game theoretical model by the rather obvious assumption that elite soldiers are more likely to follow orders involving great danger than ordinary soldiers. An assumption that has the additional advantage that it is – other than assumptions about payoff values – empirically very easily testable in comparable circumstances.

The more general lesson to be learned from this is that game theoretical models prove to be useful only in situations where we can either proceed from standard assumptions about the relevant payoff parameters or where reliable measurement procedures for the input parameters of the models exist. Apart from the fact that it leaves out too many causally relevant factors, this is the second reason why the theory of the "evolution of cooperation" fails to explain the sort of cooperation that emerged between the opposing soldiers in the trench warfare of World War I. (And with this second reason it is clear that it does not even provide a partial explanation.)

Following an influential argument from Carl Gustav Hempel (Hempel, 1965) it might still be objected that even though the game theoretical model cannot offer more than an *ex post* explanation, it is still of scientific value because it affords a *general* explanation for a course of historical events and thus increases our understanding of historical processes of a particular kind by subsuming them under general laws or principles. Unfortunately this is not the case here. For, as we have seen, the theory of the "evolution of cooperation" provides hardly an explanation for the emergence of the "live and let live"-system in World War I at all. It is not well possible to defend a wrong explanation or a theory that is not an explanation at all with the argument that it affords a generalization. To say this does not mean that historians and social scientists do not need to or should not be interested in general theories. But in the social sciences and especially in history, generalizations that are meaningful and rich in content are typically found on lower levels of abstraction. One of the standard methods for generating and testing general theories in history is the comparison of similar chains of events under different historical circumstances. For example, it might be interesting to compare the situation in the First World War with that in other wars and with the aim of deriving a generalized theory of fraternization, which could then in turn be applied to the "live and let live"-system and other comparable events. But it seems rather hopeless to seek a general theory for the explanation of the "live and let live" system that is still meaningful and rich enough in content on the level of abstraction of the theory of the "evolution of cooperation".

Summing it up, computer simulations of the "evolution of cooperation" hardly add anything to our understanding of the "live and let live" system in the trench warfare of the First World War. The emergence (or the "evolution", if this term is preferred) of "live and let live" is due to an intricate network of interlocking causes that cannot accurately be explained by reference to simulations of the repeated Prisoner's Dilemma game. At best there exists a vague metaphorical resemblance between

the situation of the soldiers in the trenches and the repeated Prisoner's Dilemma, but this alone is not sufficient for an explanation and it is hardly sufficient to justify the technical effort of a computer simulation in this particular case.

5.3 Conclusions

The previous survey of empirical studies on the evolution of altruism provided some interesting insights in how and why altruism and cooperation can evolve even under unfavorable conditions. Regarding the epistemological merits of simulation models for the explanation of evolutionary altruism, however, the insights gained from looking at the empirical research are extremely sobering: First of all, it is an undeniable fact that computer simulations on the evolution of altruism have remained largely useless for empirical research. And this does of course also mean that computer simulations of the evolution of altruism hardly provide us with any knowledge about how altruism really evolves. This seems to be especially true for repeated Prisoner's Dilemma simulations of reciprocal altruism because they rely on a setting that plays only a very marginal role in nature (see page 146 for one of the few examples where it does). Secondly, the in-depth discussion of two selected examples where the application of simulation models failed despite the serious attempts of its supporters precisely showed why the simulation models failed. In the biological example the model failed because it relies on payoff parameters that could not be measured, while the model is at the same time sensitive to changes of these parameters. That the fitness relevant payoff is very hard to measure is a general difficulty that evolutionary game theory faces in biology, though it does not always turn out to be as fatal as in this instance.[20] In the sociological example the repeated Prisoner's Dilemma model failed because from the many interlocking causes that brought about cooperation between the enemy front soldiers in World War One, it captured at best one cause that could be described as "the strategical situation" of the front soldier. But then it cannot seriously be maintained that cooperation occurred in the trenches in virtue of the very factors for which it evolves in repeated Prisoner's Dilemma simulations. Apart from that, the very same measurement problems and model stability issues that have already been encountered in the biological example reappear in the sociological example as well.

[20] See (Hammerstein, 1998, p. 9ff.) for some reflections on how to remedy this difficulty by means of clever interpretation.

It should not be considered too much of a surprise that the simulation model fared so badly in the sociological example. After all, formal mathematical models can be used in the social sciences only in a few select areas, most notably economics. The reason is that for many explanations that we give in the social sciences, we have to draw on connections for which no formal description exists. One may regret this state of affairs, but it certainly does not get any better by ignoring all factors that cannot be rendered formally. Therefore, in many cases an ordinary historiographical approach may serve the needs of the social scientist much better than a seemingly more refined simulation based approach. Other than that, part of the art of applying formal models in a sociological context certainly consists in picking out the right empirical situations for which a model based approach might indeed be appropriate. How this can possibly be achieved has been hinted at when discussing the internet auction example in section 5.2.1.

All in all, a look into the empirical literature is apt to strengthen some of the skeptical conclusions about computer simulations on the evolution of altruism that have been drawn at the end of the previous chapter (see chapter 4.4), most notably the impression is strengthened that pure model research conveys a distorted picture of how and why altruism evolves. If one really wants to understand how and why altruism evolves then designing models based on "plausible" assumptions alone and uninformed by concrete empirical research is certainly not the way to go. If the simulation based approach to the explanation of the evolution of altruism has thus been a failure then what remains to be clarified is just why it had to fail and what a possible remedy could look like. This is what will occupy us in the next chapter.

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.