Eckhart Arnold

Explaining Altruism

A Simulation-Based Approach and its Limits

www.eckhartarnold.de

Contents

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.