Biol Philos (2011) 26:385–400
DOI 10.1007/s10539-011-9261-7

Why reciprocal altruism is not a kind of group selection

Grant Ramsey · Robert Brandon

Received: 10 October 2010 / Accepted: 7 March 2011 / Published online: 20 March 2011

Abstract Reciprocal altruism was originally formulated in terms of individual selection and most theorists continue to view it in this way. However, this interpretation of reciprocal altruism has been challenged by Sober and Wilson (1998). They argue that reciprocal altruism (as well as all other forms of altruism) evolves by the process of group selection. In this paper, we argue that the original interpretation of reciprocal altruism is the correct one. We accomplish this by arguing that if fitness attaches to (at minimum) entire life cycles, then the kind of fitness exchanges needed to form the group-level in such situations is not available. Reciprocal altruism is thus a result of individual selection and when it evolves, it does so because it is individually advantageous.

Keywords Reciprocal altruism · Group selection · Fitness · Kin selection · Game Theory

Introduction

Altruism has long been seen as the central problem of sociobiology (Wilson 1975a, b). The theoretical problem it raises is *how does it evolve?*, since altruism, by definition, lowers the fitness of the altruist. Empirically, the question is *how prevalent is such behavior in nature?* That question is not, of course, entirely separable from a proper theoretical understanding of the concept of altruism—we need to know exactly what it is that we are looking for in order to see how common it is in nature. The empirical issues will not be dealt with in this paper.

G. Ramsey (✉)
Department of Philosophy, University of Notre Dame, Notre Dame, IN, USA
e-mail: grant.ramsey@nd.edu

R. Brandon
Department of Philosophy, Duke University, Durham, NC, USA

Traditionally there have been three major responses to the theoretical conundrum raised by evolutionary altruism: group selection (Darwin 1871; Wilson 1975a, b, 1980; Wade 1978), kin selection (Hamilton 1963, 1964a, b; Maynard Smith 1964) and reciprocal altruism (Trivers 1971). Although these hypotheses are not mutually exclusive, they have been thought to be distinct, at least when originally presented. But for some time it has been argued that kin selection models of the evolution of altruism can be subsumed under group selection models (Michod 1982; Uyenoyama and Feldman 1980).[1] This mathematical result should not be surprising. It is worth noting that Hamilton, the originator of kin selection theory did not like that term, which was introduced by Maynard Smith (1964), not Hamilton. Hamilton preferred the term "inclusive fitness" theory. Dispersing altruistic benefits non-randomly with respect to kin was simply one way, perhaps the likeliest way, of increasing inclusive fitness. To use Dawkins (1976) example, if the altruistic allele pleiotropically produces a green beard, then one can behave non-randomly with respect to beard color and the altruistic allele will evolve. But in either case the non-random behavior (or, more specifically, non-random dispersal of altruistic benefits) creates groups, often termed "kin groups", that compete with other groups. Furthermore, the explanandum of both theories is the same—a non-calculating form of altruism that requires no reciprocation for its evolution. It may apply equally well to plants[2] (not known for cognitive sophistication) and humans.

Reciprocal altruism, on the other hand, seems clearly distinct from group selection. Reciprocal altruism, as Trivers (1971) conceived it, evolves only within populations of animals that interact frequently and have mutual dependencies. So a moderately long life span and a population structure that brings the same individuals into contact with one another repeatedly is key. Trivers also argued that some cognitive capacity is favorable for the evolution of altruism.[3] Under these circumstances, if I can help you at a small cost, c, and your benefit, b, is large relative to c, i.e., $b \gg c$, then if you reciprocate with the same cost/benefit ratio, we both end up better off than we would have been had we not entered this coalition. Evolution will favor me if I have the propensity to engage in such coalitions and if I am good at picking out likely reciprocators. That is the basic idea and, it seems, it is individual phenotypes that are in competition.

The phenomena explained by this theory—the sorts of altruism explained—are, it seems, quite different from that of group selection theory. For reciprocal altruism to

---

[1] Whether that means they should best be interpreted as group selection is a matter with which we will not deal. It is also beyond the scope of this paper to deal with the related arguments for the unification of inclusive fitness and reciprocal altruism (e.g., Fletcher and Zwick 2006; Queller 1985).

[2] Consider the following scenario. By sequestering nitrogen and diffusing it into the soil a plant may make the very local environment better for its type. But because of the nature of seed dispersal in this species, say big seeds dispersed by wind, its nearest neighbors tend to be closely related. Therefore the benefit dispersed, nitrogen, is dispersed non-randomly with respect to kin and this trait can evolve by "kin group" selection. As Hamilton stressed, it is a mistake to think that kin selection requires any sort of kin recognition.

[3] Strictly speaking, the cognitive capacity is not a necessary condition for evolution of reciprocal altruism. Reciprocal altruism could evolve, for example, in extremely small populations where individuals have very few conspecifics to interact with (Axelrod and Hamilton 1981; Maynard Smith 1982), but such circumstances are far from the norm in nature.

be selected for it needs to be dispersed non-randomly with respect to the probability of reciprocation. (I'll help you, but only if you are likely to help me.) Some have complained that Trivers' theory takes the "altruism out of altruism", and that complaint has merit (e.g., Ghiselin 1974). In contrast, group selection theory explains a quite different phenomenon. As Wilson (1978) emphasized, there is a rift between reciprocal altruism and group-selected altruism—it is group selection but not reciprocal altruism that produces altruism with benefits that can be dispersed indiscriminately within the group and with no requirement of reciprocation. I'll throw myself on a grenade to save my comrades in battle. However, the flip side of within-group altruism is, typically at least, between-group hostility. No one would expect the theory of reciprocal altruism to explain suicidal acts of altruism, while group selection might.

Sober and Wilson (1998) disagree with this mainstream understanding of reciprocal altruism—they argue that it is just another form of group selection. So we want to note at the outset that their conclusion is, on the face of it, implausible given the two very different sorts of altruism involved. Their discussion of reciprocal altruism is subsumed within a discussion of evolutionary game theory. In game theory, they hold that "[f]or those who have become comfortable with the multilevel framework, it is child's play to see the groups in evolutionary game theory, calculate relative fitnesses within and between groups, and determine what evolves on the basis of the balance between levels of selection." (1998, 85) That is, iterated prisoners' dilemma situations, the sort of situations Trivers originally had in mind in his model of reciprocal altruism, are really just another sort of group selection.

Thus, they respond to the major theoretic problem of sociobiology by presenting an incredibly expansive theory of group selection—one that includes all of kin selection and all of what is covered by evolutionary game theory under group selection. They hold that altruism evolves only through a single mechanism, group selection. We think they go too far. Instead of maintaining that the only way that altruism can arise is group selection, we hold that altruism can arise via a plurality processes (reflecting the plurality of phenomena labeled "altruism"). The goal of this paper is not to provide a complete account of these phenomena. Instead, our focus is rather narrow. We will argue that traditional reciprocal altruism, the sort of situation modeled by iterated prisoners' dilemma, is not a form of group selection. Rather it is a form of individual selection.[4]

Our argument has a simple structure. Our main goal is to establish the following conditional: *If fitness attaches to the whole life cycle,*[5] *as opposed to some sub-part thereof, then reciprocal altruism is: (a) individually advantageous; and (b) evolves by individual level selection.* We will be satisfied if the reader is convinced of the conditional's truth. We do believe the antecedent of this conditional is true and will

---

[4] Our argument against reciprocal altruism being group selection thus differs from other challenges to Sober and Wilson, like Godfrey-Smith's (2008) argument that with the appropriate neighborhood structure (but without groups sensu Sober and Wilson), altruism can arise. Skyrms (1994) makes the related and more general point that all one really needs is correlated interaction.

[5] More precisely, the smallest unit that fitness attaches to is the whole life cycle—fitness could attach to a unit greater than the individual life cycle and our argument would go through.

offer two arguments to that effect ("The relationship between Altruism and fitness" and "Why reciprocal Altruism is not group selection"). Thus our primary goal is modest—to establish the conditional. More ambitiously, we would also like to convince some that the antecedent is also true, and then invite them to perform modus ponens. Finally, given the prima facie plausibility of the consequent, and the fact that understanding fitness as attaching to whole life cycles would explain why and how the consequent is true, we seek to add even further (abductive) support for this view of fitness.

Sober and Wilson's argument

The logic of Sober and Wilson's argument is fairly straightforward, and our reply will be straightforward as well. Sober and Wilson (henceforth S&W) follow a mainstream population genetic approach to defining groups (see Uyenoyama and Feldman 1980) in which a group is the smallest unit in which the fitness of individuals is a frequency-dependent function of the composition of the group, and therefore not a frequency-dependent function of the composition of any other group. The idea is that fitness-affecting interactions take place within a group and are contained therein. In standard mathematical models of group selection of altruism this fact could be represented in fitness equations thusly:

$$w_{\mathrm{a}} = 1 + bx - c$$

$$w_{\mathrm{s}} = 1 + bx$$

where 'w_{a}' stands for the fitness of an altruist, 'w_{s}' stands for the fitness of a non-altruist and 'b' and 'c' stand for benefit and cost as before. The variable 'x' stands for the number of altruists in the group. Thus the fitness of any particular individual is a function of the number of altruists in its local group. This is the formal conception of group in operation here. It reflects a perfectly sensible causal idea.[6]

Mereologically, groups are composed of individuals. That does not imply that they are longer lived than the individuals that compose them. This insight led D. S. Wilson to explore his models of trait-group evolution in the 1970's and 80's (Wilson 1975a, b, 1977, 1980). Although his ideas were controversial, we have no quarrel with that particular aspect of S&W's theory.[7] However, S&W have taken

---

[6] Note that these equations assume that the altruist benefits everyone in the group equally, including itself. One certainly need not make that assumption. It is quite possible for the altruism to be distributed non-randomly within the group and have group selection still work (Brandon 1990). However, these equations describe the simplest case. They might well apply to the plant example described in note 2 above. Also they would apply to a case where alarm calls have some cost (increased chance of being noticed by the predator), but everyone in the group, including the caller, benefits (say by mobbing the predator).

[7] We should mention that S&W's group selection model has been critiqued on many fronts. For example, some have argued that their models can be interpreted without invoking group selection (e.g., Gildenhuys 2003), while others have challenged the homogeneity of the category of group selection—see, e.g., the discussion of MLS1 and MLS2 in Okasha (2006). We are not here challenging the basic model of S&W, we are merely arguing that the group structure needed in their model is not present in reciprocal altruism.

this one step further in arguing that iterated reciprocal altruistic coalitions are simply multiple groups in the above sense. We think that this is wrong.

First, let us try to grasp their argument. Take the classical case of a man drowning in a lake, call him R (the potential reciprocator), and have A (the altruist) walking by with easy access to a life preserver. Now A throws a life preserver to R thereby saving R's life. A's cost is minimal, call it c, R's benefit is huge; call it b ($c \ll b$). Let us try to apply the above equations to this case. Costs and benefit are modeled as described, but the group structure needs to be made explicit. R and A form a short-lived group, which exists only for the duration of that brief interaction.[8] For modeling this one shot interaction, the altruist gets no benefit, therefore, $x = 1$ so $w(A) = 1 - c$; while the fitness of the recipient, $w(R)$, $= 1 + b$.

It is important to understand that when the situation is reversed, i.e., A is drowning and R is walking by and R decides to reciprocate, then we have a new group. It has the same members as above, but now R is the altruist and its fitness is $1 - c$. (Conversely for A.) Each temporary interaction constitutes a group. The group is a good, or not so good, group, compared to others. The outcome of selection in the meta-population (i.e. the population that consists of all of the competing groups) is a resultant of within-group and between-group selection. Between-group competition occurs and altruistic groups are favored. But within-group selection always favors selfishness. Why? Because when the group consists of a selfish individual and an altruist in need, the selfish individual avoids the cost of helping. And when the group consists of an altruist and a selfish individual in need, the selfish individual will reap the benefits from the altruist. Theoretically, there is a broad range of conditions where altruism can prevail. Our concern is not when it will prevail, but whether it should be considered group selection.

Sober and Wilson argue that this is group selection because: (1) we do have a population subdivided into groups; and (2) the trait that is evolving would not evolve otherwise, i.e., it is not individually advantageous. We hope to show that both points (1) and (2) are wrong. But before we can do this, we need to explicate the relationship between altruism and fitness.

The relationship between altruism and fitness

Although S&W are not explicit about the concept of fitness they are using in their group selection arguments, their discussion of altruism makes it clear that they are making a questionable assumption about fitness. Consider this quote of theirs: “A behavior is altruistic when it increases the fitness of others and decreases the fitness of the actor.” (1998, 17) This implies that they hold that an altruist changes the

---

[8] For example, Sober and Wilson (1998, 83) state the following: “Is it really fair to call a pair of individuals a group, especially if they interact only briefly with each other before departing, never to meet again? […] The duration of the group is left open in evolutionary game theory, which is sensitive only to the fitnesses of the ‘players’ that result from their interactions within the group. We therefore cannot object to calling an ephemeral group of size $n = 2$ a group, even though it is a far cry from the groups that Wright and Haldane imagined.”

fitness of itself and others when it performs an altruistic behavior.[9] We think this is wrong. We hold that fitness attaches to (at minimum) a whole lifetime, not, as they assume, to some subpart thereof. We will first point out how they are committed to this view of fitness. Next we will offer an analogy that supports our claim that fitness attaches to the whole life cycle.

First, recall the situation with *R* and *A* described above: *A* saves *R* from drowning. We stated that this single interaction constituted a (short-lived) group for S&W. Why? Taking the standard population genetics view of groups, *A*'s fitness at that instant is dependent on *R* and vice versa, but not dependent on the composition of any other group. Thus it is a perfectly good group from that perspective, provided that instantaneous fitness is the appropriate notion of fitness here. We think it is not (see "Objections", objection 3 below). Arguments for that will follow. For now we want to make clear how S&W are committed to this view of fitness. The point we wish to make here is that if one takes fitness as attaching to the whole life-cycle, and if organisms in the relevant population have multiple interactions with multiple partners (which is the idea behind reciprocal altruism), then *R* and *A* do not form a group. *R*'s lifetime fitness depends not just on the behavior (or behavioral dispositions) of *A*, but also on the behaviors (or dispositions) of every other member of the population with which *R* interacts. Likewise for *A*. Why not take (*R, A*) as a group based on all of their interactions during their lives, ignoring their interactions with others, then do likewise for every other pair? To do that is to invoke lifetime fitness, but then to ignore multiple components of that quantity. Remember, *R*'s fitness is not dependent on *A* to the exclusion of other members of the population. And to do so is deal with an explanatorily and predictively inadequate concept. What end would this serve, other than saving the view that reciprocal altruism evolves by group selection?[10]

Now consider the following analogy supporting our view of fitness. There are two pool players, *H* and *S*, who meet one night and begin to play a series of games. Each starts the night with a lump of cash in their pocket and they play for money. They begin the first game with a small bet of $20. *H* looses the game and they play the next game for $50 and, again, *H* looses. They play three more games with subsequent wagers of $100, $150, and $200. *H* again looses each of these games. How well are *H* and *S* faring? What is the status of what we might call their pool fitness? It is tempting to model the situation in the following way: Each player started the night with, say, $1000. A change in this fund can be understood as a change in their fitness. After the first round, *S*'s fitness increased and *H*'s decreased

---

[9] This way of understanding the relationship between fitness and altruism is not restricted to S&W; instead it is a standard way of defining altruism.

[10] For the sake of completeness, let us mention one other possibility that would save S&W's position, namely setting the relevant group to be not the single interaction between *R* and *A*, nor the complete set of their interactions over their lifetimes, but rather a *pair* of interactions. The first interaction is the one we have described: *A* saves *R* from drowning. The second? When *R* saves *A* from drowning. But one difficulty here is that this second interaction may never happen. Furthermore, *A* and *R* may interact in a fitness affecting way, say, 53 times during their lives, the first five of which *A* plays the altruist role. So in this case, the suggestion is we would have to take interactions #1 and #6 and say that that constitutes a spatially and temporally discontinuous group. Again, what motivation might we have for this move? Only to save the position of S&W with respect to reciprocal altruism.

by \$20, after the second, \$50, and so on. By the time they finish their fifth game, *S* has \$1520 and *H* has \$480. *S* has thus become much fitter/wealthier than *H*.

Let's say they play one more game and that *H* convinced *S* to bet \$1500. This final game is decisively won by *H*, which leaves *H* and *S* with evening totals of \$1980 and \$20, respectively. Should we then say that *S* has experienced a precipitous decrease in fitness, while *H*'s fitness has shot up dramatically? It turns out that *H* is a hustler and that he suckered *S* into taking himself to be the better pool player. Thus, it was part of *H*'s strategy to loose the initial games. We feel that the above description of *H* and *S*'s fitness as fluctuating with each game is mistaken. One reason for this is because *H*'s performance in each game is neither predictive nor explanatory of how well he will do in the next game or at the end of the night. His early losses are neither a reflection of poor skills—he is better than *S*, after all—nor a poor strategy. The quality of a player is a function of his ability to strategize and execute the strategy he decides upon—it is not a function of his earnings from any particular game. How, then, should we understand his fitness? The player's fitness is simply a function of his disposition to strategize and execute his strategy. Because this is neither a function of, nor revealed by, a player's performance on any particular game, his fitness should not be understood as increasing or decreasing with the outcomes of these games. There is a non-arbitrary time-scale for accounting here. We should assess *H*'s status right after he leaves town, which, presumably, is shortly after his big win.[11] That is the way pool hustling works.

Evolution is much more complicated than pool hustling, and some qualifications need to be made. But just as in the pool case there is a non-arbitrary time scale for fitness accounting. It is the life cycle. It is the time between reproductive events.[12] Just as the "fitness" of a pool player is a function of the player's disposition to strategize and execute those strategies, an organism's fitness is a function of its disposition to survive and reproduce. How well a player does in a particular game, or how well an organism does in a particular interaction, does not change its fitness for the same reasons. Similarly, it is a mistake to view each of the interactions as an episode of selection that can be multiplied together to obtain the organism's fitness. For even if one takes multiplicativity to be a criterion for episodes of selection (see Arnold and Wade 1984), it is not the case that the episodes in the pool hustling example are multiplicative since, for example, smaller payoffs earlier can make for disproportionally larger payoffs later.

Support for this view of fitness and selection comes from both biologists and philosophers. Biologists often take the fitness of an organism to be fixed by the genes (plus other epigenetic heritable features) and the features of the environment it is born with. Because these features do not change with time, fitness does not fluctuate with time. A Drosophila researcher, for this reason, might study the fitness values of particular genotypes in distinct environments (of varying temperature, say)—and single values are sought for each genotype. Instantaneous fitness values

---

[11] Of course, since fitness is based on dispositions, not actual outcomes, the hustler's tally is not a perfect indicator of the hustler's fitness. Bad luck can occasionally quash even the best hustling strategy, after all.

[12] Overlapping generations are one of the difficult complications. Also, fitness in some cases may be dependent on a super-generational time scale (based on grandoffspring, not offspring, say). But none of the complications affect our basic point against S&W.

are not desired and make little sense in this context. While the fixed nature of fitness is often implicit in biology, it has been argued for explicitly in the philosophical literature (see Ramsey 2006).

Imagine again our friends *A* and *R* by the lake, in which drowning *R* is saved by *A*. Has this act raised *R*'s or lowered or *A*'s fitness? Strictly speaking, neither has occurred. Fitness, as a probabilistic propensity, is fixed and is not affected by any particular event during the organism's life. But if this is the case, the standard way of defining an altruistic act—as decreasing the actor's fitness while increasing the recipient's fitness—will not work. Biological altruism must be redefined in terms of dispositions. Providing such a definition is beyond the scope of this paper.

But even if one is skeptical of fitness considered as a probabilistic propensity, it is still the case that altruistic acts involve neither an increase nor a decrease in fitness. Consider actualized or realized fitness. Has *A*'s fitness been lowered? Our point is that the particular event itself cannot even qualitatively address this question. Once we realize that fitness (realized or not) applies to the whole life cycle we see that *A*'s saving *R*'s life at time *t* in no way lowers its fitness. In fact, reciprocal altruism evolves precisely when *A*'s having the disposition to help others—such as *R*—makes *A* *fitter* than it would otherwise be. Recall the pool hustler. Did his losing game 3 decrease his pool fitness? Thinking of pool fitness as a probabilistic propensity, the answer is no. Losing game 3, losing $100, did not decrease his pool fitness. What if we think of pool fitness as actualized fitness? Then, in the world we described, the answer again is no. (Though there are possible worlds, say where the cops raid the pool hall right after game 3 and close it down, where the answer would be yes).

Why reciprocal altruism is not group selection

Recall the two premises S&W use in concluding that reciprocal altruism is just a species of group selection: In reciprocal altruism (1) the population is subdivided into groups and (2) the trait that is evolving (i.e., altruism) would not evolve otherwise because it is not individually advantageous. Now that we are clear on the relationship between fitness and altruism, we can show why both (1) and (2) are false.

First, if we accept this view that fitness attaches to the whole lifecycle, then *A*'s fitness is in no way a frequency-dependent function of *R*. Instead, *A*'s fitness is going to be determined by the frequency of likely reciprocators in the overall population within which *A* is interacting over its lifetime. Thus the group is the whole population and consequently there is no group selection within this population.

Second, if all of this is correct, what then of point (2)—how can a trait that is individually disadvantageous evolve if not by group selection? Our answer is, basically, that Trivers was right in saying that engaging in reciprocal altruistic coalitions is individually advantageous. As Trivers (1971, 35) put it, "…under certain conditions natural selection favors these altruistic behaviors because in the long run they benefit the organism performing them." What is it for a trait to be individually advantageous? "A trait is individually advantageous relative to some set of alternative traits and to some selective environment if and only if that trait's expected contribution to fitness is not exceeded by any of the alternative traits" (Brandon 1990, 105). When Alxerod compared various computer strategies in iterated prisoners' dilemma games (1980a, b) he was investigating just this question—he was answering the question of what, on average, over the course of the entire series of interactions, would be the fittest strategy.

The following example may make the fundamental motivation of our argument clearer. Consider this scenario: In a species of long-lived birds, the adults pair for a single season and raise their chicks together. The pair-bonds last only for one season, and females perform the mate choice. They choose on the basis of the nests that the males have prepared. The lifespan of the birds is 10 years. Among the males there is a polymorphism in nest building. Some males build nests of mud, other males build nests of sticks. The environment alternates unpredictably between cold rainy summers and hot dry summers. Mud nests are good in the hot dry summers as they provide good thermal insulation during the hot days. But during rainy summers they wash away and the young die. In contrast, the stick nests work well during the rainy summers, but provide no insulation during the hot dry summers and lead to high mortality. There are three choice rules operating among the females during mating season: (1) Always choose stick building males, (2) always choose mud building males, and (3) choose stick building males in cold wet summers, and mud building males in hot dry summers. Let us classify the females into three phenotypic classes, F_{S}, F_{M}, and F_{SM}—where these phenotypes exhibit only rules (1), (2), or (3), respectively.

Let's further assume that whatever genetic variation exists among the males for nest building is independent of the genetic variation among the females for mate choice. We will say nothing more about the underlying genetics, for our purpose here is not to specify an exact evolutionary outcome, but rather to say something about selection.

Now we can think of this example as one involving S&W-style group selection. Each season mating pairs, i.e., groups, are formed. And we will classify the males into two classes M_{M} (mud) and M_{S} (sticks). Thus we have 6 combinatorially possible groupings: (M_{M}, F_{S}); (M_{M}, F_{M}); (M_{M}, F_{SM}); (M_{S}, F_{S}); (M_{S}, F_{M}); and (M_{S}, F_{SM}). Of the six groupings two never occur, namely the (M_{M}, F_{S}) and the (M_{S}, F_{M}), because of the nature of phenotypes F_{S} and F_{M}. There are two environmental states that randomly fluctuate from year to year: E_{C} (cold and wet) and E_{H} (hot and dry).

Although we have argued for the idea that fitness attaches to the entire life cycle, which in this case lasts 10 years, to get S&W-style groups we will need to assign fitnesses on a seasonal basis. On a seasonal basis, the female phenotype turns out to be irrelevant, but the male phenotype highly relevant. To make this as dramatic and clear as possible we have the following fitnesses of the four possible groups in the two possible environments:

$$\begin{aligned} w(M_{\mathrm{M}}, F_{\mathrm{M}})/E_{\mathrm{C}} = 0 \text{ and } w(M_{\mathrm{S}}, F_{\mathrm{S}})/E_{\mathrm{H}} = 0, \quad \text{while} \\ w(M_{\mathrm{S}}, F_{\mathrm{S}})/E_{\mathrm{C}} = 1 \text{ and } w(M_{\mathrm{S}}, F_{\mathrm{SM}})/E_{\mathrm{C}} = 1, \\ w(M_{\mathrm{M}}, F_{\mathrm{M}})/E_{\mathrm{H}} = 1 \text{ and } w(M_{\mathrm{M}}, F_{\mathrm{SM}})/E_{\mathrm{H}} = 1. \end{aligned}$$

Where $w(M_M, F_M)/E_C$ denotes the fitness of a group which consists of an M_M male and a F_M female in the E_C environment, i.e. a mud nest building male and a female that always chooses such males in a cold and wet environment, and so on for the other fitness designations. Two of the possible combinations—$(M_S, F_{SM})/E_H$ and $(M_M, F_{SM})/E_C$—will not occur, since F_{SM} females will not pair with an M_S male in an E_H environment or with an M_M male in an E_C environment. Although there is no question of altruism here, one might well think of this as group selection. After all, in sexual reproduction, neither male, nor female can go it alone.[13] A successful pair depends on a good grouping. In our case a good group is determined by a good match between the nest that the male has built and the weather during the season in which the chicks are raised. So during cold rainy years (E_C), groups that have a M_S in them are favored, regardless of the female half of the group. Conversely in hot dry years.

If the environment were stable, then although we have left the underlying genetics unspecified, the evolutionary response would be easy to predict under a wide range of possible genetic realizations. If the E_C condition were to dominate then we would expect strong selection of the (M_S, F_S) and (M_S, F_{SM}) groups over the (M_M, F_M) groups. Recall that we did specify that the genetic variation of the male phenotype was independent of the female phenotype. So this group selection would, presumably, have effects on the distribution of individual phenotypes in the population, with the prediction being that M_S males would come to dominate M_M males and that F_M females should be at a large selective disadvantage compared to F_S and F_{SM} females. Similarly, were the E_H condition to dominate, then M_M males would be selectively favored over M_S males and F_S females would be at a selective disadvantage relative to F_M and F_{SM} females.

But that is not the scenario we have described, rather we have described a case where the environmental states randomly fluctuate from season to season. In this scenario no set of groups is consistently favored. Are there therefore no selective effects on the distribution of individual phenotypes? On males there is not, but on females there is. Notice that in cold rainy years (E_C) females with the F_M phenotype are disfavored while in hot dry years females with the F_S phenotype are disfavored. In other words, in E_C selection does not discriminate between F_S and F_{SM} females while selecting against F_M females. Likewise in E_C selection does not discriminate between F_M and F_{SM} females while selecting against F_S females. Put in terms of group fitness: Females of phenotype F_{SM} disproportionately find themselves in good groups. One can clearly see this by looking at the group fitness functions above.

Now this should remind one of group selection for altruism. Groups with higher numbers of altruists are favored in competition with groups with lower numbers of altruists. This is the process of group selection. By the very nature of this process, altruists find themselves disproportionately represented in the selectively favored groups. This is not luck. Rather, it is exactly what makes a group a good group. As so our case is a good case of group selection and supports the S&W view—or so it would seem.

---

[13] It is for this reason that Evelyn Fox Keller (1987) argued that sexual reproduction involves group selection.

But we think the causal story sketched in the last paragraph concerning group selection for altruism has no analogue in our bird case. That, furthermore, the attempt to give an analogous causal story in this case will severely muddy some rather clear causal facts.

Trying to account for this on a seasonal basis puts the causal pants on the males, but—as is perhaps already clear—they are just along for the ride. Perhaps drift will eliminate one of the two variants, but barring that, both will persist. Conceived of as a lifetime probabilistic expectation, $w(M_M) = w(M_S)$. There is no fitness difference among the males. This is not to say that what they do is not causally crucial in determining life and death of chicks each year, just that being one type instead of the other does not differentially determine lifetime chick production. That differential determination is what selection acts on, and that is what selection picks out in our scenario. Over the whole life cycle the female phenotypes are competing. In the environment we have envisioned—one that randomly fluctuates back and forth between E_C and E_H—phenotype F_{SM} is clearly superior to F_S and F_M. It picks the right sort of male to mate with for the season based on the prevailing environmental conditions, whereas phenotypes F_S and F_M are stuck with always picking one certain type of mate, regardless of prevailing conditions. That is a strategy that would work in a stable environment, but not in the case we described.

We maintain that the above scenario is a classic one of individual selection. There is competition among individuals—here among females. The phenotype in question is mate choice and the alternative trait values are F_S, F_M and F_{SM} above. Clearly F_{SM} is the best trait value. It is the best among these alternatives, and so, relative to them, it is individually advantageous.

Furthermore, there is no group selection going on here. Why not? The S&W-style groups we described, i.e., the seasonal mated pairs, are not groups in the classical population genetic sense once one treats fitness as attaching to the whole life cycle. Why not? Because the fitness of a particular female depends not on her mate in one particular season, but rather on the relative frequency of potential mates over her lifetime. This is particularly obvious in the case of phenotype F_{SM}. She “wants” to be in a population with a good supply of M_M and M_S males so that she can make the appropriate choice each season.[14] The particular mate she is with in one particular season is no more determinative of her fitness than is the result on a single game of pool on a night when I decide to hustle you. In this case one can give a S&W-style group selection account of what is going on in our population of birds, but such an account gets the causal facts wrong.

Our argument against reciprocal altruism being group selection is in effect an argument against all iterated game theoretic interactions being modeled as group selection. There is a group selection story to be told in such cases, but the story gets the causal facts wrong. In such instances, only an individual selection story truly captures the causal structure of the process.

---

[14] Although our simple model does not have costs associated with searching for rare mate types, one could add the cost of a female, say, finding their appropriate male. Such costs would make it more realistic, but would not alter the basic point that we are making.

Objections

There are five key objections that we will reply to in this section.

Objection 1: You have redefined altruism in terms of dispositions to engage in certain interactions in specific ways. But this redefinition of altruism looses the key sense of what altruism is, which is an exchange of something—if not fitness, than resources like food or energy.

Reply: We are not denying that reciprocal altruism—or other iterated games—involve exchanges of some kind or another. Because these games involve exchanges, it is tempting to define altruism in terms of such exchanges. But there are two reasons why one should not do this.

First, not all unequal exchanges should be considered altruistic. If one organism gives all of its food to the other, we are tempted to label this altruism, but if the organism gives all of its parasites to the other, we would hesitate to classify this "gift" as altruistic. Why is this so? Because being disposed to receive food makes one fitter than being disposed to receive parasites. Our point, then, is that any attempt to define altruism with respect to exchanges of things other than fitness will invariantly be reliant on the concept of fitness. And if this is the case, there is no way to redefine altruism in such a way that is not based on the concept of fitness.

Second, such redescriptions based on exchanges are myopic. Consider the following example: A chimpanzee is observed to climb a tree, procure fruit, descend the tree with fruit in hand, and give a portion of the fruit to an older invalid chimp, too weak to procure the fruit itself. This undoubtedly is a case of altruism, right? We feel that such cases are too thinly described. What are the effects on the supposed altruist in such a situation? If the other members of the group observe this act and consequently hold the altruist in high regard, increasing its status in the group, then such an act might give the chimp a better chance at reproductive success. Thus, only a thick description—one that goes beyond the locus of the exchange and traces the full impact of the act—can be used to determine whether an act is altruistic.

Objection 2: Your reply to Objection 1 shows that fitness needs to play a central role in the definition of altruism. And you have argued above that fitness attaches to whole lifetimes, not subparts thereof. But an organism that becomes sterile from an altruistic act has surely lowered its fitness. There is no way that becoming sterile is part of a successful strategy like the hustler's early losses. An organism cannot recover from sterility. Similarly, there are many kinds of irreversible bodily or psychological traumas that may be the result of altruistic acts. Therefore, in some cases fitness does in fact decrease. And it is these decreases and the concomitant increases in the recipients' fitness that altruism can be based on.

Reply: Although it is counterintuitive that sterility does not involve a change in fitness, we hold that it does not. Lifetime fitness is based on the dispositions of the organisms to survive and reproduce. Thus, an organism's fitness is based not on the outcomes it happens to realize, but instead on the set of possible lives it might live. (This account of fitness was articulated by Ramsey (2006) and is dubbed "Block Fitness.") This set of possibilities includes some lives in which the individual dies or becomes sterile while still young, while other lives are long and include many successful acts of reproduction. What explains why the organism realized one possible life and not the other is not fitness, but chance. It happened to encounter the heterogeneity in the environment in one way and not the other.

Objection 3: Although it may be true that the correct organismic fitness concept is Block Fitness, it is nonetheless true that an organism's instantaneous probability to survive and reproduce does change. It undoubtedly decreases, for example, if an organism becomes ill or sterile.

Reply: We will not deny that such a measure does exist. In fact, Ramsey (2006) dubbed this *Flux Fitness*. Flux Fitness tracks the instantaneous probability of survival and reproduction of individuals. The trouble with this concept is not its coherence or lack of operationalizability. Rather, it is the fact that Flux Fitness cannot be the fitness concept at the center of natural selection. If fitness is to explain and predict the fates of organisms or organism types, then a different fitness concept is needed. Instead of explaining how well individuals do, Flux Fitness merely tracks how well individuals are doing. Because we want to understand how altruism can evolve by natural selection, the concept of fitness that altruism is based on must be the same as the concept of fitness in the theory of natural selection. For this reason, it makes no sense to define altruism in terms of Flux Fitness. (Again recall our pool hustler analogy.)

Objection 4: In calculating Block Fitness, you are averaging over many contexts. You are thus committing the averaging fallacy. As S&W put it: "The behavior that confers the highest fitness is assumed to evolve by individual selection, without requiring group selection. The presence of groups in the model is simply ignored. This is the averaging fallacy, pure and simple." (1998, 83).

Reply: The "averaging fallacy," as described by S&W, is the fallacy of averaging over the within and between group fitnesses. S&W's averaging fallacy is an example of the more general fallacy of comparing fitnesses across selectively heterogeneous environments (Brandon 1990). So when there is genuine group structure, i.e., when individual fitnesses are a frequency-dependent function of local groups, then it is indeed uninformative to average across those local groups. However, one of our main points has been that this sort of group structure is lacking in the traditional models of reciprocal altruism. The iteration of interaction homogenizes the selective environment for the interacting individuals—making this genuine individual selection and not group selection.

Furthermore, we should point out that one may be tempted to argue that the fitness of a particular individual is context-sensitive, that group selection just is context-sensitive selection in which the differing contexts are realized by differing compositions of conspecifics, and therefore that because reciprocal altruism is context-sensitive, it involves group selection. The reason that this argument is specious is that mere context sensitivity is too permissive a criterion for group selection. As Brandon (1990) has argued, context-sensitive selection includes group selection, but it also includes clearly distinct selection regimes, such as frequency-dependent selection.

Objection 5: Your arguments against reciprocal altruism being group selection are in fact not just arguments against all of evolutionary game theory being group selection, but against the possibility of group selection in general. Your argument is thus an a priori *rejection of the possibility of group selection.*

Reply: We are in no way arguing against the possibility of group selection. Our arguments are not a priori ways of showing that all of what S&W take to be group selection are cases of individual selection. On the contrary, we feel that S&W's models are quite tenable. Their mistake is just to be too permissive in their criteria for what constitutes a group. We hold that in order to say that two individuals belong to different groups, it must be the case that there is a difference in their *lifetime* fitness based on this difference. In iterated games like reciprocal altruism, it is only the constitution of the individuals in an organism's population—percent altruists, say—and the way it is disposed to interact with them that determine its fitness.

But there are cases in which being in one group rather than another is genuinely relevant to an individual's lifetime fitness. The trematode parasite—the "brain worm"—example that S&W (1998) use in their first chapter is such a case. In this instance, it is not merely the proportion of individuals in the entire population that determines the individual's fitness. Rather, it is also the constituents of the group it finds itself in. Thus, unlike reciprocal altruism, the group structure is an important determinant of fitness. This is therefore a genuine case of group selection.

Furthermore, there are cases involving repeated, reciprocal interactions that are justifiably considered group selection. If, for example, a pair of birds mates for life, these birds could have repeated (perhaps seasonal) interactions over the course of their lives. Whether or not the interactions are iterated or not is immaterial. The question is whether lifetime fitness values are affected by the pairing and not just by population-level criteria, such as the overall frequencies of types in the population. If lifetime fitness values are affected, then group selection is in fact occurring.

As we said above, we do not believe (as have some critics of S&W) that groups must be long-lived relative to the organisms that compose them. We are perfectly happy with the possibility of group selection acting on groups, say of caterpillars living in small local groups on plants, which exist for only a small portion of the life cycle of the organism. What we insist on, however, is that the frequency-dependence of fitness within the group that defines that group be so defined with a coherent, explanatory, defensible notion of fitness. That, we have suggested here, is a notion that attaches to the whole life cycle, not to some sub part thereof.

Conclusion

We have argued that a proper understanding of biological fitness is one that attaches to the whole lifetime of organisms, and that it follows from this that iterated game theoretic interactions—in particular reciprocal altruism—are not cases of group selection. Instead, only individual selection is involved.

Our arguments are not arguments against group selection in general and are not arguments against the possibility of altruism evolving by group selection in particular. Like S&W, we think group selection has been too often dismissed by faulty arguments; like them we think the models for group selection are logically/mathematically cogent and need to be taken seriously. But by classifying nearly every interaction between individuals as group selection, the empirical question of the role of group selection in, say, human evolution becomes much less interesting.

We think that the following are interesting and important empirical questions: (1) *To what extent has altruism evolved in nature?* (2) *Of the altruism that has evolved, how much has evolved by group selection and how much by individual selection?* Sober and Wilson are correct in arguing that the first question is a genuine empirical question—that attempts to define altruism out of existence, to give a priori reasons why altruism cannot exist, are mistaken. But the account of altruism that S&W provide does not allow one to empirically investigate the second question. This is because S&W's account understands altruism only in terms of group selection.[15] Under our account the second question becomes a genuine empirical question, allowing one to frame these important questions about the evolution of altruism and providing for the possibility of answers to them.

Acknowledgments We wish to thank Charles Goodnight, David Sloan Wilson, and an anonymous referee for helpful critical comments on an earlier draft of this paper. A previous version of this essay was presented to the Duke Philosophy of Biology Reading Group—we thank them for their valuable discussion of our paper.

References

Arnold SJ, Wade MJ (1984) On the measurement of natural and sexual selection: theory. Evol 38:709–719
Axelrod R (1980a) Effective choice in the prisoner's dilemma. J Confl Resolut 24:3–25
Axelrod R (1980b) More effective choice in the prisoner's dilemma. J Confl Resolut 24:379–403
Axelrod R, Hamilton WD (1981) The evolution of cooperation. Science 211:1390–1396
Brandon RN (1990) Adaptation and environment. Princeton University Press, Princeton NJ
Darwin C (1871) The descent of man, and selection in relation to sex. J. Murray, London
Dawkins R (1976) The selfish gene. Oxford University Press, Oxford
Fletcher J, Zwick M (2006) Unifying the theories of inclusive fitness and reciprocal Altruism. Am Nat 168:252–262
Ghiselin MT (1974) The economy of nature and the evolution of sex. University of California Press, Berkeley
Gildenhuys P (2003) The evolution of Altruism: the Sober/Wilson model. Phil Sci 70:27–48
Godfrey-Smith P (2008) Varieties of population structure and the levels of selection. Brit J Phil Sci 59:25–50
Hamilton WD (1963) The evolution of altruistic behavior. Am Nat 97:354–356
Hamilton WD (1964a) The genetical evolution of social behavior I. J Theor Biol 7:1–16
Hamilton WD (1964b) The genetical evolution of social behavior II. J Theor Biol 7:17–52
Keller EF (1987) Reproduction and the Central Project of Evolutionary Theory. Biol Philos 2:383–396
Maynard Smith J (1964) Group selection and kin selection. Nature 201:1145–1146
Michod RE (1982) The theory of kin selection. Annu Rev Ecol Syst 13:23–55
Okasha O (2006) Evolution and the levels of selection. Oxford University Press, Oxford
Queller DC (1985) Kinship, reciprocity and synergism in the evolution of social behavior. Nature 318:366–367
Ramsey G (2006) Block fitness. Stud Hist Phil Sci C 37:484–498
Skyrms B (1994) Darwin meets the logic of decision: correlation in evolutionary game theory. Phil Sci 61:503–528
Smith MJ (1982) Evolution and the theory of games. Cambridge University Press

---

[15] Wilson (1978), for example, argued that most human altruism is reciprocal altruism and not group-selected altruism. Whether or not this is true, we think that it is a genuine empirical question.

Sober E, Wilson DS (1998) Unto others: the evolution and psychology of unselfish behavior. Harvard University Press, Cambridge, Mass

Trivers RL (1971) The evolution of reciprocal altruism. Q Rev Biol 46:35–57

Uyenoyama M, Feldman MW (1980) Theories of Kin and group selection a population genetics perspective. Theoretical Population Biol

Wade MJ (1978) A critical review of models of group selection. Q Rev Biol 53:101–114

Wilson DS (1975a) A theory of group selection. Proc Natl Acad Sci 72:143–146

Wilson EO (1975b) Sociobiology: the new synthesis. Belknap Press of Harvard University Press, Cambridge, Mass

Wilson DS (1977) Structured demes and the evolution of group advantageous traits. Am Nat 111:157–185

Wilson EO (1978) On Human Nature. Harvard University Press, Cambridge

Wilson DS (1980) The natural selection of populations and communities. Benjamin/Cummings, Menlo Park, CA