*Brit. J. Phil. Sci.* **50** (1999), 457–466

REVIEW ARTICLE

# Altruism[1]

Neven Sesardic

## 1 Introduction

The belief in the existence of genuine altruism is still widely regarded as an underdog theory. This is well reflected in the fact that the whole debate about egoism and altruism is frequently conceptualized as being about the so-called *paradox of altruism*. The obvious suggestion here is that the cards are so heavily stacked against altruism that the easiest way to resolve the controversy would be to simply agree that altruism does not exist at all. In their book *Unto Others*, the philosopher Elliott Sober and the biologist David Sloan Wilson make a strong effort to swim against this current.

The battle between altruism and egoism is fought on two separate fronts: in evolutionary biology and in psychology. The book covers both aspects of the debate: the first part deals with biology, the second part with psychology. Although the definition of altruism in biology significantly differs from the concept of altruism in psychology, the authors have shown that the two strands of the discussion nevertheless remain interrelated to such a degree that the integration of both topics into one book makes perfect sense.

[1] Review of Elliott Sober and David Sloan Wilson [1998]: *Unto Others: The Evolution and Psychology of Unselfish Behavior*, London/Cambridge, MA: Harvard University Press, cloth £19.95/$29.95, ISBN: 0 674 93046 0.

## 2 Evolutionary altruism

For defining evolutionary altruism crucial concepts are *behavioural effects* and *fitness*: an organism is behaving altruistically in the *evolutionary* sense if and only if the effect of A’s behaviour is an increase of fitness of some other organisms at the expense of its own fitness. If this kind of altruistic behavioural disposition is selected for, it must be the product of *group selection*. Does this ever happen, and if yes, how often? Sober and Wilson argue that the change of paradigm is already under way, and that on the basis of accumulated empirical evidence group selection should be recognized as an important evolutionary force. In their opinion, the remaining reluctance of some biologists and philosophers to accept the existence of the group selection processes springs from two sources: (1) a historical confusion and (2) a conceptual mistake they dub ‘the averaging fallacy’.

### 2.1 Historical confusion

The group selection controversy started with all the parties agreeing that group selection is *empirically possible*. The opponents of group selection claimed only that the conditions necessary for the operation of group selection are so fine-tuned and rarely met in nature that this kind of evolutionary explanation should always be regarded as highly suspect (at best). In other words, the critics’ charge was merely that group-selectionists tended to accept uncritically a very implausible theory, and certainly not that the theory itself could be dismissed *a priori*. In a fascinating and persuasive historical reconstruction S. & W. show that in the course of a decade or two the fundamental reason for opposing group selection changed drastically. The most surprising thing about it is that, apparently, the participants of the debate themselves were (and still are) largely unaware of that momentous argumentative shift.

Essentially what happened is the following. After empirical research started to turn up a number of *prima facie* clear examples of group selection (under the definition of that term universally accepted at the time) the opponents of group selection found a strange way to remain unconvinced. They simply reinterpreted all these examples so that in the new account no groups were mentioned, and the description was in each case entirely restricted to describing what happened to individuals. The reinterpretation was meant to demonstrate that, contrary to the first appearance, there was no need to invoke group selection even under this best case scenario. S. & W. point out, quite correctly, that if this kind of move is allowed then the existence of group selection actually ceases to be an empirical issue. For we know in advance that no matter what kind of natural selection process is occurring it is always possible to describe the facts of the evolutionary change just in terms of differential reproduction of individuals or differential replication of genes—i.e. without resorting to 'group talk' at all. Doing this, however, amounts to changing the definition of 'group selection', and looks very much like shifting the goalposts during the game. It is easy to win under such conditions: Heads, group selection is empirically unlikely; tails, it is conceptually ruled out.

S. & W. give many illustrations of that peculiar gestalt switch. Let me try to supplement their database with the following curious example involving Richard Dawkins. When S. & W. first presented the basic ideas of *Unto Others* in a target essay for *Behavioral and Brain Sciences* in 1994, Dawkins was one of the most critical voices in the open peer commentary. Genes are replicators (the main actors in any selection story), he said, and groups are mere *vehicles*, entities too ephemeral and transient to be evolutionarily interesting. Clearly implying that he has always been consistent in denying the theoretical importance of groups-as-vehicles, Dawkins wrote: 'I coined the [term] vehicle not to praise it but to bury it.' But as a matter of fact he does not seem to be a very reliable witness about this. For from some of his earlier discussions of group selection it transpires, quite to the contrary, that Dawkins did *not* coin the term 'vehicle' with the intention to bury it immediately, but that he was prepared to let further empirical inquiry answer the question about the selection of vehicles at a given level, one way or the other. For instance, commenting on the philosopher John Mackie's striking suggestion ([1978]) that the well-known cheat-sucker-grudger example from the *Selfish Gene* actually represented a group selection process, Dawkins wrote in 1981: 'It is too early to say, yet, whether formal mathematical models will uphold this possibility, but if they do, Mackie's paper in *Philosophy* will have to be seen as a useful contribution to biology' (p. 564). In the same context, Dawkins also mentioned very favourably the work of M. J. Wade who started the group selection revival in the late 1970s. All this shows that at the time, at least for a while, Dawkins kept an open mind towards the group selection theory and regarded it as a very interesting biological hypothesis. His later dismissive tone is never based on a detailed empirical criticism of the work of Wade (or later Goodnight), but is typically grounded in a newly constructed aprioristic argument against the very idea of group selection. Dawkins is only one of those who went through such a radical conversion without actually noticing it. When these people now look back at the first stage of the group selection debate they must feel like the person in Steven Wright's joke who said: 'Right now I'm having *déjà vu* and amnesia at the same time. I think I've forgotten this before.'

## 2.2 Averaging fallacy

One thing should be clear, though. Even if we agree that the critics of group selection are guilty of the aforementioned historical confusion (i.e. that they changed their theoretical argument against group selection without acknowledging it, or even without being aware of it), this by itself does not show that there is something wrong with their *currently adopted* views on the matter. Therefore, S. & W. develop a separate argument to show that the contemporary opposition to group selection is based on a fallacy ('the averaging fallacy'). But somehow their diagnosis seems to raise more questions than their historical analysis.

The contemporary opponents of group selection ask a simple question: if any description of evolutionary change involving group-perspective can be replaced by an equivalent individualistic or gene-centred description, why should we then ever adopt the group-level viewpoint? S. & W. reply: the individualistic and the group description of the same evolutionary process are indeed equivalent in terms of *what* has changed over time, but not necessarily in terms of *why* it changed. They argue that for certain processes the group-level description may be the only way to disclose the evolutionary *forces* at work, and that in such cases strictly individualistic stories will remain essentially incomplete, as explanations. That is, if it is group-level properties that are *causing* the change, an individualistic redescription of the process may well yield the final result completely accurately (in numerical terms), but it will still be inherently lacking in explanatory force.

So, what is the 'averaging fallacy'? According to S. & W., it is the mistaken assumption that one can use fitness averaged across groups to define individual selection, and that a higher-level account of an evolutionary change can always be replaced by the corresponding lower-level description. In some cases, they concede, the two approaches may be equivalent, and the group-level perspective may indeed be dispensable. But in the case of group adaptations, they insist, a lower-level description that 'averages' the group process over individuals will inevitably omit the most important *causal* aspects of the story.

S. & W. say two things about the 'averaging fallacy' that create a problem for accepting their diagnosis. They claim, on one hand, that 'the controversy over group selection and altruism in biology can be largely resolved simply by avoiding the averaging fallacy' (p. 34), but on the other hand they say that 'no one has tried to defend the averaging fallacy in its general form' (p. 157; *cf*. p. 33). This sounds strange for two reasons. First, how can the averaging fallacy be the main source of resistance to group selection if no one has defended it in its general form? And second, what then *are* the real grounds of those who oppose group selection in general terms? The first question is easier to deal with. S. & W. would probably say that the averaging fallacy is typically committed in particular cases without its ever being defended (or even being thought of) as a general claim. The suggestion would be that it is the example of not seeing the forest (the fallacy) for the trees (particular cases). Well, perhaps. However, I think the second question is tougher. Contrary to what S. & W. say, some biologists and philosophers *did* try to defend the averaging fallacy in its general form. The reason, of course, is that they did not agree that it really was a fallacy. Somewhat inconsistently, S. & W. themselves attribute the general averaging fallacy to Dawkins (p. 125). Also, in a footnote on p. 341 they mention a well-known article by Kim Sterelny and Philip Kitcher ([1988]) in which the strategy of averaging was declared to be entirely appropriate. So, after all, at least some resistance to group selection seems to be based on the fact that certain people do not see anything wrong with the averaging approach.

The basic suggestion as to how to make do with only one-level (gene-level) selection is to allow for different *components* of genic selection (e.g. 'which environment a gene is in' and 'how well it does in its environment'). In this way one commits no fallacy since one shows that one is aware of the obligation to distinguish different causal processes, and that one is indeed trying to distinguish them. But a different problem arises. If genetic selectionism armed with this talk of components of genic selection offers nothing more than what Robert Brandon so aptly called 'post hoc redescriptions of selection scenarios that have already been developed within the [hierarchical] framework' ([1990], p. 157), then it is not really an alternative, self-contained approach but rather a theory entirely parasitic on the hierarchical selection perspective. Therefore, genic selectionists face a dilemma. If they stick to the austere, purely genetic account they will be unable to explain organismic and group adaptations: in that case their theory will be seriously inadequate because it will fail to deal with important biological *explananda*. On the other hand, if they broaden their account to include 'components' of genic selection, and if in the next, logical step they start to rely heavily on the (multi-level) concept of vehicle, then they may find out in the end that they are in fact just duplicating the very pluralistic theory that they thought they were attacking: in that case their enterprise will have 'all the advantages of theft over honest toil'.

This kind of ambiguity can perhaps explain, at least in part, why the thesis of genic selectionism still has appeal among some biologists and philosophers despite much criticism directed against it. Namely, if you are not aware of the ambiguity, it is easy to switch perspectives without noticing it and get a false feeling that your standpoint is invulnerable to objections. So, if you are, for instance, warned about the inability of the austere theory to explain many biological adaptations, you can readily fall back on the more inclusive view that is in effect a thinly disguised pluralistic perspective. If, on the contrary, you are criticized for taking from the hierarchical approach everything except the name, you can shift to the austere theory and then cogently argue that your theory is interesting and bold (some would say, *too* bold).

With these two attitudes changing back and forth, depending on circumstances, it looks as if the multi-level selection perspective is occasionally resisted by a kind of multiple personality: on the one hand, the aggressive and uncompromising genic selectionist, and on the other hand, the mild-tempered and open-minded person who pays lip-service to genic selectionism but who actually accepts much of the hierarchical picture (under a different label, of course). One is out to kill group selection while the other tries to hide his capitulation to the same idea (Dr Kill and Mr Hide, as it were).

## 3 Psychological altruism

According to a standard definition, an act is *psychologically* altruistic if the agent is acting with an ultimate intention to advance the interests of others at the expense of his own interests. In a number of disciplines (e.g. economics, rational choice theory, and until quite recently in psychology) the existence of psychological altruism has been vehemently denied. Moreover, it became usual to think that until a purely self-interested reason was uncovered beneath an apparently altruistic act the behaviour in question remained incompletely understood or even somehow unintelligible. It is exactly this kind of genuine inability to make sense of a truly non-egoistic motivation that was ridiculed in the following Monty Python episode:

> Asked to contribute to the orphans' fund, the banker becomes increasingly puzzled when told this would be neither a loan nor a tax dodge. After hearing that he is simply being asked to donate a pound that will be given to the orphans, the man frowns and shakes his head. 'I don't follow this at all,' he says. 'I mean, I don't want to seem stupid but it looks to me as though I'm a pound down on the whole deal' (Kohn [1990], p. 187).

S. & W. try to break the spell of psychological egoism. They first devote a lot of attention to important conceptual issues (the definitions of altruism and egoism), and then they address the question about the empirical plausibility of genuine altruism.

### 3.1 Conceptual issues

One can define egoism in such a way that it occupies the entire logical space, and the whole debate is thereby short-circuited. Some of the attraction of the thesis of psychological egoism may indeed spring from this kind of conceptual slip. In a number of papers published over the last decade or so Elliott Sober has warned about this and many other conceptual confusions that impede the fruitful discussion about egoism and altruism. *Unto Others* heavily relies on this earlier work but it also contains a lot of new ideas and arguments. I will concentrate here on some claims that I regard as central but less than convincing.

In the current psychological literature altruism is closely linked to the feeling of empathy (e.g. Batson [1991]). Empathy itself is explained as requiring perspective-taking. The idea is, simply, that I can empathize with someone's feeling only if I can imagine myself being in his shoes. S. & W. think, however, that perspective-taking is not an essential component of empathy. Instead, they offer the following definition of that crucial concept: '*S* empathizes with *O*'s experience of emotion *E* if and only if *O* feels *E*, *S* believes that *O* feels *E*, and this causes *S* to feel *E* for *O*' (p. 254). In my opinion, this definition fails to capture what we mean by empathy. For, even after all three conditions of the definiens are met (*O* feels *E*, *S* believes that *O* feels *E*, and this causes *S* to feel *E* for *O*) it is still possible that *S* does *not* empathize with *O*'s experience of *E*. Suppose that, indeed, *S*'s belief that *O* feels *E* causes *S* to feel *E* for *O*, *but that S is completely unaware of that causal connection*. Imagine, for example, that a brain scientist observes S's belief that O feels E, and for some reason *because of this* decides to manipulate S's brain and produce in him a feeling of E for O. The causal connection requirement for empathy stipulated by S. & W. would be satisfied in this scenario, but it seems to me that the connection would be too 'external' for recognizing this as a real case of empathy. For, if S indeed thought that his feeling E for O would still be there even if he did not believe that O felt E, then for S, the matching of the emotions would be purely accidental. I would argue that there is no empathy unless the subject of empathy makes at least some, however weak, subjective connection between the two. But I agree with Sober that my criticism of this definition does not affect the main argument in the second part of the book.

In discussing the meaning of 'egoism' and 'altruism' S. & W. start by distinguishing the four preference structures (pure egoists, pure altruists, E-over-A pluralists, and A-over-E pluralists). According to this classification, pure egoists are moved only by what they expect to advance their own well-being, the well-being of others constituting for them absolutely no reason for action. Pure altruists are their mirror image: in a complete self-abnegation they are driven exclusively by an effort to help others. The E-over-A pluralists and A-over-E pluralists have a more complex preference structure. Each of them places some value on both the well-being of self and of others; they are distinguished by which of the two types of consequences (self-regarding or other-regarding) carries more weight with them. They both value most the outcome with two 'pluses' (for self and other), but the difference between them shows up in what they prefer when faced with the so-called anti-diagonal choice. Having to choose between X ('plus' for other and 'minus' for self) and Y ('minus' for other and 'plus' for self), an E-over-A pluralist will opt for Y, whereas an A-over-E pluralist will opt for X.

In one respect this is an important improvement over the taxonomy from

Sober's earlier work. Namely, he used to label as 'moderate *egoist*' the preference structure 'E-over-A pluralist', and this was very misleading. Now S. & W. realize that the E-over-A pluralist is really an altruist of sorts, and they say explicitly that the existence of E-over-A pluralists is *inconsistent* with the thesis of universal egoism. Although this is definitely the step in the right direction, I would argue that they should have gone even further. In my opinion, there is no good reason to keep the distinction between E-over-A pluralists and A-over-E pluralists, because it is a distinction without a difference. Let me try to show this with a simple example. If in a particular decision situation I choose to play tennis instead of going to do some volunteer work in my daughter's school, it may seem that I am an E-over-A pluralist, because I show that I value 'plus' for self and 'minus' for other as opposed to 'plus' for other and 'minus' for self. But if the situation is changed a little and I am faced with a different choice, I look more like an A-over-E pluralist: If I had to choose, say, between playing tennis and helping the children and teachers in the school during an emergency, the 'plus' for other would loom larger in my mind and it would easily trump the selfish 'plus' of playing tennis. I hope this demonstrates that it is better to postulate only one pluralist preference structure, and that as to which of the two 'pluses' (egoist and altruist) will win over will depend on the circumstances of a particular choice. So, instead of the too elaborate fourfold taxonomy proposed by S. & W., we could follow common sense and keep the tripartite 'natural classification': pure egoists, pure altruists, and those with both motivational components.

## 3.2 Empirical issues

The most important empirical work on egoism and altruism has been done by the psychologist Daniel C. Batson ([1991]). Batson has shown much ingenuity in setting up different experimental scenarios controlling for the presence of different egoistic motivations, and so testing a number of proposed egoistic hypotheses against the empathy-altruism hypothesis. S. & W. discuss Batson's research programme in some detail, but their overall evaluation of that line of investigation is pessimistic: 'Observation and experiment to date have not decided the question, nor is it easy to see how new experiments of the type already deployed will be able to break through the impasse' (p. 272). By virtually dismissing the prospects of this kind of empirical investigation the authors in fact prepare the ground for their valiant effort to defend the existence of altruism on the basis of some very general evolutionary considerations. They say: 'Without an evolutionary perspective, the conflict between the two hypotheses [psychological egoism and its denial] seems unresolvable' (p. 333).

Does it, really? Sober thinks (personal communication) that some of the egoistic hypotheses disproved by Batson's laboratory experiments were not actually very plausible in the first place, and that the theoretical importance of some of his results is for this reason quite limited. This is a valid point, yet some other rival-hypotheses tested and eliminated in the course of Batson's research seem to have had more intellectual respectability, and surely the negative outcome of these experiments could not have been predicted beforehand. Besides, speaking in general terms, S. & W. offer no good argument for their extreme scepticism about the future promise of Batson-type experiments. In stark contrast, they put much faith in their own evolutionary argument developed in the chapter 'The Evolution of Psychological Altruism'. The argument has two steps. They first show, taking the example of parental care, that organisms with an altruist (or semi-altruist) psychological structure would be fitter than pure hedonists because they would have a more reliable mental mechanism for taking care of their offspring. Next, they argue that we have good reasons to believe that this kind of evolutionary improvement was also ancestrally available, on the grounds that the implementation would not require any novel resources (not accessible to the hedonist). In a way, it would be more like building an altruist personality by merely reorganizing the psychological inventory already at the disposal to the egoist.

The argument is an original and interesting contribution to the literature on altruism and evolution. Nevertheless, as S. & W. acknowledge, it is largely a speculation, because it works at the very abstract level of beliefs and desires, with most of the things about underlying physiological reality being presently very poorly understood or even completely unknown. As more empirical details about human motivational and cognitive organization come to light in the future many of our current judgements about reliability and ancestral availability of different psychological mechanisms will probably have to be significantly revised in the light of new evidence. It is somewhat surprising that S. & W. put all their money on such a fragile conjecture that may look to some like a little more than a shot in the dark, while they are at the same time so harshly critical of the budding and conceptually sophisticated research programme in social psychology that has already produced some empirical results.

## 4 Conclusion

*Unto Others* is a book that is unlikely to leave anyone indifferent. The first reactions have shown this already. According to one opinion, the book 'should come with a health warning' because of its possible 'disastrous effects', whereas another reviewer thought that it was 'one of the most important books of the decade'. I believe that the truth in this case is not in the middle but much more in the direction of the latter judgement.

Although *Unto Others* strongly relies on the previously published work of both authors, it expands their thinking in many new directions. It even opens up some novel lines of investigation: for instance, in the section that discusses human groups as adaptive units S. & W. show how a possible unconscious bias in the choice of cultural examples can be removed by randomly selecting the data from the ethnographic record HRAF (Human Relations Area File). Spanning three fields (philosophy, biology and psychology), the book addresses an impressive number of issues, and the opinions defended are thoughtful, well argued, often challenging and always very clearly expressed. Even those who continue to strongly disagree with Sober and Wilson will have a lot to learn from them.

## Acknowledgements

I would like to thank Elliott Sober and David Sloan Wilson for a very useful email discussion that significantly improved my understanding of their views. I am also grateful to Michael Levin for his comments on the first draft.

*Department of Philosophy*
*King's College London*
*Strand*
*London WC2R 2LS, UK*
*E-mail: neven.sesardic@kcl.ac.uk*

## References

Batson, D. C. [1991]: *The Altruism Question: Toward a Social-Psychological Answer*, Hillsdale, NJ: Lawrence Erlbaum.

Brandon, R. N. [1990]: *Adaptation and Environment*, Princeton: Princeton University Press.

Dawkins, R. [1981]: 'In Defense of Selfish Genes', *Philosophy*, **56**, pp. 556–73.

Dawkins, R. [1994]: 'Burying the Vehicle', *Behavioral and Brain Sciences*, **17**, pp. 616–17.

Kohn, A. [1990]: *The Brighter Side of Human Nature*, New York: Basic Books.

Mackie, J. L. [1978]: 'Law of the Jungle', *Philosophy*, **53**, pp. 455–61.

Sterelny, K. and Kitcher, P. [1988]: 'The Return of the Gene', *Journal of Philosophy*, **85**, pp. 339–61.