What is the problem of evolutionary altruism?

Abstract: This paper addresses the "problem of altruism" within evolutionary theory. An alternative conceptualisation is presented in order to show that standard presentations of the matter are slightly off the mark in some respects: They draw the line between evolutionary models in the wrong place, they produce odd anachronisms by presenting the matter from within a superseded position, and they promote theoretically ungrounded verdicts to the effect that altruism is merely apparent. It is suggested that the discourse suffers from "terminologically driven" distorting pre-theoretical influence, and that an alternative framework therefore might be useful.

1. Introduction.

In evolutionary biology we encounter what Wilson and Wilson (2007) call "the fundamental problem of social life", which problem involves dealing with traits that seem pointless or even outright costly for the individuals that display them, and useful only some neighbouring individuals. This problem area is generally addressed in terms of *evolutionary altruism*. The target of this paper is this well-known conceptual practice of discussing the evolution of social life in terms of altruism and selfishness. This is a firmly established practice but also one that exhibits individual variation in the application of the terms. That is, there is no consensus concerning which traits qualify as altruistic or selfish in the evolutionary sense. A given explanation may be understood, by some, as accounting for the evolution of altruism and, by others, as explaining away the appearance of altruism.

In this paper I will present a plausible account of what the theoretical point of introducing the notion of evolutionary altruism was. I will argue that this rationalisation continues to make good sense. Of course, this is consistent with there being alternative, at least equally well-motivated proposals regarding the use of 'altruism' and 'selfishness' in evolutionary contexts. I will argue, however, that there is as of yet no theoretically relevant motivation for understanding the issue of evolutionary altruism differently. As a consequence, commonly occurring proposals to the effect that altruism is explained away are theoretically misguided. My conclusion in this respect is not original. So for instance, my view is in conclusion just what Sober and Wilson (1998) are suggesting. However, there are problems with the way these authors argue their point as they do so in terms of group selection, which is a controversial subject, empirically and conceptually. The demarcation of evolutionary models that they endorse can be argued for without taking a stand on the question of group selection.

The key substantial point of the paper is that there are aspects of the altruism-framework as commonly presented that lack theoretical underpinnings. I suggest an alternative conceptualisation to avoid these shortcomings. To this I add, more speculatively, an account of why the ungrounded proposals are nevertheless intuitive. This account appeals to the pre-theoretical "loads" of 'altruism' and 'selfishness' as explanatorily relevant. That there is a dubious pre-theoretical influence on the debate has been suggested repeatedly. So, for instance, David Sloan Wilson once wrote (1983: 184):

> Many humans have a strong philosophical bias towards seeing purpose and order above the level of the individual. Others have an equally strong bias towards ascribing selfish motives to all behaviors. Not surprisingly, both of these dispositions are represented in evolutionary thinking and both have been stated axiomatically. The process of converting them into legitimate scientific hypotheses about evolution has not been smooth.

This perceived lack of smoothness has been reported repeatedly (Midgley 1983, Kitcher 1985, Sober 1988, Kerr et al. 2004, Stich 2007, West et al. 2007). It is noteworthy that Richard Dawkins, in his preface to the 30th anniversary edition of *The Selfish Gene*, admits to having strayed against his very own stipulation regarding 'selfish' (Dawkins 2006/1976 *ix*).

In the next section (2) I will present what I take to be a very plausible account of the rationale for introducing the term 'altruism' in evolutionary discourse. In section 3 it is argued that 'altruism' is frequently defined in a way that draws the line between evolutionary models in the wrong place, and/or that the categorisation of models is incomplete. Section 4 argues that common definitions/presentations of the problem must be understood as tacitly counterfactual, and that an odd anachronism results as this is hidden from view. In section 5 I argue that the option, taken by some, to define altruism out of existence is misconceived in light of the only reasonable account of the point of introducing the notion we have.

I will now turn to suggesting a vantage point from which to address the prevailing altruism/selfishness conceptualisation.

2. Default model – Assortativity model

Traits like having sharp teeth, being faster, or being less conspicuous could rather easily be conceived of as advantageous from the point of view of a sole mutant. In a great deal of cases, evolutionary hypotheses engage with traits such that we would expect a single mutant with

the trait to be fitter. By "*the default model*" I will understand the intuitive evolutionary model that is capable of dealing with such traits. As an informal negative characterisation we may say that the default model deals with selection scenarios such that the superior fitness and predicted spread of the focal trait does *not* depend on its providing benefits to others. I intend this characterisation to capture a kind of case that early evolutionists intuitively considered as being theoretically clear, one where the accumulation of useful properties basically could be conceived of as the result of the occasional appearance of superior mutants exhibiting traits that are beneficial to the mutants themselves. I think the default model, understood in this way, is an appropriate starting point for seeing how the fundamental problem of social life arises within biology. The thing is that not all established traits yield to this easily accessible model. Social life involves traits such that being beneficial to others is crucial for becoming established. One important feature of models that have been developed to deal with such traits is what we may call *assortativity*. The main idea is that if a population consists of individuals that differ as regards a social trait T, one that benefits others, selection for T requires that those with T (or at least with the relevant gene) benefit to a greater extent than those without T (or without the relevant gene). Even if T is costly for the agent it may spread if it predominantly distributes, to a sufficient degree, its benefits to others with T. When it comes to traits such as speed and sharp teeth, on the other hand, no demands are made for considering benefits to others, and assortativity is thus not an issue.[1]

The basic idea of assortativity, if not the term (in this context, that is), has clearly been around for very long. Darwin himself hinted at it in an oft-cited passage on human morality in *The Descent of Man* (1871: 166). R. A. Fisher (1930), J. B. S. Haldane (1932), and Sewall Wright (1945) made proposals relevant for the topic. Still, it seems fair to say that the idea wasn't developed in much detail until the contributions by Williams and Williams (1957) and, in particular, William Hamilton (1963, 1964a, b). The same basic consideration, or at least closely related considerations, can be framed in various ways; the issue may be expressed in terms of group-level benefits (Wilson 1980, Sober and Wilson 1998), correlated interactions (Hamilton 1975, Sober 1992, Okasha 2005), indirect effects on fitness (Lehmann and Keller 2006), and population-structured evolution (Sterelny 1996, Kerr and Godfrey-Smith 2002). This is not claiming that all these different modes of expression are equivalent. The claim is rather that they are all mainly motivated by the need to come to grips with the fundamental

[1] Strictly speaking, I assume that there may be an assortativity component to many "default traits" as well. Given parental care, offspring will benefit not only from their own sharp teeth but from those of their parents as well.

problem of social life, which problem I suggest can be conveniently stated as the problem of modelling assortativity.

Although I have provided only an informal characterisation of the key terms 'default model' and ' assortativity model', I think there is consensus enough about the extensions. I believe, for instance, that theorists who disagree about whether Hamilton's kin selection model (Hamilton 1964a,b) involves altruism and/or group selection nevertheless agree that it essentially involves taking into account benefits to others, and thus assortativity. Although the terms 'default' and ' assortativity ' are not commonly used in the context, there seems to be wide agreement, at least implicitly, that there is an important distinction to be made between models that need assortativity (that is, where benefiting others of the right kind is essential) and models that don't (Dawkins 2006/1976, Futuyma 1998, Sober and Wilson 1998, Michod 1999, Bourke 2001). However, this consensus is much obscured by the use of 'individual selection', or 'genic selection', to cover both default and (some) assortativity models, which terminological practice is controversial (e.g. Sober and Wilson 1998, esp. ch. 2). In using 'assortativity' I hope to steer clear of this conceptual controversy.

A model that demands assortativity is significantly different than one that makes no such demands. Evolutionists started out with a firm enough grasp of cases that made no demands concerning assortativity. However, it was clear ever since Darwin that if the default model was all there was to evolutionary theory the theory would be seriously incomplete. There was a firm conviction that the world presents us with cases that don't yield to this model. Given this, the so-called problem of altruism can be viewed as a problem that stems from the insufficiency of the default model. I will argue that there is no other point of evolutionary significance to me made concerning the use of 'altruism'.[2] On this understanding, any trait that requires an assortativity model counts as altruistic. From this vantage point, some shortcomings related to extant presentations of the problem area can be readily seen and addressed. This will be the topic of the next three sections.

3. How not to draw the line between evolutionary models

The by far most common informal definition takes 'evolutionary altruism' to denote traits that benefit, fitness-wise, others at the expense of the focal individual. So, for instance, E. O. Wilson defines it as "self-destructive behavior performed for the benefit of others" (1975:

[2] One may of course want to further subdivide the class of assortativity models. However, the problem faced by evolutionists, and which called out for a name, was simply that the default model was insufficient for many cases.

578). Samir Okasha writes: "In evolutionary biology, an organism is said to behave altruistically when its behaviour benefits other organisms, at a cost to itself." (2008). A more general definition, without mention of behaviour, is provided by Mark Ridley who describes altruism as "the transfer of some benefit from the altruist to the recipient, at a cost to the altruist." (1995: 234).

A shortcoming of these definitions as they stand, one that has been addressed by Peter Gildenhuys (2003), is that they allow rather pointless verdicts. For instance, it seems quite unnecessary to speak of altruism whenever a variety does a bad job at competing for resources by having, say, bad memory or reduced sex drive, although the resulting "performances" will be beneficial to others. A special case is when circumstances have changed such that a default model trait has come to fit the description. So, for instance, it is likely that there are cases where human hunting amounts to selection against the most aggressive variety. This variety has, we assume, become established in accordance with the default model but is nevertheless, in current surroundings, benefiting others by offering itself as an easy target and thus seemingly qualifies as altruistic.

We get rid of this abundance of altruism by holding that there is no point to applying 'altruism' to whatever traits confer benefits to others at a cost to the focal individuals - which perhaps any less fit trait can be claimed to do - but only to traits that seem to have spread in spite of being resistant to the default model.

A feature of common definitions of altruism, as the ones cited above, that deviates from the default model/assortativity model framework that I am proposing is the cost criterion. We saw that Wilson defines 'altruism' as "self-destructive behavior performed for the benefit of others" (1975: 578), and both Okasha and Ridley take altruism to involve costs. Consequently, on this understanding cases with no costs to focal individuals but that are nevertheless as problematic from the point of view of the default model will be excluded from the problem area. The default model clearly faces problems with traits that are seemingly useful and widespread but of no particular benefit to the actor, even if not costly. We may appeal to R. A. Fisher's example involving distasteful larvae to illustrate this. The case was considered a problem because it was assumed that a bad-tasting individual is likely to die whether it is spat out or not; the trait appeared pointless from the point of view of individual

survival (Fisher, 1999/1930: 158-159, Hamilton 1964b: 19).[3] The seeming usefulness of the trait, as compared to the alternative, suggested that its persistence is not a matter of drift. However, there is no indication that Fisher assumed the trait to be costly, and Hamilton's presentation of the problem is straightforward in this respect (1964b: 19): "That this phenomenon presents a difficulty, namely an apparent absence of positive selection, is obvious as soon as we reject the pseudo-explanations based on the "benefit to the species"...." He clearly thinks that there is a problem whether there is a cost or not. Now, I am not perplexed by the assumption that substantial changes in tastefulness generally come at a cost, but whatever we think about this empirical matter we should acknowledge that the case at hand is problematic whether or not distastefulness is costly. So, the cost criterion that is commonly associated with the notion of altruism is not an essential ingredient to have the relevant kind of theoretical trouble. If you have a model capable of dealing with cases involving a mere lack of benefit, without costs, quantitative changes will suffice to deal with cases with costs.

Behaviour that is of no particular advantage to the actor needs to benefit the "right" kind of others more than the "wrong" kind of others in order to spread selection-wise. Thus, it requires assortativity. And, again, once benefits to others are brought into the picture, accounting for costs rather than mere lack of advantage requires no more than boosting those benefits. From a modelling perspective, then, cases involving "self-sacrifice" do not give rise to a theoretical puzzle of their own.

Nevertheless, it appears to be a persistent desideratum for proposals in the domain that they accommodate the idea that altruism is costly. When theorists furthermore pair altruism with a notion of *cooperation* that is taken to involve mutual benefit the upshot is conceptual schemes with unmotivated gaps. For instance, Lehmann and Keller propose a framework, presented as a general one, for dealing with altruism and cooperation in terms of direct and indirect fitness. They claim (2006: 1367):

> [W]e categorize as cooperation those cases where the act of helping is associated with an increase in the FI's [focal individual's] direct fitness ... and as altruism cases where helping is associated with a decrease in the FI's direct fitness...

[3] Fisher mentions having been informed by a professor Poulton that some distasteful organisms are actually robust enough to survive being tasted. However, Fisher rests confident that a non-classical alternative "...will certainly be effective in a usefully large class of cases." (1999/1930: 159).

This then conforms to the understanding of cooperation as mutual benefit and altruism as costly. However, in the context of evolutionary modelling it comes with the unmotivated non-categorisation of cases where there is neither cost nor benefit in direct fitness. In a similar manner, West et al. (2007) work with a scheme in which behaviour that benefits others may be either (+/+) or (–/+), with the direct effect on the actor on the left hand side of the dash and the effect on neighbours on the right. Thus, cases where there is neither direct cost nor direct benefit to the actor, the (0/+)-cases, are left out of the conceptual scheme. These approaches thus cannot be considered complete with regard to the modelling requirements set by the evolution of sociality.

It is reasonable to hold that models should be distinguished primarily on the basis of the parameters they need to include, not on the basis of the exact quantities of those parameters. The difference between benefiting others at a minor cost to oneself and doing it without cost to oneself is presumably a very slight difference on a continuous parameter, and it shouldn't be overwhelmingly important from the point of view of evolutionary conceptualisation. The true borders of the fundamental problem of social life should be conveyed clearly by our conceptual scheme, but many proposals that adopt the altruism framework fall somewhat short of this desideratum.

It seems to me that from the default model/assortativity model perspective one would have no inclination to make much conceptually of the difference between, on the one hand, benefiting others at a (slight) cost and, on the other, benefiting others at no cost (and no benefit to oneself).

It may be that the aforementioned shortcomings of the altruism-framework, as commonly presented, are "terminologically driven". That is, assuming that 'altruism' in the vernacular sense is typically associated with costs, rather than a mere lack of benefit,[4] the cost requirement may then transfer to influence the evolutionary discussion simply in virtue of the fact that the term to be used ('altruism') brings this criterion with it. In any case, the cost criterion is not called for from the point of view of delineating evolutionary models.

4. Hidden counterfactuals

Another noteworthy feature of standard informal definitions of altruism is that it is unclear how it could hold of any trait that actually manages to evolve. Again, these definitions tell us

[4] Not that this is entirely obvious. Intuitions about the exact reading of 'altruism' in the vernacular sense may vary. My hunch is that it is frequently associated with costliness, however.

that altruistic traits increase the fitness of others while inducing a cost to the focal individual. The definition then presumably states that the agent would be fitter not being altruistic. If this net loss is characteristic of the trait and thus true on average, then altruism shouldn't increase in frequency on any model. Thus, common definitions may seem to suggest that evolutionary altruism is bound to be non-existent as far as established traits are concerned.

Things look considerably different if we interpret the definition not as dealing with what characterises so-called altruistic traits when these are selectively advantageous, but with what would be the case under the constraints of the default model. So, for instance, if we were to assess trait fitness by averaging across simulations involving sole mutants, so-called altruistic traits would come out looking unpromising. Thus, the definition makes sense if understood as tacitly counterfactual in this respect.[5] If it isn't so understood, then 'altruism' would seem to denote traits that won't evolve on any model, and it is difficult to see the point of using 'altruism' for that purpose, especially since we have an alternative interpretation that makes sense of the need for the term and which is considerably less redundant.

As the counterfactual character is commonly hidden from view in introductions to the problem area we get misleadingly anachronistic presentations. Here are four examples to illustrate this claim:

Sober and Wilson give the following account of why so-called altruistic behaviour seems hard to explain (1998: 18-19, emphasis is original):

> After all, natural selection evolves traits that cause individuals to have *more* offspring than their competitors, not fewer. There is a selective advantage in being selfish, just as there is a selective advantage in having strong teeth and keen eyesight.

Douglas Futuyma begins a section on social interactions by stating (1998: 594, emphasis is original):

> Because natural selection is based on *individual* advantage, "selfish" traits should increase in frequency if they are heritable. Thus cooperative interactions in which individuals apparently dispense benefits to others, often at a cost to themselves, seem antithetical to evolution by natural selection.

[5] Others would prefer other ways of spelling out the counterfactual: as what would happen without group selection (Wilson 1980, Sober and Wilson 1998), without correlated interactions (Hamilton 1975, Sober 1992, Okasha 2005), without indirect effects on fitness (Lehmann and Keller 2006), or without population-structured evolution (Sterelny 1996, Kerr and Godfrey-Smith 2002).

Kim Sterelny and Paul Griffiths write (1999: 153):

> Altruism is a puzzle. Imagine, for example, that you are a male vervet monkey in a tree, and you notice an eagle. Do you give an alarm call, warning all the monkeys around you, or do you quietly hide? Selection should favor quite hiding. (…) Over time, we would expect selection to weed out the trait of warning others about predators, as well as signalling the presence of food, contributing to collective defence (as defenders would lose out to cowards and skulkers), reproductive restraint, and caring for others´young.

In a similar vein, Samir Okasha writes in an encyclopaedic entry on evolutionary altruism (2008, emphasis is original):

> Natural selection leads us to expect animals to behave in ways that increase their *own* chances of survival and reproduction, not those of others. But by behaving altruistically an animal reduces its own fitness, so should be at a selective disadvantage vis-à-vis one which behaves selfishly.

We are told that so-called altruistic behaviour is not to be expected given the "nature" of natural selection. Are these authors thereby stating how things really seem to them? No, to the contrary; they think that there are conditions under which the "unexpected" kind of trait can spread. But then, why are they claiming that such traits seem "antithetical" to natural selection or that they "should be at a selective disadvantage"? Because, I suggest, they are not stating what they in fact expect given what they know, but what they would have expected given the default model of selection. What they say would have been true had the default model been all there was by way of evolutionary modelling. Still, it is by no means clear why they choose to pretend to be confined to the default model. Consider the following: Physicists, or their subject's historians and philosophers, sometimes need to talk about things that were anomalies on the Newtonian theory but are accommodated by current physics. I assume that we expect them to express this explicitly. That is, they should say things like "On a Newtonian view we would predict that..." rather than "Physics predicts that....". Newton's physics is super-seeded, and it would be distinctively odd to speak as if it still reigned in the domains where it is currently taken to give the wrong answers. But this oddity is exactly what the cited

presentations of the problem of altruism exhibit. The view that the default model exhausts the theory of natural selection is no less superseded than Newton's physics. Given this, there is no point today in speaking from within the constraints of the clearly insufficient default model when stating the problem. It is true that we wouldn't expect to see so-called altruistic traits if we thought that all selection processes conform to the default model, but we have no reason to present that counterfactual condition as actual.

Now, perhaps it may be argued that the oddity involved is a consequence of a benevolent pedagogically motivated decision to level with the average readers presumed assumptions. Perhaps the average reader thinks of evolution from within the default model, and authors merely want to frame the problem from this imaginary readers point of view, not their own. If so, this pedagogical device has become quite popular. I very much doubt, however, that there is anything pedagogically virtuous about this approach. I would think it better to state the problem by first presenting the default model and then adding that not all traits could be accounted for that way.

Another possibility is that this way of presenting the problem derives from the pre-theoretical context of the terms 'selfishness' and 'altruism'. Psychological selfishness is frequently held to be obvious enough and so the pressing question is: are people ever truly altruistic? Whatever our beliefs about vernacular altruism we are all well acquainted with proposals to the effect that there are mostly, or always, hidden selfish motives to be uncovered under seemingly altruistic feats. That is, we are used to viewing altruism as carrying the burden of proof. It seems that the presentations above are offering a presentation in this very vein. However, from my proposed perspective this pattern doesn't make much sense in the current evolutionary context. Some assortativity models are widely accepted, and have been so for some time, and it is quite misleading to suggest that the idea that some traits have evolved in accordance with such models carries the onus of proof.

5. Explaining altruism away

If we understand 'evolutionary altruism' to denote widespread traits with an air of utility about them but otherwise such that the default model was deemed incapable of accounting for them, then this gives us a clear criterion for when we can justifiably claim to have explained away the appearance of altruism. The only way it can be shown, on this view, that a trait isn't altruistic is to show that it yields to a default model, initial appearances notwithstanding. Sterelny and Griffiths (1999, 154) provide an illustration of this situation:

> [R]avens that give loud yells when they find large carcasses turn out to be acting altruistically after all. These ravens are young birds with no territories of their own. Though their calling recruits others with whom they must share their food bonanza, if they did not call, they would be expelled by the territory owners. Recruiting other ravens swamps the territory owners' defenses (Heinrich 1990).

The benefit involved is apparently such that a single mutant raven of this kind could be fitter than non-calling conspecifics. Thus, the default model is sufficient. However, Sterelny and Griffiths do not rely on the default model/assortativity model distinction in drawing the line between altruism and non-altruism. This is evident as they also claim that so-called reciprocal altruism is really a matter of explaining away altruism (154-155), which verdict is not uncommon. This is not the result delivered by the default model/assortativity model approach. The blood sharing practice of vampire bats is an oft-cited example of reciprocal altruism. Successful hunters will share blood with unsuccessful hunters, but individual bats will take turns being donors and receivers. A single mutant blood-sharing bat would be good for some neighbours, but in a population of non-sharers it would not make for superior fitness of the donor. We need a model that incorporates, as an essential ingredient, the presence of other similarly inclined individuals. In cases that yield to the default model, on the other hand, we don't need to bother about benefits to similar others, whether accounting for sharp teeth, sophisticated camouflage or resistance to parasites. So, it seems that we can hold reciprocal altruism to be altruistic in virtue of the distinction between default model and assortativity model. If not, what is at stake theoretically?

Consider also the not uncommon inclination to assume that Hamilton's kin selection model amounts to showing that traits that at first appeared altruistic turned out to be nothing of the kind (e.g. Futuyma 1998: 596). Hamilton's model essentially involves the fitness benefit to a recipient, weighted by degree of relatedness between donor and recipient[6], and so it is clearly an assortativity model. Then we must ask: What is at stake in these denials of altruism?

A person who employs the default model/assortativity model approach to guide the application of 'altruism' is not thereby misunderstanding the models that are under discussion. That is, it is not as if one cannot understand reciprocal altruism models or kin selection models if one takes them to model altruism rather than showing that altruism is merely apparent. The complaint that we are not really dealing with altruism in these cases needs to be

[6] Which is, in the context, taken to represent the probability of sharing the relevant gene.

underwritten by something of theoretical value. If the verdict is, implicitly, that he distinction between default models and assortativity models isn't significant enough to motivate the use of 'altruism' about traits that require the latter, then what is significant enough?

It can hardly be more fruitful to use 'altruism' only about traits that cannot evolve on any model. However, this is apparently sometimes the interpretation of 'altruism' that is favoured. The verdict that kin selection explains away the appearance of altruism is commonly reached by appeal to a selfish gene perspective on evolution, a perspective that introduces a switch from tokens to types the reason for which is quite elusive. At the organism level token individuals forfeit fitness benefits to other tokens. Although this relation between tokens is no different at the genetic level the case is held to be different, since now tokens don't count. An early example of this is Hamilton (1963) in which the author claims that "...the ultimate criterion which determines whether [a gene for altruism] will spread is not whether the behavior is to the benefit of the behaver but whether it is to the benefit of the gene ..." (Hamilton 1963: 354). The token cost for the behaver is here contrasted with the type benefit for the gene. For some (e.g. Dawkins 1976, Futuyma 1999), the switch from token organisms to gene types means that in order for the gene (qua type) to be altruistic it would need to benefit an allelic competitor at its own expense. It is hard to see any point to having 'altruism' indicate a trait that won't evolve on any model. It is clear that the value of Hamilton's model is entirely unaffected by our decisions regarding the use of 'altruism' and 'selfishness' in evolutionary contexts. As far as historical priority is concerned we have no good reason to doubt that the notion of altruism has made its way into evolutionary discourse due to problems, as seen from the default model, stemming from observed interactions between token organisms. The term was almost certainly *not* suggested to indicate the belief that the traits in question simply couldn't have evolved at all. Given this understanding, if it becomes evident that an assortativity model is required to capture the evolution of some social trait then the assumptions that motivated talk of altruism are, if anything, vindicated. Using 'altruism' in this latter manner isn't a sign of getting anything of evolutionary interest wrong.

Employing the default/assortativity framework makes it clear that there is no point to performing a switch from organism tokens to gene types only then to claim that altruism has been found to be merely apparent. If the perceived problem is that an established trait cannot be accommodated on the default model, then the only way this could be an illusion is that we find out that, contrary to appearances, the trait does yield to the default model after all. But this is not what anyone takes the situation to be like in most cases where altruism is claimed to be merely apparent (e.g. reciprocal altruism and kin selection).

Now, there may be another prima facie reason for resistance to applying 'altruism' to all assortativity traits, one having to do with how the debate about altruism has become linked to the debate about group selection. It seems clear, for instance, that Sober and Wilson (1998) assume both 1) that the key distinction to be made is the one suggested here, the one between the default model and assortativity models and 2) that all assortativity traits a) count as altruistic and b) require group selection. I agree with 1 and 2a but remain uncommitted as regards 2b. It could be, then, that if theorists think that altruism and group selection come in a package, and if they do not think that kin selection qualifies as group selection, then they will thereby deny that it involves altruism.

Regarding the relation between altruism and group selection the following is worth pointing out. On a very reasonable understanding, 'altruism' was useful in picking out a problem category ("the problem of altruism"). As should be clear by now, I take the problem to best be characterised negatively: the problem emanating from the insufficiency of the default model. Traits qualify as altruistic merely in virtue of fitting this negative characterisation. Group selection (on whichever interpretation) is, or may be, a purported partial solution to the problem. There is no clear reason why in designating a trait as pertaining to the problem category we should bother about what might be the solution. That is, it seems quite uncalled for to insist that 'altruism' be conceptually linked to 'group selection', such that we might have to re-categorise the trait as non-altruistic depending on how we decide to label solutions.

In the context of debating group selection 'altruism' has frequently been taken to mean a trait that is less fit within groups (e.g. Sober and Wilson 1998). This interpretation looms large in "subversion from within"-arguments. Such arguments are raised concerning group selection models and claim that although benefits may accrue to groups with many altruists, within such groups selfish individuals are fitter than altruistic ones. When, in this context, it is said that selfish traits will subvert groups with many altruists this is not, as might first appear, much of an empirical generalisation. It is a probabilistically certain result given that 'selfish' by definition denotes a trait that is fitter within groups. Now, how then might altruism persist, and perhaps even reach fixation? Well, one proposal (e.g. Sober and Wilson 1998) is that if altruism has been around for some time there may have evolved defences against subversion; perhaps altruists have become extremely selective and apt at detecting and mobbing freeloaders. However, we now face a conceptual problem. In the context of the subversion issue, 'altruism' is understood as signifying being less fit within groups. However, the defence argument seems to conclude that the so-called altruists are no longer threatened. That is, they

are no longer less fit within groups. Consequently, the trait is selfish after all, not altruistic. So, what was presented as a manner in which altruism might be protected from subversion turns out to be nothing of the kind; the trait is no longer altruistic on this understanding. There is something very unhelpful about this way of using 'selfishness' and 'altruism'. The trait is not usefully reclassified from altruistic to selfish merely because it turns out not to be less fit within groups. It should rather continue be called altruistic in virtue of the fact, assuming it is one, that the evolution of the trait requires an assortativity model. The reasonable formulation of the subversion issue is asking under which conditions traits that require an assortativity model to spread can become stable, and perhaps even fixed. It is trivial that altruistic traits will be forever challenged if 'altruism' is by definition indicating being less fit within groups. It is not trivial, and perhaps not true at all, however, that assortativity traits are invariably facing serious challenge. An assortativity trait doesn't end being such a trait merely because potentially challenging mutants are invariably less fit due to, say, the aggressive selectiveness of the assortativity trait. It is still true, in such a case, that had we reversed the situation, letting the assortativity trait appear as a single mutant in a population of non-donors, the assortativity trait would be less fit.

So, to the extent that the explaining away of altruism relies on the idea that altruism is, by definition, less fit within groups, it relies on a very unhelpful usage of the term 'altruism'. In this context it is noteworthy that Sober and Wilson run into conceptual problems as they attempt to address the issue about how altruism might defeat selfishness. They write (1998, 135):

> Imagine a very large population of unrelated individuals who vary in their altruistic tendencies. Suppose that each individual's degree of altruism is observable and that membership in a social group requires the consent of all parties. Each individual wants to associate with the most altruistic partners it can find. Thus, if the groups are of size *n*, the *n* most altruistic individuals in the population will form one group, the *n* next most altruistic individuals will form another group, and so on down to the *n* least altruistic individuals, who associate not by choice but by default (alternatively, the least altruistic individuals could choose to remain solitary...)

It is evident that the so-called altruists in the example are a valuable resource. Now, when it comes to valuable resources in general, such as food, shelter, or mates, it is not the case that one is less self-sacrificing as regards fitness if one, whether by choice or default,

systematically forfeits such benefits to others. So, it remains to be explained why the individuals who systematically fail to reap the gains of generous company, or even choose a solitary life, would be the least altruistic ones. If we were discussing motivational profiles then the individuals who Sober and Wilson hold to be the most selfish ones would presumably earn that title if they cared only about themselves (disregarding the fact that they are not very successful at securing resources). However, if the issue is evolutionary matters then it concerns fitness gains, not motivation, and Sober and Wilson do not state in what sense those that consistently fail to exploit the valuable company of the "altruists" could count as selfish. The designation is not in line with their commitment to the idea that being selfish in the evolutionary sense is a matter of having a fitness advantage over altruists that altruism needs to counter on a group level. In their stated example so-called selfish individuals are not favoured at any "level" of fitness ascription. I take it that the trouble could be avoided by speaking directly in terms of how traits that require an assortativity model may persist and even become fixed.

6. Concluding Remarks.

I have argued that the prevalent manner of presenting what Wilson and Wilson (2007) have termed "the fundamental problem of social life" in terms of altruism and selfishness is associated with a number of problems. Although this is quite widely acknowledged I think some shortcomings have not been given sufficient attention. The basic problem of so-called evolutionary altruism can very reasonably be seen as one stemming from the insufficiency of models that lack resources to capture the role of benefits to others. That is, the problem should be seen as one of the need for assortativity models in addition to the default model. Given this understanding the temptation to have altruism come out as a mere appearance is misconceived, as the application of assortativity models clearly do not show the insufficiency of the default framework to be illusory. Also, standard presentations of the problem area unnecessarily invite confusion by failing to make the counterfactual nature of the problem explicit. It is true that we wouldn't expect so-called altruism if we were confined to the default view of selection, but there is hardly any point in presenting that as our current predicament. Furthermore, many proposals utilising the targeted framework give us incomplete categorisation schemes as far as the problem of social life is concerned.

One way to develop an alternative scheme, one that is no doubt already inherent in extant approaches, terminology aside, is simply to use 'assortativity trait' of traits that are held to demand some assortativity model. Such a label can hopefully be used to sidestep most of the

terminological issues there are surrounding notions like 'altruism' and 'cooperation' (see Kerr et al. 2004, and West et al. 2007), and 'individual selection' and 'group selection' (see Sober and Wilson 1998, and Okasha 2001). Also, it may be convenient in that, in dealing with traits in the field, biologists are likely to be more confident that a particular trait is an assortativity trait than they are about the specific cost-benefit pattern (e.g. whether it is (-/+), (0/+), or (+/+), to adapt the scheme presented by West et al. (2007)). The category can clearly be partitioned with respect to such patterns in a variety of manners without giving a misleading image of the fundamental problem of social life.

Another advantage of such a scheme is that it would serve to eliminate what there is of disturbing pre-theoretical "waste" surrounding the notions altruism and selfishness in evolutionary contexts. As mentioned in the introduction, quite a few authors have indicated that there are illicit transfers between domains in the context of evolutionary altruism and selfishness. It is at least a plausible hypothesis that, frequently, the urge to explain altruism away in evolutionary contexts reflects a pre-theoretical bias bearing no relation to issues of evolutionary significance. It should be borne in mind in this context that, given the inter-disciplinary reach of evolutionary biology, the concepts we wish to promote need to work reasonably well across a wide range, including most, or even all, of the social sciences. It seems to me that that the altruism/selfishness framework has performed quite badly in this respect.

Even if my terminological proposal is rejected, and so that the altruism framework will continue to be preferred, I have nevertheless indicated ways in which discourse utilising that framework can, and should, be improved.

Literature Cited:

Bourke, A. F. G. 2001. Social Insects and Selfish Genes. Biologist 48 (5): 205-208.

Darwin, C. 1871. The Descent of Man and Selection in Relation to Sex. London: Murray.

Dawkins, R. 2006/1976. The Selfish Gene: The 30th Anniversary Edition. Oxford and New York: Oxford University Press.

Dugatkin, L. A. and H. K. Reeve. 1994. Behavioral Ecology and Levels of Selection: Dissolving the Group Selection Controversy. Advances in the Study of Behavior, 23: 101-133.

Fisher, R.A. 1999/1930. The Genetical Theory of Natural Selection. Oxford and New York: Oxford University Press.

Futuyma, D. 1998. Evolutionary Biology. Sunderland, MA: Sinauer.

Gildenhuys, P. 2003. The Evolution of Altruism: The Sober/Wilson Model. Philosophy of Science 70: 27-48.

Haldane, J. B. S. 1932. The Causes of Evolution. London: Longman.

Hamilton, W. 1963. The Evolution of Altruistic Behaviour. American Naturalist 97: 354-356.

Hamilton, W. 1964a. The Genetical Evolution of Social behaviour. I. Journal of Theoretical Biology 7: 1-16.

Hamilton, W. 1964b. The Genetical Evolution of Social behaviour. II. Journal of Theoretical Biology 7: 17-52.

Hamilton, W. 1975. Innate Social Aptitudes of Man: An Approach from Evolutionary Genetics. In Biosocial Anthropology. Ed. by R. Fox. New York: Wiley and Sons.

Kerr, B. and P. Godfrey-Smith. 2002. Individualist and Multi-level Perspectives on Selection in Structured Populations. Biology and Philosophy 17: 477-517.

Kerr, B. et al. 2004. What is Altruism? Trends in Ecology and Evolution 19 (3): 135-140.

Kitcher, P. 1985. Vaulting Ambition. Cambridge, MA: MIT Press.

Lehmann, L. and L. Keller. 2006. The Evolution of Cooperation and Altruism – a General framework and a Classification of Models. Journal of Evolutionary Biology 19: 1365-1725.

Maynard Smith, J. 1998. Commentary on Kerr and Godfrey-Smith. Biology and Philosophy 17: 523-527.

Michod, R. E. 1999. Darwinian Dynamics. Princeton: Princeton University Press.

Midgley, M. 1983. Selfish Genes and Social Darwinism. Philosophy 58: 365-377.

Okasha, S. 2001. Why Won't the Group Selection Controversy Go Away? British Journal for the Philosophy of Science 52: 25-50.

Okasha, S. 2005. Altruism, Group Selection and Correlated Interaction. British Journal for the Philosophy of Science 56: 703-725.

Okasha, S. Biological Altruism. Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/entries/altruism-biological. Last updated on 28 October 2008.

Ridley, M. 1995. Animal Behavior. Malden, MA, Oxford, Melbourne and Berlin: Blackwell.

Sober, E. 1988. What is Evolutionary Altruism? Canadian Journal of Philosophy 14 (suppl.): 75-99.

Sober, E. 1992. The Evolution of Altruism: Correlation, Cost and Benefit. Biology and Philosophy 7: 177-187.

Sober E. and D. S. Wilson. 1998. Unto Others. Cambridge, MA, and London: Harvard University Press.

Sterelny, K. 1996. The Return of the Group. Philosophy of Science 63: 562-584.

Sterelny K. and P. Griffiths. 1999. Sex and Death: An Introduction to Philosophy of Biology. Chicago and London: University of Chicago Press.

Stich, S. 2007. Evolution, Altruism and Cognitive Architecture: A Critique of Sober and Wilson's Argument for Psychologichal Altruism. Biology and Philosophy 22 (2): 267-281.

West, et al. 2007. Social Semantics: Altruism, Cooperation, Mutualism, Strong Reciprocity and Group Selection, Journal of Evolutionary Biology 20: 415-432.

Williams, G. C. and D. C. Williams. 1957. Natural Selection of Individually Harmful Social Adaptations among Sibs with Special Reference to Social Insects. Evolution 11: 32-39.

Wilson, D. S. 1980. The Natural Selection of Populations and Communities. Menlo Park, CA: Benjamin/Cummings.

Wilson, D. S. 1983. The Group Selection Controversy: History and Current Status. Annual Review of Ecology and Systematics 14: 159-187.

Wilson, D. S. and E. O. Wilson. 2007. Rethinking the Theoretical Foundation of Sociobiology. The Quarterly Review of Biology 82 (4): 327-348.

Wilson, E. O. 1975. Sociobiology: The New Synthesis. Cambridge, MA: Harvard University Press.

Wright, S. 1945. Tempo and Mode in Evolution: A Critical Review. Ecology, 26: 415-419.