Studies in History and Philosophy of Biological and Biomedical Sciences 56 (2016) 32–38

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Can altruism be unified?

Grant Ramsey

Institute of Philosophy, KU Leuven, Belgium

ARTICLE INFO

Article history:
Available online 14 November 2015

Keywords:
Altruism
Psychological altruism
Biological altruism
Helping
Evolution
Fitness

ABSTRACT

There is clearly a plurality of forms of altruism. Classically, biological altruism is distinguished from psychological altruism. Recent discussions of altruism have attempted to distinguish even more forms of altruism. I will focus on three altruism concepts, biological altruism, psychological altruism, and helping altruism. The questions I am concerned with here are, first, *how should we understand these concepts?* and second, *what relationship do these concepts bear to one another?* In particular, is there an essence to altruism that unifies these concepts? I suggest that while there is no essence to altruism, this does not mean that the array of altruism concepts is completely disunified. Instead, I propose we place all the concepts into a common framework—an altruism space—that could lead to new questions about how this space can be filled.

1. Introduction

'Altruism' clearly has diverse meanings. In discussions of altruism, many are careful to distinguish between biological (or evolutionary) altruism and psychological altruism. Biological altruism is often understood to center on fitness exchanges, whereas psychological altruism is based on intentions—an act is psychologically altruistic not because of the outcomes, but because of particular intentions of the actor. This distinction has become all but standard in the study of altruism (Sober & Wilson, 1998).

The psychological–biological distinction, however, does not appear to exhaust the range of varieties of altruism. The reason for this is that there are forms of altruism that are not clearly either biological or psychological in nature. Some instances of helping, for example, count as altruism independently of both the psychological mechanisms driving the behavior and their fitness consequences. Such 'helping altruism', as I will call it, is a genuinely distinct form of altruism.

In this essay, my goal is to clarify the taxonomy of altruism concepts and to consider whether this diversity merely constitutes distinct concepts loosely related and collected under the rubric of altruism, or whether there is a deeper unity. I propose that while there is no essence to altruism, one can take what I suggest are the three central altruism concepts, render them as single scalar values, and construct a three-dimensional altruism space.[1] This space will open up new empirical questions about how the space can be filled and why particular regions are, or are expected to be, empty.

2. A taxonomy of altruism concepts

How many concepts of altruism are in circulation and what are their natures? This question, it turns out, is not an easy one to answer. The reason is that there is no standard array of altruism concepts and associated terms that can be relied upon to answer this. Instead, one must read the literature carefully to attempt to extract implied meanings in the various uses of 'altruism'. There has, however, been a recent attempt to do just that. Clavien and Chapuisat (2013) have identified what they take to be four distinct concepts of altruism. I will thus begin with their framework and modify it in several ways.

E-mail address: grant@theramseylab.org.

[1] Although his project is quite different, this is in the spirit of Godfrey-Smith's (2009) Darwinian space.

http://dx.doi.org/10.1016/j.shpsc.2015.10.007
1369-8486/

2.1. Biological altruism

Let's begin with biological altruism, the concept of altruism tied to biological fitness. This form of altruism is also called evolutionary or, by Clavien and Chapuisat, reproductive altruism. None of these terms is entirely appropriate. While this form of altruism is linked to fitness/selection, it is a mistake to think that it is tied to just an individual's reproductive output: Measures of reproductive success such as lifetime reproductive success (LRS) can serve as imperfect proxies for fitness, but fitness can change without changes in LRS. To see how LRS can deviate from fitness, consider two types of individuals, where one type is disposed to reproduce earlier in their life history than the other, but they are otherwise identical in terms of their longevity, health, etc. Assuming that the organisms have overlapping generations (they are not, say, restricted to reproducing once each spring), the type that reproduces earlier in its life history will increase in proportion over time. This is true because over a given span of time, the early reproducing type will wedge in more generations and each individual of that type will, on average, have more descendants (assuming, of course, that there are no countervailing effects of early reproduction). The early reproducing type will thus be the fitter type of individual in spite of having the same LRS.

Just as LRS is too restrictive, tying biological altruism to evolution does not work either. The term 'evolutionary altruism' points correctly to the link between this form of altruism and core evolutionary concepts. But the dispositions to behave altruistically in the other senses discussed below are not somehow outside of evolution—they can certainly be evolved traits. A more appropriate term would be 'fitness altruism' or 'selection altruism' since fitness/selection are definitionally linked with this form of altruism. But because I hesitate to coin yet another synonym for this form of altruism, I will henceforth use what is perhaps the most common term, 'biological altruism'.

Clavien and Chapuisat define biological altruism thus: "A behaviour is altruistic if it increases other organisms' fitness and permanently decreases the actor's own fitness" (2013, p. 128). Similarly, Sober and Wilson (1998) hold that "A behavior is altruistic when it increases the fitness of others and decreases the fitness of the actor" (p. 17).[2] There are two things to notice about these definitions. First, they involve a loss to the actor and a benefit to the recipient(s)—it is not enough that the actor loses or that the recipient benefits, both must occur. Second, the fact that the actor relinquishes some of its fitness to boost the fitness of the recipient means that the fitness of individual organisms is something capable of changing as a result of these behaviors. Let's consider whether fitness can change in these ways.

If fitness is to causally explain evolutionary outcomes, then it cannot be equivalent to those outcomes. One way that philosophers have proposed to avoid the equation of fitness and outcome is to consider fitness to be a probabilistic propensity to produce offspring, not actual offspring produced (Brandon, 1978; Mills & Beatty, 1979). A corollary of this view, or so argue Ramsey (2006) and Abrams (2009), is that the fitness value a particular organism has does not change from moment to moment. (Ramsey coined the term 'block fitness' for this understanding of fitness and I will follow his nomenclature.) The block fitness concept has urged some to rethink the way that biological altruism should be understood (Ramsey & Brandon, 2011). The core idea is that organisms have particular fitness values, and that these values are based on their hereditary material, the environment that they are born into, the possible future states of this environment, and their possible interactions with it. If organismic fitness is a function of these properties, then it will be fixed over the life history of the individual. While it is true that the organism's health can fluctuate—it can become ill or remain healthy—its fitness does not fluctuate accordingly. Even ending up sterile does not lower one's fitness. Furthermore, while it is true that bearing viable offspring will raise an individual's realized fitness, it will not raise its fitness. Realized fitness is a tally of outcomes, whereas fitness is the weighted probability distribution over the possible outcomes.

For those who are skeptical of the block fitness idea, consider this analogy: If we have a coin and a coin-flipping device and we flip the coin a number of times, we can produce several interesting quantities: (1) the probability that the coin has of landing head up prior to being flipped, (2) the instantaneous probability of landing head up at each moment throughout the course of its flips, and (3) the number of times the coin lands head up. The last of these is what we can analogize with the realized fitness of the coin. It is the result of the coin's propensity, combined with the chance features of particular coin flips. These outcomes are not identical with the coin's chances of landing head up, but serve as evidence for it. The second of these quantities is neither realized fitness nor fitness, though is sometimes confused with the latter. If the world is fundamentally indeterministic, or if the probabilities are based on partial information, then the values for (2) can vary over the life of a coin flip. But such a quantity (an instantaneous probability) will be of little use for predicting or explaining or understanding the outcome of entire coin flips, though it could be useful in understanding some elements of the dynamics of coin flips. Like (3), it is an outcome—it is an outcome of the chance path that the coin has taken, combined with the coin's weighted possible future paths. Such a measure partway through the flip of a coin may provide a useful estimate for the probable fate of the coin, but it is not a good estimate for how the coin will do when flipped again, or what the outcome is likely to be from a large number of such flips. For estimates of this kind, we need quantity (1).

Quantity (1) is given by the properties of the coin (its symmetry, etc.) and the environment (whether it acts differentially with respect to each side of the coin). It does not fluctuate from moment to moment. The tallies of flip outcomes do, of course, change—they are ratcheted up over generations of coin flips. The first quantity, the probability of landing head up, is analogous to the block fitness of organisms. Like block fitness, it does not fluctuate from moment to moment. If a coin has a 0.5 probability of landing head up, this is true of the coin even if its instantaneous probability changes, and even if the coin is damaged or otherwise transformed during its flip (see Ramsey, 2006 for a more extensive discussion of this point).

Quantity (1) is what is analogous to fitness. Just as the fitness of a coin is a function of the set of possible ways it can undergo its flip—and the associated probability-weighted outcomes—so is the fitness of organisms based on their possible life histories. And while fitness is based on the set of possibilities, realized fitness is based on the one life history that the organism realizes. If this is true, then biological altruism needs to be reconceived: Altruistic acts are no longer acts whose performance lowers the fitness of the actor and raises the fitness of the recipient. How then should we reconceive biological altruism?

A full explication and defense of a revised account of biological altruism is well beyond the scope of this paper, but what I will say here is this: Biological altruism should not be taken to be based on

[2] A further distinction can be made between biologically strong altruism and weak altruism. The strong variety requires a cost to the actor and benefit to the recipient(s), whereas the weak variety includes a benefit for the recipient(s) and a more modest benefit for the actor (see Kerr, Godfrey-Smith, & Feldman, 2004 & Wilson, 1990 for a discussion). Both strong and weak altruism thus require that the fitness benefit to others does not exceed a fitness benefit to the actor. Because of the relative unity of these concepts I will not further discuss weak altruism in this paper.

the outcomes of individual actions, but should instead be based on the organism's propensities to act and the probable outcomes of its actions. Altruistic acts are therefore ones that issue from a particular kind of disposition and are associated with a particular kind of outcome. Thus, instead of taking its outcome to be determinative of being an altruistic act, it is better to understand it in the following way:

> An act is altruistic if and only if having the act in the behavioral repertoire of the individual (1) lowers its fitness and (2) increases the fitness of the recipient(s) (or the group to which the altruist belongs).

This concept of altruism, then, follows directly from the fact that fitness is a propensity, and that the value of an organism's fitness is invariant over its lifetime. Consider again Clavien and Chapuisat's definition: "A behaviour is altruistic if it increases other organisms' fitness and permanently decreases the actor's own fitness" (2013, p. 128). They are correct about the permanence, but if this is understood as the act having a permanent effect, this is wrong. Instead, it is permanent in that it is a disposition that arises from the fixed organism-environment setup—the organism's genes, the environment into which it is born, and the probable future environmental features.

2.2. Psychological altruism

Now consider psychological altruism. Stich (2007) offers a clear definition of this concept: "A behavior is psychologically altruistic if and only if it is motivated by an ultimate desire for the well-being of some other organism, and as a first pass, we can say that a desire is ultimate if its object is desired for its own sake, rather than because the agent thinks that satisfying the desire will lead to the satisfaction of some other desire" (p. 286). Let's consider the details of this definition and how it differs from biological altruism.

Whereas biological altruism is linked to core behavioral dispositions, psychological altruism is based on facts about the psychological states of individuals. And although the desires adverted to in Stich's definition are desires for future states, these states do not need to be realized in order for the behavior to be altruistic. An individual who gives money to a charity, where this gift is motivated by an ultimate desire to help others, will be performing an altruistic act even if the money is mishandled and ends up doing no good. And an individual who gives money to a charity that does in fact increase the well-being of others, but whose gift was made only in order to quash the rumors that they are greedy, is not acting altruistically in this sense.

While psychologically altruistic behaviors necessarily stem from certain kinds of ultimate desires, just as the behaviors do not in fact have to result in increasing the well-being of others, the behaviors also bear no necessary relationship with biological fitness. Being disposed to act for the well-being of others does not necessarily mean that one is less fit than an individual who is psychologically selfish—it could be that psychologically altruistic individuals fare better in terms of biological fitness. That is, there is no conceptual link between biological and psychological altruism. This is not to say that there are no empirical links between them. It could, for example, be the case that psychologically altruistic mechanisms are a good way to realize biological altruism and that, as Sober and Wilson (1998) suggest, "natural selection is unlikely to have given us purely egoistic motives" (p. 12). These empirical points are of interest, and below I will reflect on the sort of empirical questions that one can (or should) investigate with respect to altruism. But for the sake of this section, only the conceptual points are relevant.

Psychological and biological altruism are the two forms of altruism chiefly discussed in the literature. And from these discussions, it might appear that these are the only two forms of altruism. Despite appearances, however, these are not the only forms of altruism in existence. I will now articulate another category of altruism.[3]

2.3. Helping altruism

Although biological and psychological altruism are the prototypical forms of altruism discussed in the literature, there is another form of altruism that has significant currency, what I will label *helping altruism*. To understand helping altruism, consider this quote of Warneken and Tomasello (2008): "we may ask whether human altruism is intrinsically or extrinsically motivated; that is, do human beings help one another because the helpful act itself is inherently rewarding or only because the helpful act is instrumental in bringing about separate outcomes such as material rewards or the avoidance of punishment?" (p. 1785). Here, you see an unqualified slip from "human altruism" to "human beings help one another." This implies that human altruism in this sense is fundamentally based on—or definitionally intertwined with—human helping.

Helping in their sense is not necessarily linked to psychological states or fitness impacts—helping does not need to be helping procure food, raise young, etc. Instead, it could be as simple as aiding another in reaching their goal or otherwise improving their life conditions. In the paper from which this quote was drawn, their tests for helping consisted of adults dropping objects, reaching for them, but not being successful in retrieving them. The experiment gathered data on the extent to which the children would pick up the objects. Such helping behaviors of course may be fitness relevant, or may be linked with particular psychological states. But what is important is that psychology and fitness are not definitionally linked to altruism in this sense.

This concept of altruism is widespread and we see similar definitions by other researchers. Consider de Waal's (2008) definition of what he is calling *directed altruism*: "helping or comforting behavior directed at an individual in need, pain, or distress" (p. 281). This form of altruism can arise in either of three ways, "1. Altruistic impulse. Spontaneous, disinterested helping and caring in reaction to begging or distress signals or the sight of another in pain or need. 2. Learned altruism. Helping as a conditioned response reinforced by positive outcomes for the actor. 3. Intentional altruism. Help based on the prediction of behavioral effects" (p. 281). Again, we see that this form of altruism may, and may even typically, be associated with psychological states of the actor or recipient, or may be relevant to fitness outcomes, but it is not definitionally linked with either psychological states or fitness outcomes.

Helping altruism is closely associated with Clavien and Chapuisat's *behavioral altruism* since they both center on benefiting others, where this benefit is not necessarily fitness-relevant and is

[3] Interestingly, Clavien and Chapuisat break psychological altruism into two related concepts, psychological and preference altruism. They define psychological altruism similarly to Stich's definition given above. For them, "An action is [psychologically] altruistic if it results only from motivations directed towards the goal of improving others' interests and welfare" (2013, p. 127). Preference altruism is defined in the following way: "An action is altruistic if it results from preferences for improving others' interests and welfare at some cost to oneself" (p. 131). These notions are clearly closely related. Both critically depend on the etiology of the behavior, whether or not it issues from a particular sort of desire or preference. Because they share this similarity, I will lump them together here, since the purpose is to trace out the broad categories of altruism. But of course for some studies, it may be beneficial to recognize this distinction.

not by definition linked to psychological desires: "A behaviour is altruistic if it brings any kind of benefit to other individuals at some cost for the agent, and if there is no foreseeable way for the agent to reap compensatory benefits from her behaviour" (2013, p. 131). They propose this form of altruism because they recognize that in some fields altruism is measured in ways that are not necessarily connected to psychology or fitness, for example, "In experimental economics, costs and benefits are usually translated into monetary units" (p. 132). Clavien and Chapuisat's definition does a good job capturing the concept used in experimental economics, but it would need to be modified to some degree if it were to include the forms of helping just described. For one thing, the requirement that the behavior be costly does not seem to fit with how the definition is used by de Waal, Warneken, Tomasello, and others. There is of course always an opportunity cost to performing actions (in terms of time and energy), but such costs cannot be the ones Clavien and Chapuisat are referring to, since by pointing out that the behavior must be costly, they are implicitly contrasting it with noncostly (that is, trivially costly) behavior.

The second difficulty with Clavien and Chapuisat's definition is that it subsumes mistakes within behavioral altruism. The salmon that accidentally leaps into the mouth of a wading bear is not being altruistic even though this leap is beneficial to the recipient and there is no way that the bear can repay the salmon for its behavior. Similarly, if an Olympic gymnast tragically falls down at the end of her routine, she benefits her competitors who are vying for medals. But this is no altruistic act on her part. The fact that Clavien and Chapuisat's definition renders both of these cases acts of altruism shows that their definition is problematic if the intent is to separate mistakes from altruistic acts. This, however, should not be understood as a critique of Clavien and Chapuisat's paper. On the contrary, their project is to try to capture the definitions in circulation in the literature, not to offer definitions that they think should supplant those definitions. Thus, their project is descriptive, not prescriptive or stipulative. The project here is to some extent prescriptive; I am offering what I think are better alternative definitions, which accord with research practices and help to clarify and unify altruistic phenomena.

How might one head off these difficulties? One way is to view token actions only within the context of the adaptations that bring them about. This is the approach pursued by Tooby and Cosmides (1996): "An adaptationist definition of altruism would focus on whether there was a highly nonrandom phenotypic complexity that is organized in such a way that it reliably causes an organism to deliver benefits to others, rather than on whether the delivery was costly" (p. 123). This solves problems like the salmon or gymnast. Salmon are adapted to swim upstream and to occasionally jump over small waterfalls. A side effect of this is that they occasionally leap into the mouths of bears. Similarly, the gymnast does not have an adaptation for falling in Olympic competitions, thus the behavior does not count as altruistic.

If an adult drops a ball, reaches for it, perhaps making an effortful grunt, and then a twenty-month-old walks over, picks it up, and hands it to the adult, the infant helped the adult. Although, such studies may not specifically identify these behaviors as adaptations, there is nevertheless reason to think that they center on adaptations, in this case adaptations for reading goals of others and helping others in achieving these goals. Tying helping altruism to adaptations also does not rule out learned behavior. Consider again de Waal's conception of learned altruism: "Helping as a conditioned response reinforced by positive outcomes for the actor." Here, the adaptations are in terms of adaptations for conditioning, for taking particular sorts of outcomes as positive, etc.

Clearly this suggestion to base helping altruism on evolved adaptations needs further elaboration and a broad survey of the uses of helping altruism to see if it captures the right phenomena. But for the purposes here, what matters is that helping altruism is helping behavior that is not a mistake. I believe that the adaptation-centered way of eliminating mistakes is promising, but other ways may do so even better.

3. How are these altruism concepts related to one another?

Now that we have an overview of the three basic categories of altruism, we can ask, *what relationship do these forms of altruism bear to one another?* One response is to argue that there is a nested relationship among them. A second response is to argue that while there may not be a nested relationship, there is nevertheless an essence to these altruism concepts, that they bear some essential property in common. By this, I mean that there is a property or set of properties that are necessary and sufficient for being altruism. A third is to argue that 'altruism' is simply polysemous, that the same word is used for what are clearly distinct—and not nested or essentially linked—concepts. I will argue for the polysemy position, but I will suggest that this does not mean that we should not study altruism more generally. Instead, in the following section, I will offer a way to subsume the diversity of altruism into a single altruism space.

The fact that we refer to all these forms of altruism as altruism lends at least weak evidence to the idea that there is some thread running through each. In order to investigate their relationship, let's make a systematic comparison of their central qualities. To do so, consider Table 1.

Does the table imply a nested relationship among the altruism concepts? It appears that a nested relationship is argued for by Clavien and Chapuisat for at least some of their altruism concepts: "To summarize the relationship between these two notions, behavioural altruism is much broader than reproductive altruism. The latter is used in a specific way in evolutionary biology and may be seen as a special case nested within the broader category of behavioural altruism" (2013, p. 133). For the way that I have articulated the three core concepts of altruism, it appears that a nested relationship does not hold. While there may be cases of biological altruism that are also helping altruism, as I have described the concepts above, there can also be biological altruism that is not helping altruism and helping altruism that is not biological altruism. Assisted suicide could count as helping, despite being detrimental to fitness; tampering with another's birth control may be quite unhelpful, despite promoting biological fitness. In fact, as Table 1 makes plain, each of these forms of altruism are independent and can be realized without necessarily realizing the other forms of altruism. A nested view is thus wrong.

If the concepts are not nested, perhaps there are one or more core features that all of them share. An easy contender for an essence of altruism would be a row in Table 1 in which each slot is the same, in this case either all "Not necessarily" or all "Yes." But not only is there no such row, if we consider the places in which "Yes" occurs, there is no overlap whatsoever in the rows: none of the forms of altruism has a "Yes" for the same property. This provides

Table 1
The three basic forms of altruism along with some of their properties.

	Biological altruism	Psychological altruism	Helping altruism
Fitness benefit to recipient/group	Yes	Not necessarily	Not necessarily
Fitness detriment to actor	Yes	Not necessarily	Not necessarily
Help for recipient	Not necessarily	Not necessarily	Yes
Desire to benefit others	Not necessarily	Yes	Not necessarily

no definite proof of a lack of an altruism essence, but does imply that there is no essence related to the core altruism properties. Not all of the forms of altruism require a benefit (fitness or otherwise) to the recipient, not all of them require a detriment (fitness or otherwise) for the actor. Given this fact, there is a burden of proof placed on the essentialist to produce a convincing essential property. I am skeptical that there is any such property forthcoming, since there is none present in the core features of altruism in Table 1, and any other factors in common would be unlikely to be central to altruism.

If there is neither a hierarchy nor an essence to altruism, we should understand 'altruism' as polysemic, referring to quite distinct concepts: biological altruism centers on behavioral dispositions (and their fitness impacts), psychological altruism centers on the motivations behind the behaviors, and helping altruism centers on the helpfulness of the behavioral outcomes. The absence of an essence or hierarchy, however, does not mean that the altruism concepts are sufficiently unrelated that we should not look for connections between them—it is not merely the label 'altruism' that is what is tying all of the forms of altruism together. If, for example, people began to refer to each truism about Al Gore—Al Gore was vice president, Al Gore is concerned about climate change—as 'altruism' this does not mean that this is another form of altruism.[4] Instead, altruism is a family of concepts centering on benefiting others and possibly costing the actor. It is thus worth continuing to pursue the question of how these concepts are related to one another, and to do this, I will introduce an altruism framework.

4. An altruism framework

The proposed essentialist and nested relationships between the three forms of altruism are conceptual relationships; they offer a view of how these three concepts are related to one another. The nested proposal argues that they are in some way tied together in virtue of their nested relationship. The essentialist proposal is that there is one or more core features that these concepts bear that unite them as altruism. My proposal is that we should search for empirical, not conceptual ties between these forms of altruism.

The idea is this: If each of these forms of altruism can be captured by a single scalar quantity from, say, 0 to 1, then we could think of each form of altruism as representing an axis in a three-dimensional altruism space (see Fig. 1). The representation of this altruism space will allow for formulating and testing interesting empirical questions, ones that may not have been considered otherwise. But before we can get there, let's consider whether such a space is indeed coherent. In particular, can each form of altruism be quantified by a single scalar quantity? And if so, what is the best way to scalarize them?

In order to scalarize altruism, we must first consider whether we are focused on token actions or on token individuals. That is, we could be considering how altruistic individuals are, or we could consider how altruistic their particular actions are. Thus, before we create a rank ordering of each of the forms of altruism, we need to first decide whether we want the space to be a space of organisms or their behaviors. While both are possible, for the purposes here, we should pursue a space of actions, not organisms. The reason is that defining how altruistic an individual is requires an understanding of the degree to which that individual is disposed to perform altruistic actions, which in turn requires some sort of ranking of how altruistic these actions are. Furthermore, such ranking of individuals is made all the more difficult by the fact that

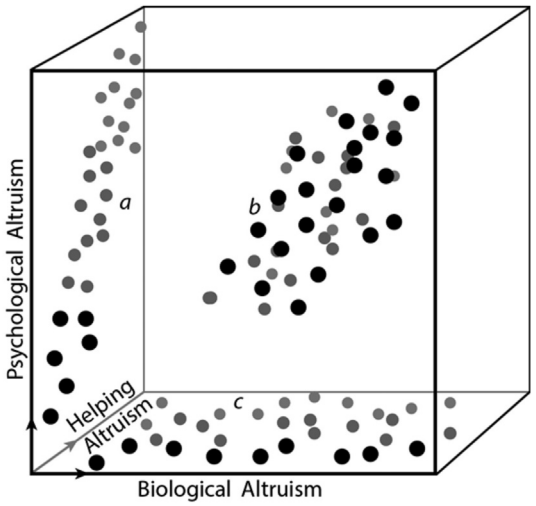

Fig. 1. The three basic forms of altruism represented as axes in a three-dimensional space.

patterns of exhibited behaviors can vary in multiple dimensions, including their frequency, intensity, the variance in their intensity, etc. Distilling all these variables into a single value is thus no easy task, and it is a task we can sidestep if we focus on actions alone. Therefore, the space I will consider here will be for actions only, not individuals.

4.1. Helping altruism

What would a rank-ordering of helpfulness consist in? Is there a common metric for quantifying helpfulness such that comparisons within and across species are possible? Is an ant helping to drag a beetle larva to its nest being more or less helpful than a chimp picking lice from the back of a fellow chimp? Such a question is all but impossible to answer, at least when presented with these two behaviors in isolation. A better approach is instead to consider a single domain, such as the ant helping drag the grub. For such a case, it is fairly clear what it would be to have a helpfulness value of 0, it would be not helping at all in the process of dragging the grub. (There could of course be unhelpful behavior—pulling in the wrong direction in such a case.) And maximal helpfulness could be understood as taking over the entire task of hauling the grub or, if this is impossible given the size of the grub, maximizing the amount of time and effort it gives to the task.

Understood in this way, the values along the helpfulness axis in the figure are relative helpfulness, not absolute helpfulness. Thus, if we fill the space with ant, chimp, and human behaviors, we must understand that each of these is quantified within its own relative scale. This does not, however, mean that such cross-taxa comparisons are spurious. Instead, it is merely that they have to be interpreted carefully. Discovering that helpfulness is positively correlated with another form of altruism will be of interest, even if these are relative notions of altruism—in fact, they may be of more interest because this is the case.

4.2. Psychological altruism

Now consider psychological altruism. Psychological altruism is based on an act for which the ultimate desire is the wellbeing of another. How can such acts be ordered from more to less psychologically altruistic? A challenge in making this ordering is that one's desire can vary independently of the degree to which the behavior

[4] Thank you to one of the reviewers for the Al Gore example.

boosts the well-being of the recipient. One can have a strong desire to effect a modest boost in another's wellbeing, and one could have a moderate desire to effect a strong boost to their wellbeing.

There are of course multiple solutions to the puzzle of distilling psychological altruism into a single value, but I would suggest the simple solution of considering the desire and the anticipated wellbeing as two scalar components of psychological altruism. A single scalar value could thus be achieved by combining them in some nonadditive way, multiplying them together being a simple possibility. This single scalar is represented as the vertical axis of Fig. 1.

4.3. Biological altruism

Finally, consider biological altruism. This is the most challenging form of altruism to place into a single scalar quantity. The reason for this is that while psychological and helping altruism answer "yes" only to one each of the properties in Table 1 ("desire to benefit others" and "other benefit to recipient," respectively), biological altruism answers yes to both fitness benefit to recipient/group and fitness detriment to actor. Fitness cost to actor and benefit to recipients/group seem sufficiently heterogeneous that this form of altruism would require (at least) two dimensions. But while there may be times when such a two-dimensional representation would be useful, for the purposes of inquiring into the broader relationships between the three forms of altruism, it is beneficial if each of these forms of altruism can be represented by a single value.

A good way to reduce biological altruism to a single value is to do what was done for psychological altruism (and the two factors of the degree of desire and the degree of well-being), to mathematically combine them to make a single scalar. Again, one could multiply the two factors together or perform some more sophisticated operation on them. Some ways of reducing the variables into a single scalar will be better than others. Adding them, for example, will be problematic. Consider the case where there is no detriment to the actor. Such a case is clearly not an instance of biological altruism. But if there were still a benefit to the recipient(s), and if one derived the scalar from the sum of this benefit and the detriment to the actor, then the scalar would be positive (it would register as biologically altruistic). If, on the other hand, the values were multiplied, then the correct value of zero would be registered for such actions that bear no cost to the actor. I will thus consider the horizontal axis in Fig. 1 to be the product (or some other mathematical scalarization) of the fitness detriment to the actor and the fitness benefit to the recipient(s)/group.

4.4. The altruism space

Now that we have each of the dimensions fleshed out, we can begin to consider what empirical questions can be asked of the framework. Consider the question of the link between psychological and biological altruism. As mentioned, Sober and Wilson (1998) argue that while there is logical independence of psychological and biological altruism, the two may be empirically linked: "natural selection is unlikely to have given us purely egoistic motives" (p. 12). Others, such as Stich (2007), have challenged Sober and Wilson's reasons for holding that psychological altruism is a probable evolutionary outcome. Regardless of which side of this debate is correct, the altruism space of Fig. 1 allows us to ask many more empirical questions, and to be able to move far beyond simple, two-variable comparisons.

The addition of helping altruism opens up a third dimension of possible relationships between the varieties of altruism, and allows one to see the complex relationships that the forms of altruism bear to one another. Instead of focusing on simple comparisons between pairs of variables, we can investigate the infinite number of ways that the three forms of altruism could be related to one another. For example, a study of behavior might reveal a tight link between psychological and helping altruism, but the behaviors may have little or no fitness consequence (*a* in Fig. 1). Another possibility is that the group of behaviors cluster close to the middle, such that many of them exhibit intermediate amounts of helping, psychological, and biological altruism (*b* in Fig. 1). Yet another possibility is that the behaviors do not exhibit any psychological altruism (perhaps the organism is unable to form the requisite forms of desire), and that helping and biological altruism are not associated (*c* in Fig. 1).

In addition to the three distinct possibilities represented by *a, b,* and *c* in Fig. 1, one could also imagine a single study producing a scattering of points like that represented by the totality of *a*, *b,* and *c.* The fact that such a possibility exists shows that by moving beyond mere dyadic comparisons and using a space like that of Fig. 1, one can represent and evaluate the richness and complexity of the relationship between the three forms of altruism.

This is not to say that visually depicting altruism in three dimensions will always be the most useful representation. Conducting the analyses mathematically and then presenting multiple two-dimensional graphs may in some cases be visually more informative, since three-dimensional graphs projected onto two dimensions have considerable limitations. But however the data are presented visually, the point nevertheless holds that there is use in investigating the patterns within this larger altruism space.

4.5. Is this altruism space the best altruism space?

In distilling each form of altruism into a single scalar, one might wonder if too much information is lost in the formation of the space. Might there be a better information-preserving space that retains all independent dimensions related to altruism? Such a space could be constructed, not from the columns in Table 1, but from the rows: fitness benefit to recipient/group, fitness detriment to actor, help for recipient, desire to benefit others. In fact, one could unpack these categories into further dimensions. Other benefit to recipient could be several dimensions, one for each kind of benefit in some baroque taxonomy of benefits.

There are five reasons why these complex n-dimensional spaces, built on more atomic instead of composite dimensions, may not be more illuminating. First, while in some cases atomic dimensions may be more information preserving than composite ones, the very act of breaking composites down into atoms is laden with assumptions. One might, say, break down the helping category into a number of distinct forms of helping. But what justifies this taxonomy? There are many ways to create a taxonomy such as this and each carries its own set of assumptions and interpretations. Thus, the added assumptions built into the more complex n-dimensional space may obscure as much as clarify the phenomenon of altruism.

Second, it is not clear that any array of putatively atomic variables will be truly atomic. Consider fitness benefit to others. Fitness benefits can be distributed in an infinite number of ways. For example, there could be a large benefit to a small number of individuals and a small benefit to others. Or perhaps each individual will receive a unique quantity of benefit, directly proportional to their age, say. How do we create atomic dimensions for such a case? For such a distribution of benefits, we could have one dimension be the mean benefit, another the variance, another the kurtosis, etc. But of course these are not atomic, they are statistics pooled from sets of data across a population. Perhaps instead we should have a distinct axis for each individual, which records the benefit received.

This suggestion leads to the third problem with the atomic proposal.

Third, by proliferating dimensions, we may be able to maximize data preservation, but we end up with a space merely containing the raw data, and not one that illuminates their connections. The space is then not a model of altruism, it is just a container for raw data. One could of course use statistical techniques like cluster analysis to reveal patterns within the multidimensional space, but this then leaves the difficult task of relating these patterns to discussions of altruism in the literature. Related to this is the fourth problem, which is that if this space is supposed to illuminate these various forms of altruism, then there must be a way of delimitating regions within the space as biological altruism, helping altruism, etc. But in order to do so, the same kind of variable composites inherent in the space I propose must be created in order to accomplish this.

Finally, by limiting the altruism space to only three dimensions, it is something that can be graphically represented in a way that a higher n-dimensional space cannot. This allows for visual inspection of the patterns and, thus, greater insight into the structure of the space. Although the question of which set of dimensions is the most insightful is ultimately an empirical question, these reasons lend support for the claim that the three-dimensional space proposed here is illuminating and may even occupy an ideal place between overly crude mere measures of associations between pairs of variables, and a complex n-dimensional space.

5. Conclusions

Can altruism be unified? The answer of course critically depends on how one understands altruism and what one means by unification. If unification requires one or more essential properties, then unification may not be possible. What I have instead argued for here is that there are three fundamentally distinct forms of altruism, and that there is no unique property that they all share. But despite this disunity, there is a way to make a unified altruism space. This space can be used to explore whether the forms of altruism are empirically linked and, if so, what relationships they bear to one another. The rich possibilities offered by the space allow for an unlimited number of ways that these forms of altruism can be related. And only through such a space can the full complexity of altruism be revealed.

Acknowledgments

I would like to thank the editors for this special issue, Justin Garson and Armin Schulz, for their guidance, and to the anonymous reviewers for their insightful comments and critiques. This paper was completed while a fellow at the National Humanities Center. I thank the Center for its support.

References

Abrams, M. (2009). Fitness "Kinematics": Biological function, altruism, and organism-environment development. *Biology and Philosophy, 24*(4), 487–504.
Brandon, R. N. (1978). Adaptation and evolutionary theory. *Studies in History and Philosophy of Science Part A, 9*(3), 181–206.
Clavien, C., & Chapuisat, M. (2013). Altruism across disciplines: One word, multiple meanings. *Biology & Philosophy, 28*(1), 125–140.
De Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*, 279–300.
Godfrey-Smith, P. (2009). *Darwinian populations*. Oxford University Press.
Kerr, B., Godfrey-Smith, P., & Feldman, M. W. (2004). What is altruism? *Trends in Ecology & Evolution, 19*(3), 135–140.
Mills, S. K., & Beatty, J. H. (1979). The propensity interpretation of fitness. *Philosophy of Science, 46*, 263–286.
Ramsey, G. (2006). Block fitness. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 37*(3), 484–498.
Ramsey, G., & Brandon, R. (2011). Why reciprocal altruism is not a kind of group selection. *Biology and Philosophy, 26*(3), 385–400.
Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, Mass: Harvard University Press.
Stich, S. (2007). Evolution, altruism and cognitive architecture: A critique of Sober and Wilson's argument for psychological altruism. *Biology & Philosophy, 22*(2), 267–281.
Tooby, J., & Cosmides, L. (1996). Friendship and the Banker's paradox: Other pathways to the evolution of adaptations for altruism. In W. G. Runciman, J. M. Smith, & R. I. M. Dunbar (Eds.), *Evolution of social behaviour patterns in primates and man*. Proceedings of The British Academy, *88*, 119-143, Oxford University Press.
Warneken, F., & Tomasello, M. (2008). Extrinsic rewards undermine altruistic tendencies in 20-month-olds. *Developmental Psychology, 44*(6), 1785–1788.
Wilson, D. S. (1990). Weak altruism, strong group selection. *Oikos, 59*, 135–140.