Open access • Journal Article • DOI:10.1017/S0266267110000180

Altruism as a thick concept — Source link

Michael Schefczyk, Mark S. Peacock

Institutions: University of Zurich, York University

Published on: 01 Jul 2010 - Economics and Philosophy (Cambridge University Press)

Topics: Altruism (ethics), Reflective equilibrium and Normative

Altruism as a thick concept

Schefczyk, Michael

DOI: https://doi.org/10.1017/S0266267110000180

Economics and Philosophy, **26** (2010) 165–187
doi:10.1017/S0266267110000180

ALTRUISM AS A THICK CONCEPT

MICHAEL SCHEFCZYK
Zurich University

MARK PEACOCK
York University

In this paper, we examine different forms of altruism. We commence by analysing the 'behavioural' definition and, after clarifying its conditions for altruism, we argue that it is not in 'reflective equilibrium' with everyday linguistic usage of the term. We therefore consider a 'psychological' definition, which we likewise refine, and argue that it better reflects ordinary language use. Both behavioural and psychological approaches define altruism descriptively and thus fail to capture an important aspect of altruism, namely its normative component. Altruism, we argue, is a 'thick concept', i.e. one which embodies both positive and normative components. We discuss and compare various formulations of this normative component.

INTRODUCTION

Philosophers sometimes refer to lines of reasoning which use words idiosyncratically as Humpty Dumpty arguments, alluding to a dialogue in Lewis Carroll's *Through the Looking-Glass* in which Humpty Dumpty asserts that a word 'means just what I choose it to mean'. His interlocutor, Alice, is not convinced: 'Glory', she insists, 'doesn't mean 'a nice knock-down argument', as Humpty Dumpty chooses it to mean. We have often experienced that philosophers and economists follow the dramaturgy of Carroll's dialogue in their debates about altruism; philosophers play Alice's part and complain that 'altruism' does not mean what behavioural economists choose it to mean, whereas the latter assert their right to define a concept in a way that makes its instantiations observable in

We would like to thank two referees from the journal and, in particular, Christian List for many valuable comments on a first draft.

empirical studies. Such exchanges often end with the conclusion that the parties are arguing past, not with, one another. Economists strive for an enhancement of their behavioural model in the light of experimental results, whereas philosophers seek a deeper understanding of the concept, or, put somewhat naively, they want to settle what altruism (really) is.

The latter aim raises the question how one can acquire knowledge about what altruism is. The notion that there is a truth about the concept which research in economics fails to grasp is dangerously close to Platonism. But settling what altruism is cannot plausibly mean that one strives to make the term correspond to a Platonic idea. An alternative approach is to take established usage in ordinary language as the authority for what a concept means. This seems to be what many philosophers have in mind when they criticize behavioural definitions of altruism; they thus take a leaf out of John Stuart Mill (1965/1871: 49):

> [W]hen employing terms which common usage has taken complete possession of, it seems advisable so to employ them as to do the least possible violence to usage; since any improvement in terminology obtained by straining the received meaning of a popular phrase, is generally purchased beyond its value, by the obscurity arising from the conflict between new and old associations.

This passage contains an important insight and we point out, in Section I, that the behavioural definition does, indeed, strain the received meaning of altruism. However, Mill's dictum has limits when applied to altruism, for the term seems not to be one of 'which common usage has taken complete possession'; in this way, it differs from terms like 'glory'. Our linguistic intuitions may be definite enough to admit the identification of clear cases of appropriate or inappropriate usage of the term, yet they are too indistinct to serve as a basis for claims about what altruism really is in a broad range of cases. If, as we assume, an analysis of ordinary language does not suffice to settle the meaning of altruism, and if the Platonic road is not open any more, how are we to proceed?

We propose to employ an approach which is similar to Rawls' 'process of mutual adjustment of principles and considered judgments'. Rawls uses such a process in order to arrive at a description of the original position 'that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted'. If nobody sees the need for further adjustment, the model is in 'reflective equilibrium' (Rawls 1971/1999: 18). A similar approach can be used when it comes to calibrating terms like altruism: a definition of altruism is appropriate if it is in reflective equilibrium. Thus, we aim at defining the term so that no further adjustment in the light of well established theories, core examples and thought experiments is necessary. What Rawls describes as 'going back and forth' between principles and

considered judgements is, in our context, a process of mutual alignment between stipulative definitions and linguistic intuitions. 'Going back and forth' is necessary because our intuitions about the correct use of the word 'altruism' (and its cognates) are not sufficiently clear and distinct; we have to correct these intuitions in light of the formal definitions with which we confront them and vice versa.

Section 1 examines the behavioural definition of altruism which, we argue, fails to do justice to our linguistic intuitions, above all because it lacks any mention of the goals of those whose actions are supposedly altruistic. To overcome the deficiencies of the behavioural definition, we turn, in Section 2, to *psychological altruism* which, unlike its behavioural counterpart, puts the goals of the agent at the centre of altruism research. We deviate, however, from orthodox definitions of psychological altruism by relinquishing the notion of the agent's 'ultimate goal' as the decisive concept in determining whether an action be altruistic. In the final section, we consider an aspect of altruism which is lacking in both behavioural and psychological definitions, namely its normative content. Altruism, we argue, is a *thick concept* which consists of a descriptive and an evaluative component. Descriptive definitions which do not heed this evaluative component are not in reflective equilibrium.

1. BEHAVIOURAL ALTRUISM

Like biologists, experimental economists define altruism in terms of the consequences of actions. This explains their contention that they use a 'biological definition' (e.g. de Quervain *et al*. 2004: 1257; for biological definitions, see Trivers 1971: 35; Wilson 1975: 117; Dawkins 1976: 4). On closer inspection, it is evident that this claim is to be taken with a pinch of salt. According to the biological definition, the behaviour of an organism, *O*, is altruistic if it enhances the reproductive fitness of another organism but diminishes the reproductive fitness of *O* itself.[1] The puzzle, then, is how altruism persists if it is a selectively disadvantageous trait?[2] Behavioural economists do not focus exclusively on the effects of actions on reproductive fitness; instead they conceptualize costs and benefits in terms of utility, or, in experimental settings, financial payoffs (e.g. Fehr and Fischbacher 2003: 785). In order to emphasize this difference, we distinguish a behavioural from a biological definition of altruism. The

[1] Dawkins (1976: 4) is an exception; he writes of increasing another's 'welfare' rather than reproductive fitness when he describes the benefits conferred by altruistic actions.

[2] Thus Robert Trivers' claim that evolutionary explanations of altruism aim to 'take the altruism out of altruism' (Sesardic 1995: 130). His theory of 'reciprocal altruism' is designed to show that seemingly altruistic behaviour is, in fact, advantageous in evolutionary terms.

former imposes two conditions:

> **Behavioural altruism:** An action, Φ, is altruistic if and only if it is costly to the agent who performs Φ (costliness condition), and it benefits another party (benefit condition).

We offer three remarks on how costs and benefits are best conceived in the behavioural definition of altruism.

Remark (i): The behavioural definition conceives the utility of the agent in terms of personal goods, services and resources. It thereby abstracts from hedonic rewards ('warm glow'), which the agent might experience in consequence of an altruistic action. Hence, helping a friend or making a donation is altruistic in the behavioural sense, even if the agent derives great pleasure from these activities.

Remark (ii): The behavioural definition refers to opportunity costs. Thus, an action Φ_1 can be costly to the agent A, even if $U_A(\Phi_1) = B_A(\Phi_1) - C_A(\Phi_1) > 0$, where $B_A(\Phi_1)$ and $C_A(\Phi_1)$ refer respectively to the benefits and costs to A of Φ_1. This is the case when an agent performs action Φ_1 when Φ_2 would yield a higher utility to the agent. The difference between $U_A(\Phi_1)$ and $U_A(\Phi_2)$ gives us the opportunity cost of Φ_1.

Remark (iii): The behavioural definition refers to *total opportunity costs* for the benefactor B. That is, there will be no compensating benefits to B in the future which will outweigh the opportunity costs that B carries by performing Φ_1. It is on this issue that proponents of 'strong reciprocity' (Fehr *et al*.) and 'reciprocal altruism' (Trivers) differ, for the latter, but not the former, 'relies on the idea that altruistic behaviour creates economic benefits for the altruist in the future' which outweigh the present costs of acting altruistically (Fehr and Fischbacher 2003: 789).

We now offer a criticism of behavioural definitions which lead us to reject them in favour of a psychological definition. The behavioural definition of altruism is silent about the agent's goal(s), motives and intentions. This indifference is understandable in non-human contexts but less so in the human realm. To be sure, the *definitional* irrelevance of motives and intentions in behavioural studies does not entail a *theoretical* disinterest in them. Behavioural researchers are, for instance, aware that human altruism is not 'automatic' but 'based on deliberation and intent' (de Quervain *et al*. 2004: 1254), and behavioural studies aim to uncover the motives which engender altruistic behaviour (Fehr and Fischbacher 2003: 785). Thus, behavioural altruism among subjects who punish non-cooperative behaviour is hypothesized to stem from 'negative emotions' such as anger or from a desire to 'retaliate' or to take 'revenge' on those of whose behaviour one disapproves (Fehr and Gächter 2002: 130; de Quervain *et al*. 2004: 1254). So while motivations and intentions are absent from the *definition* of behavioural altruism, behavioural researchers do not overlook their *causal* role in producing altruistic behaviour. But despite

behavioural researchers' attentiveness to motives and intentions, there is reason to introduce 'psychological' components at the definitional phase, because some acts fulfil the costliness and benefit conditions but should intuitively be excluded from the class of altruistic acts. If, by mistake, I leave a cherished book on the bus and someone else takes it home and reads it with pleasure, my action fulfils the costliness and benefit conditions and is, therefore, to be classified as (behaviourally) altruistic. But it is implausible to call an action altruistic if its proximate cause is inattentiveness, for such 'altruism' is accidental.[3]

Actions which fulfil the benefit and the costliness conditions but rest on a misunderstanding of the situation on the part of the agent should also be seen as accidental and, thus, not be classified as altruistic either. In this spirit, Andreoni (1995) separates cases of cooperative behaviour according to two different sources. In public goods experiments, he determines subjects' experimental rewards in accordance with their rank among their fellow subjects (not according to their absolute experimental payoff); he thus distinguishes two causes for cooperation, 'kindness' and 'confusion'. Like Andreoni, we see a need to divide the set of acts which fulfil the behavioural conditions for altruism into those which have the goal to further the welfare of another party and those which do not. Goals are indispensable for any plausible definition of altruism (Peacock *et al*. 2005). Thus, we propose that definitions of altruism contain a *goal condition*:

> **Goal condition (GOAL):** For an act to be altruistic, the agent must have the goal to benefit another party at his own expense.

With the addition of GOAL, we take our leave of behavioural definitions. Instead, we turn, in the following section, to psychological approaches which, traditionally, have incorporated goals into the definition of altruism. Whether a person has the goal of benefiting another party is not a directly observable phenomenon, so to determine whether an act is altruistic requires both inference from actions and context as well as speculation about the agent's goal. This should not be anathema to behavioural researchers because, although their definition makes no mention of goals, they, as we have seen, aim to uncover the motives which cause altruism. Indeed, psychological altruism should be attractive to behavioural social scientists because, whether GOAL is satisfied can, in many cases, be inferred from contextual data with relative ease. One of the achievements of Andreoni's experimental design is that it facilitates inferences regarding the agent's goal. Psychological altruism as it is usually formulated, however, demands that, for an act to be altruistic, the

[3] See Wilson's (2002) concept of 'accidental altruism'. Shine *et al*. (2002) use the term in order to describe the behaviour of pit vipers in China.

agent must have the 'ultimate goal' of benefiting others. In the following section, we enquire into the correct specification of psychological altruism.

2. PSYCHOLOGICAL ALTRUISM

The term 'psychological altruism' is ambiguous. On the one hand, it refers to a *theory of motivation*; on the other, it designates a *type of motivation*. The *theory* claims that psychological altruism as a type of motivation is possible. Nagel's *The Possibility of Altruism* is emblematic of psychological altruism (theory). Psychological egoism (theory), in contrast, declares that all actions are egoistic, that psychological altruism (type) cannot occur and that psychological altruism (theory) is wrong. Psychological altruism and psychological egoism make competing claims about intrinsic (ultimate, non-instrumental) goals. While psychological egoists do not deny that people sometimes take pains to help each other, that they share, do each other favours and are considerate of others' interests, they allege that we benefit others only if we expect or, at least, hope to profit thereby. Psychological altruism, in contrast, asserts that humans can and sometimes do act with the 'ultimate goal' of benefiting others at their own expense. A goal is 'ultimate' if it serves no further end but is desired for its own sake; it is thus a *non-instrumental* goal (cf. Batson, 1991: 64–65). For psychological altruists, it is therefore possible that the agent's willingness to bear costs in benefiting others is not motivated by the expectation of receiving a (more than) compensating gain later.

In what follows, we embrace the idea of the ultimate goal criterion but do not avail ourselves of the term. Not only is the concept of ultimate goal a difficult one (see remark (ii), below); it can moreover be replaced by the three conditions with which we define psychological altruism. The first consists in the goal condition (GOAL) we gave at the end of the previous section. The second and third conditions formulate two requirements regarding the reasonableness of the agent's expectations.

The importance of the first reasonableness requirement is illustrated by the following example. Prior to your university examination, I present you with a lucky mascot which, I believe, will improve your mark. Hence the costliness condition of behavioural altruism and GOAL are fulfilled. Nevertheless, we hesitate to describe the act as altruistic because we (and many others) do not regard giving the mascot as a 'serious attempt' to benefit another (however much I might protest that the mascot has improved my grades in numerous exams). This raises the question whether an attempt to be altruistic is to be deemed 'non-serious' if it is based on a false belief about the efficacy of the attempt to bring about the desired benefits for others. Without examining the issue in detail here, we hold the foregoing suggestion to be too sweeping. Instead, we call an attempt 'serious' if we can ascribe to the agent sufficiently good

reasons to believe her action to have its desired effect, even if this belief transpires to be false. An attempt is 'serious' if we can ascribe to the agent sufficiently good reasons to expect her action to have its desired effect. We thus introduce a *reasonable expectation to benefit condition* (REB) for altruism.

> **Reasonable expectation to benefit condition (REB)**: An action, Φ, is altruistic if and only if we can ascribe to the agent sufficiently good reasons for believing that Φ will benefit others.

This condition replaces and refines the benefit condition of behavioural altruism.

This brings us to the second reasonableness requirement. The altruistic status of an action is not compromised if it *unexpectedly* brings its perpetrator net benefits (recalling that benefits do not include internal, hedonic rewards). For instance, after helping someone escape from a house fire at risk to himself, a person, *B*, contrary to expectation, is financially rewarded by the person she rescued; the reward is so munificent that it outweighs the risk and effort *B* bore by the rescue, and consequently the costliness condition is no longer satisfied. The non-fulfilment of this condition should not, however, vitiate the claim that *B*'s act was altruistic because, *ex hypothesi*, *B* could not have speculated on receiving the reward which transpired to make the act profitable to her. We can thus conceive cases in which the costliness condition of the behavioural definition does not apply but the act is nevertheless altruistic. We therefore propose a *reasonable expectation of cost condition* (REC) to replace the costliness condition.

> **Reasonable expectation of cost condition (REC)**: An action, Φ, is altruistic if and only if the agent reasonably expects to bear net costs from performing Φ.

REC allows the reach of altruism to extend to cases like that just enumerated because it allows the agent to derive net benefits from the altruistic act if receiving those benefits is not the goal of *B*'s action.

We are now in a position to define psychological altruism:

> **Psychological altruism:** An action Φ is altruistic in the psychological sense if and only if it satisfies the reasonable expectation to benefit condition (REB), the reasonable expectation of cost (REC) condition and the goal condition (GOAL).

To elucidate this definition, we offer four remarks, the second of which gives us occasion to reformulate the above definition of psychological altruism. (We give this reformulation of psychological altruism after our first two remarks, before turning to the third and fourth remarks.)

Remark (i): To clarify the nature and importance of GOAL, let us consider an example in which both REB and REC, but not GOAL,

are fulfilled. A person, *P*, buys tickets for her country's National Lottery. *P* knows that the proceeds of the Lottery redound to charitable organisations. *P* also knows that she has only a *very* slim chance of winning a considerable amount of money; the mathematical expectation of winning is so slight that *P* knows she should spare her money, but she plays nevertheless. What if *P* were to characterize her action as 'altruistic' for the reason that, in all likelihood, the money she pays for her tickets will end up in the coffers of a charity (thus fulfilling REB and REC)? The ascription of altruism here is inappropriate, and what makes it so is *P*'s prospect of winning a prize (perhaps a very large one). That is, although P's chances of winning are very small, if she does not have the goal of benefiting others by purchasing lottery tickets, we may not say that she acts altruistically (if she did have this goal, why does she not give directly to charity?). In other words, her goal, in buying lottery tickets, is to win (setting aside very exceptional circumstances in which purchasing a lottery ticket is the best way to give to charity), and so *P*'s action does not (typically) fulfil GOAL. That *P*'s goal is not plausibly depicted as benefiting others might be betrayed by the fact that she would surely be elated if, contrary to her expectation, she were to win and, instead of helping charities, she were to withdraw, as winnings, more money from the lottery than she put in.

Remark (ii): With this remark, we call into question the usefulness of the concept 'ultimate goal'. Most straightforward psychological definitions require that, to be altruistic, an action be motivated by the ultimate goal of benefiting another party. We see a need to revise this formulation because the question of whether an other-regarding goal is 'ultimate' is extremely difficult to decide. Elliott Sober and David Sloan Wilson (1998: 260–274) have argued that even Daniel Batson's ingenious experiments cannot prove the existence of actions that *ultimately* pursue other-regarding goals: 'It is easy to invent egoistic explanations for even the most harrowing acts of self-sacrifice' (Sober & Wilson 2000: 198). Moreover, the definite article which qualifies 'ultimate goal' (in the singular form) implies that actions can be motivated by one and only one ultimate goal. Actions, however, can be performed in pursuit of multiple goals. In what follows, we examine multiple goal cases and their implications for altruism.

To commence our discussion, we offer the following example: if I donate blood non-commercially, my blood is screened and I will be informed whether it has deficiencies or diseases; obtaining this information is one reason for donating; but I also have the other-regarding goal of helping anonymous others with my donation. We may say of my donation that it is motivated by two goals, one self-regarding, one other-regarding. How does this affect the status of the action? The answer depends on the details of the case.

Consider a case in which the goal of helping others is non-instrumental (and therefore 'ultimate'). This goal is not, however, sufficient to motivate the action. That is, the other-regarding goal does not meet the following *simple sufficiency condition*:

> **Simple sufficiency condition:** In multiple goal cases, an action, Φ, is altruistic if, in addition to fulfilling REB and REC, the other-regarding goal(s) is (are) sufficient to motivate Φ.

Applied to the example above, the sufficiency condition requires that the other-regarding motivation be sufficient to bring about my donation, even in the absence of any 'external' reward (e.g. having my blood screened).

The simple sufficiency condition leaves room for cases of over-determination in which the self-regarding and other-regarding goals would *each* (alone) be sufficient to motivate the action. Some people may find this too weak because it allows actions to be classed as altruistic when the agent has sufficient self-regarding grounds for performing them; that is, some actions which would be performed even in the absence of the other-regarding goal would be classed as altruistic. Consider again the blood donation example: my desire to have my blood screened could be so great that I would donate blood, even if I were indifferent to the plight of others; the fact that I am *not* indifferent is not necessary for me to perform the action. This reservation inspires what we call the *strong sufficiency condition*:

> **Strong sufficiency condition**: In multiple goal cases, an action, Φ, is altruistic if, in addition to fulfilling REB and REC, the other-regarding goal(s) is (are) sufficient to motivate Φ and the self-regarding goal(s) is (are) insufficient to motivate Φ.

We favour the simple sufficiency condition for the reason that, in cases of over-determination, the other-regarding goal is, *in fact*, present *and* sufficient to trigger the action. A counterfactual reflection to the effect that the agent would have been sufficiently motivated in the absence of the other-regarding goal should not count against classifying the action as altruistic.

Both versions of the sufficiency condition exclude actions regarding which the other-regarding goal has inadequate motivational force. This, however, could be deemed too strict, as the following example shows. A friend who lives abroad asks you for help in a delicate family affair; helping would involve considerable costs, e.g. taking a week off work, travelling abroad and becoming entangled in familial strife. You are, in principle, willing to help your friend whenever possible, but on this occasion you consider the costs of helping to be prohibitively high. On second thoughts, though, you notice that helping your friend would

harbour benefits to you – improving your negotiating skills, meeting interesting people and spending time in a country which you had hitherto not visited. These benefits would not, in themselves, motivate you to travel abroad, but they lower the expected net costs of helping your friend such that you are indeed motivated to travel. Here, the self-regarding and the other-regarding goals *together* are necessary to bring about your helping, although neither, alone, is sufficient. For such cases, we introduce the concept of *degrees of altruism*: if *both* self-regarding and other-regarding goals together are necessary to motivate an action, then the agent has a lower degree of altruism the higher the 'self-regarding top-up' she requires if she is to be motivated to perform the action. This inverse relation between degree of altruism and magnitude of self-regarding top-up may also be expressed as follows: an agent's degree of altruism regarding an altruistic action, Φ, is high if the agent is willing to bear high net expected opportunity costs in order to benefit others; and her degree of altruism with respect to Φ is low if the agent is willing to bear only low net expected opportunity costs. (We reiterate parenthetically here, and in line with REC, that the overall expected costs to the agent must outweigh the expected benefits (excluding hedonic benefits) if Φ is to be altruistic.)

To capture the above thought, we introduce a *necessity condition* for such cases if the action in question is to be altruistic:

> **Necessity condition:** In multiple goal cases, an action, Φ, is altruistic if, in addition to fulfilling REB and REC, it is motivated by both other-regarding and self-regarding goals, both of which are necessary and together sufficient to motivate Φ. In such cases, the agent's *degree of altruism* is inversely related to the magnitude of the self-regarding top-up which, in addition to the agent's other-regarding motivation, is required for the performance of Φ.

An altruist, *A*, whose altruism with regard to Φ is of high degree is more likely to fulfil the simple sufficiency condition (whereby the goal of helping others is sufficient to motivate *A* to Φ and hence no self-regarding top-up is required). Altruist *B*, on the other hand, whose degree of altruism vis-à-vis Φ is low, is more likely to satisfy the necessity condition (whereby an egoistic goal is necessary to supplement the altruistic goal if *B* is to be motivated to Φ). Some economists argue that the degree of altruism of ordinary human beings is generally relatively low (e.g. Kirchgässner 1992). Typically, one's degree of altruism is higher with respect to relatives or friends and lower with respect to strangers, but it is likely to be context-dependent and, to some extent, unpredictable. Oskar Schindler might be a case in point. Apparently, he led a rather selfish life before he displayed extraordinary altruism in rescuing hundreds of Jews from death in the Shoa.

What arises from this discussion is that, in cases of multiple goals, there are two different scenarios in which an action, Φ, can be altruistic:

> **Scenario (1)**: The other-regarding goal(s) *alone* is (are) sufficient to motivate Φ.
> **Scenario (2)**: Both (other-regarding and self-regarding) goals *together* are necessary and sufficient, but neither goal alone is sufficient, to motivate Φ.

We are now in a position to give a *specified definition of psychological altruism:*

> **Psychological altruism (specified)**: An action Φ is altruistic in the psychological sense if, in addition to fulfilling REB and REC, the goal of benefiting another party is either alone sufficient to motivate Φ (scenario 1), or is, together with a self-regarding goal, necessary and sufficient to motivate Φ (scenario 2).

Remark (iii): A sceptical suspicion holds that an action can be reasonably presumed to be egoistically motivated if it brings benefits to the agent. Batson (1991: 64–65) apparently wishes to rebut this suspicion in his defence of psychological altruism. He allows an altruistic action to yield benefits to the altruist but only if they are 'unintended consequences' of her action. If the benefits to the agent are 'intended', then, on Batson's conceptualization, benefiting another person cannot be the agent's ultimate goal (cf. Batson and Shaw, 1991: 109). (Having the ultimate goal of benefiting another person and intending benefits to oneself, the altruist, are antipodes for Batson.) We disagree with Batson's formulation. When an agent pursues other-regarding and self-regarding goals simultaneously, the agent's intention to benefit herself by performing action Φ does not exclude Φ from being altruistic. Instead of imposing Batson's 'unintended consequences' condition on the benefits to the agent, we impose the reasonable expectation of cost condition (REC). That is, in multiple goal cases, for which an other-regarding goal must be present, the agent must expect (but not necessarily intend) his action to carry net costs. Even if you decide to help your friend in the family affair, and thereby fully intend to reap benefits from making the requested trip, we still have a case of altruism. Your degree of altruism may be low if, in order partially to compensate for the loss of time, money and effort, the benefits of making the trip must be substantial, but nevertheless your action is altruistic. A difference between Batson's and our approach to altruism is that, whereas we impose a costliness condition (in the form of REC) on altruistic actions, he imposes none; altruism may, according to Batson, be costly but it does not have to be so (see Batson and Shaw, 1991: 109).

To be sure, there are cases in which an observer is hard-pushed to answer the question whether an action is *really* altruistic. Kant once remarked that questions of motivation are even difficult to judge for the agent herself since introspection is unreliable in this regard. One can never be certain, according to Kant, that one has a genuine concern for others; our noble ends may be mere pretension, and in truth, we 'secretly' or 'unconsciously' strive for self-benefits all along. We concede this point regarding the 'opacity of human motivation' but doubt that this gives us reason to favour psychological egoism over psychological altruism. The motives of a person can be nobler or less noble than they appear. Assessing the motivation of an agent is, to a large degree, a normative matter rather than one of empirical investigation. Whether one is in the habit of ascribing ulterior motives to an apparent altruist depends partly on social conventions regarding the appropriate exercise and expression of cynicism or benevolence regarding the motives of others. Some people or some cultures might be more cynical about the claims of altruism than others, and the determinants of cynicism, and not just the allegedly ulterior motives of the altruist, are worthy of investigation.[4]

Remark (iv): The self-referential desire of experiencing pleasure and avoiding pain plays a crucial role for psychological egoism insofar as they are supposed to explain why seemingly altruistic actions are, in fact, selfish. We call this the *hedonistic challenge to psychological altruism*. Some authors conjecture that recent neuroscientific research supports psychological egoism since it shows that altruistic acts (in the behavioural sense) are connected with hedonic rewards (de Quervain *et al*. 2004: 1257). Our response to the hedonic challenge consists in an Aristotelian consideration to the effect that an agent's taking pleasure in benefiting others, far from showing that her actions are not really altruistic at all, is a sign of a true altruist, that is, of someone who is disposed to promote welfare of others at his own expense from a firm character disposition.[5] The time-honoured problem whether a person enjoys altruistic acts because she has other-regarding goals or whether she has other-regarding goals because she enjoys altruistic acts, is irrelevant for our understanding of altruism. Although our approach does not hinge on this point, we presume that Bishop Butler's rebuttal of psychological egoism was essentially correct. If a benefactor expects to gain a hedonic reward (a

[4] In his *Autobiography*, John Stuart Mill complains about 'the low moral tone of what, in England, is called society; the habit, of not indeed professing, but taking for granted in every mode of implication, that conduct is of course always directed towards low and petty objects (. . .)' (Mill 1981/1873: 61) Mill's point is that the society of his age and nation, by taking egoism for granted as a matter of social convention, *causes* 'the absence of things of an unselfish kind'.

[5] We are aware that pleasure, for Aristotle, is not simply a hedonic phenomenon which is one of the reasons we have called our argument against the hedonic challenge 'Aristotelian'.

	Conditions for Altruism				
	Costliness	Benefit	REB	REC	GOAL
Behavioural altruism	√	√			
Psychological altruism			√	√	√

TABLE 1. Comparison of behavioural and psychological altruism.

'warm glow') from fulfilling an other-regarding desire, it is because, to paraphrase Butler, his passions are directed to external things (the well-being of others) and not to the pleasure which he derives therefrom (Butler 1983/1726: 417). If the benefactor did not value his beneficiary's flourishing independently of the pleasure he gets from contributing to that flourishing, he would not feel a warm glow from his contributions. Hence, hedonic rewards connected to benefiting others seem to supervene on the goal of benefiting others, and their accrual to the agent is compatible with classifying his action as altruistic.[6]

Before we proceed to discuss the normative content of altruism, we present Table 1 in which we compare behavioural and psychological altruism according to the different conditions each poses on altruism.

3. ETHICAL ALTRUISM

Thus far we have considered 'positive' or 'descriptive' definitions, which leave the moral status of altruism open. This contrasts with a usage that seems to prevail in everyday speech, namely that 'altruism' expresses a moral pro-attitude of the speaker. Surprisingly, this aspect of the term has escaped closer inspection in the existing philosophical literature. 'Altruism', however, is often used as a 'thick concept', that is, a concept which 'express(es) a union of fact and value' (Williams, 1985: 129) or an 'entanglement' of the two (Putnam, 2002: 34). An example of a thick concept is the term 'cruelty' which combines descriptive content (of the types of action which constitute instances of cruelty) and normative content (namely a negative attitude towards the perpetrator of cruelty). 'Altruism', too, frequently refuses to respect the fact/value distinction and therefore we introduce a further requirement that captures the evaluative aspect of the concept.

That altruism frequently functions as a thick concept explains why the usage of the word can contain information about the speaker's moral outlook. Calling a man who helped Nazi war criminals escape prosecution

[6] See Brunero (2002) for a recent argument against the hedonistic challenge.

an altruist would in many contexts suggest that the speaker finds such help morally in order. If the speaker were to insist that her characterization is purely descriptive and not intended to express moral approval, it would seem, in many contexts, appropriate to criticize the use of the word 'altruism'. The reason for this is that, in common usage, 'altruism' is not value-neutral. To capture this evaluative dimension, we introduce the notion of *ethical altruism*.

A short excursus on the history of the concept is in order here. The word 'altruisme' was coined by the French sociologist and philosopher Auguste Comte in his *Cours de Philosophie Positive*, in which he advocated radical selflessness as an absolute ethical ideal. In the words of George Lewes, a follower of Comte whose use of the word 'altruism' is the earliest in the English language: '*To live for others* is thus the natural conclusion of all Positive Morality' (1875/1853: 222). Roughly put, Comte and his Positivist followers used the concept to propagate a moral ideal which they considered to be necessary in order to create a perfect society. This origin of the concept, it seems, has left its mark on contemporary usage.

Selflessly helping persons, like Mother Theresa and rescuers of Jews during the Shoa (see Monroe *et al.* 1990), are paradigmatic cases of altruism. Borrowing Urmson's phrase, altruism seems to be situated in the 'higher flights of morality' (Urmson 1958: 211, 215). Few people, however, would argue that selflessness as such is a valuable character trait or that selfless actions are always laudable, for the moral worth of selflessness depends on the moral worth of the goals that an agent selflessly promotes.[7] The moral status of selfless acts hinges on factors that have to be specified by moral background convictions.

In the following, we endeavour to specify the normative component of ethical altruism in a way that captures the normative teint of common usage. We assume that moral common sense distinguishes between actions that are morally eligible (permissible), compulsory (obligatory) and laudable (supererogatory). All supererogatory and all obligatory actions are permissible, but not all permissible actions are supererogatory or obligatory. No obligatory action is supererogatory and

[7] Moreover, autonomy has moral worth which should not be sacrificed in the name of 'living for others'. Mill made this point in a discussion of Comte (Mill 1969/1865). He argues that altruism has moral worth only if the agent's desire to 'live for others' is formed autonomously, i.e. not as a result of adaptation to the pressures of power, public opinion or social conventions. This view resonates in some contemporary discussions. Jean Hampton, for instance, argues that selflessness on the part of women is morally bad if it is conducive to forms of domestic exploitation, keeps them in inferior social positions and, thereby, reinforces power structures in society that are not desirable from a moral point of view (Hampton 1993). Modern usage is arguably closer to Mill's than to Comte's account since selflessness is not considered to be an absolute ideal.

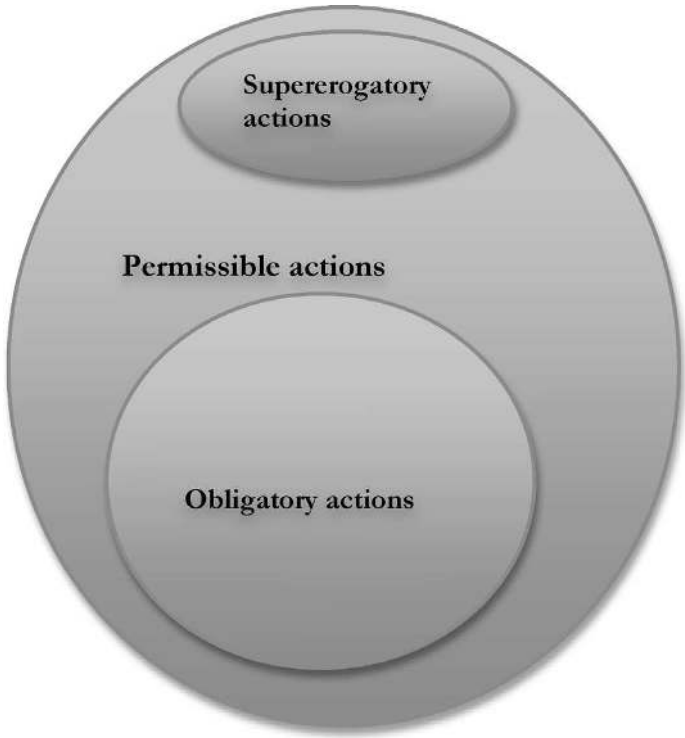

FIGURE 1. The relationships between permissible, obligatory and supererogatory actions.

no supererogatory action is obligatory (see Figure 1).[8] Correspondingly, moral pro-attitudes can express the proponent's conviction that an action is permissible, obligatory or supererogatory.

A proponent has a moral pro-attitude regarding an action, Φ, if he thinks that Φ is

- morally permissible or
- morally obligatory or
- supererogatory ('beyond the call of duty').

The attribution of moral permissibility is the weakest form of expressing a moral pro-attitude.

[8] The class of permissible actions contains all actions which are not forbidden (see Raz 1975). For an understanding of permissible in the sense of 'neither obligatory nor forbidden' see Urmson (1958: 198).

To make a 'first stab' at defining the concept of ethical altruism, we start with permissibility as a normative criterion:

> **First stab**: An action, Φ, is altruistic in the ethical sense if and only if Φ fulfils the criteria for psychological altruism (*descriptive component*), and Φ is morally permissible (according to a reasonable moral outlook) (*normative component*).

One remark regarding *First stab* is required: since there are reasonable disagreements regarding the moral permissibility of actions, there is room for reasonable disagreement as to whether an action fulfils the normative component and qualifies as altruistic in the ethical sense. With *First stab* (and the two subsequent stabs below), we do not endeavour to delineate a unique and universally valid set of actions which qualify as ethically altruistic. Those of differing moral outlooks will disagree on what constitutes a permissible action. We turn now to two criticisms of *First stab*.

Criticism (i): Consider the following example from an act utilitarian perspective. A father sacrifices his life in order to save that of his child. In saving his child, he thereby foregoes the option of saving five other children (all unrelated to him). In letting the five children die, the father, according to the act utilitarian, acts morally wrongly because saving one child does not maximize social utility. The act utilitarian can nevertheless plausibly insist that the father's saving of his child is altruistic.[9] This seems to contradict *First stab*. The latter can be defended against this objection because if the act utilitarian considers the action to be altruistic *and* morally wrong, she does not express a moral pro-attitude but conceives altruism in a descriptive, not an ethical, sense. The objection therefore misses its target, for *First stab* states that impermissible actions cannot be altruistic *in the ethical sense*; it does not state that impermissible actions cannot be altruistic in the descriptive sense.

Criticism (ii): A more serious problem for *First stab* is that it seems to be too inclusive. As we stated above, all obligatory actions are permissible. Thus, *First stab*'s normative component must include all obligatory acts as morally permissible. But there is reason to exclude obligatory actions from the set of (ethically) altruistic actions, as the following example makes clear. An agent, *A*, enters into a contract with another, *B*, in which *A* promises to Φ at time *t*. *A* expects to benefit from the mutual execution of the contract. Between making and fulfilling his contractual obligation, however, circumstances change in a way *A* had not anticipated. Consequently at time *t*, Φ-ing no longer redounds to *A*'s benefit. Nevertheless, *A* keeps his contractual word and Φ-s at *t*, even though he reasonably expects that he could, with impunity, have broken his promise. Consequently, *A* bears net costs (while *B* benefits). The conditions of psychological altruism are hereby fulfilled. It would conflict

[9] We thank one of the referees for suggesting this point.

with standard usage, however, to call the fulfilment of a contractual obligation altruistic, because altruism is not *owed* to others.[10] We therefore hold *First stab* to be an inadequate formulation of ethical altruism and pass to a revised version which we call *Second stab*.

> **Second stab**: An action, Φ, is altruistic in the ethical sense if and only if Φ fulfils the criteria for psychological altruism, and Φ is permissible but not obligatory (according to a reasonable moral outlook).

Unlike its predecessor, *Second stab* excludes all obligatory actions under its penumbra of permissibility. But like its predecessor, it includes all supererogatory actions (for, as we stated above, all supererogatory actions are permissible, although the converse is not true). In contrast to the inclusion of obligatory actions, the inclusion of supererogatory acts in the normative component appears to be unproblematic.

But before we accept *Second stab*, we wish to investigate a third possibility, which holds out the prospect of being more precise than its predecessor.

> **Third stab**: An action, Φ, is altruistic in the ethical sense if and only if Φ fulfils the criteria for psychological altruism, and Φ is supererogatory (according to a reasonable moral outlook).

Third stab is more precise than Second stab in the sense that the set of supererogatory actions is a subset of the actions, which count as altruistic under *Second stab*. In the following, we examine three criticisms of **Third stab**.

Criticism (i): The first notes that some reasonable ethical theories leave no room for supererogation. Certain versions of Kantianism and act utilitarianism are cases in point. If a Kantian or act utilitarian who adheres to such a version were to accept *Third stab*, she would be committed to denying the possibility of ethical altruism because a proponent of such a theory denies the existence of supererogatory actions. In defence of *Third stab*, one may argue that this is no reason to revise the definition since there are in fact people who deny the existence of altruism. Think of those who decline the thanks of others because 'they have just done their duty'. Perhaps these people agree with Kant that an action is either in accordance with moral duty (and thus obligatory) or has no moral worth. It would certainly be unfortunate if *Third stab* implied the non-existence of ethical altruism for all or almost all reasonable moral theories. This, however, does not seem to be the case, as we now argue for two widely held moral theories, both of which may be held, by some, to exclude the

[10] If someone were to insist that the fulfilment of an unenforceable contractual obligation is altruistic in the ethical sense, one may press the question what is meant by 'obligation' here. If one assumes that the fulfilment of an unenforceable obligation is a mere pleasantry, we see no objection to our point. For a mere pleasantry is not obligatory.

possibility of ethical altruism. The theories in question are Kantianism and act utilitarianism.

In the *Groundwork*, Kant remarks that a perfect duty is one that admits of no exception in favour of inclination. An imperfect duty, by implication, permits the agent, to a certain degree, to do what he is inclined to do without regard to the demands of morality. Thomas Hill argues that Kant's notion of imperfect duties leaves room for discretionary judgement and that an agent uses the 'moral latitude' in a commendable manner if he does more than absolutely required (Hill 1992/1971). The rationale for this discretion is to be found in Kant's derivation of imperfect duties. The universalizability test reveals that a maxim never to help a fellow-creature (maxim of non-beneficence) cannot reasonably be willed; it is thus morally forbidden to be indifferent to the needs of others. This, though, leaves room for moral judgement as to the *amount* of help one finds appropriate on particular occasions. As Henry Allison (1996/1993: 166) points out, if one rejects the maxim of non-beneficence because it is non-universalizable (and therefore morally forbidden), morality does not require that one adopt a maxim of beneficence in the sense of making a concern for the welfare of others a (or *the*) central project in one's life. Following Allison's interpretation of Kant, it is permissible to 'acknowledge a requirement to act beneficently, but to construe this as something to be gotten out of the way, to be discharged as painlessly as possible, so as to be able to get on to one's real projects' (Allison 1996/1993: 166). Thus, if one has a less than full-blooded concern for others and attends to one's own projects after fulfilling one's modestly determined duty to benefit others, one acts in a morally permissible way. If, despite having fulfilled one's duty of beneficence by helping others in a modest way, one continues to help and thereby exceeds this duty, one is acting in a commendable manner. As an illustration, one may think of a Kantian who, convinced that world poverty is the most urgent moral problem of our time, arrives at the conclusion, after careful moral deliberation, that he must help the world's poor. Regarding the extent of his obligation, he concludes that his share of helping must be such that world hunger could be ended if everyone else, who is in a similar position, were to deliver the same share of help. Let us assume that this share would amount to a percentage, p, of his income; the Kantian would fulfil his imperfect moral duty if he were to donate p but he would act 'beyond the call of duty' if he were to give more. Since helping the world's poorest by donating p is obligatory, giving more than p is *ceteris paribus* commendable because it goes beyond the obligation to give p.[11] Giving more than p is therefore be supererogatory and thus altruistic.

[11] We add the *ceteris paribus* clause here because it is conceivable that helpers are 'doing too much' by creating dependency and passivity or rent-seeking behaviour on the part of the

We have addressed this point in order to put into context the objection that *Third stab* implies the impossibility of ethical altruism for certain moral theories. The foregoing depiction of Kantianism shows that this moral doctrine is not irremediably wedded to the impossibility of ethical altruism (in the sense of *Third stab*). Some, but not all, Kantians would nevertheless be committed to claim that ethical altruism does not exist, but this is by no means fatal, for some, but certainly not all psychologists deny the possibility of psychological altruism as well.

Another moral theory that instils scepticism about supererogatory actions is act utilitarianism. For act utilitarianism, if an action maximizes social utility, it is the morally right action and is obligatory. Consequently, if the maximizing action happens to be 'altruistic', it, too, must be obligatory, but, being obligatory, it cannot be supererogatory. With the exception of the case of single-level act utilitarianism, however, there is arguably space for actions 'beyond the call of duty' in act utilitarianism. Let us take John Stuart Mill as an example. There is much disagreement as to whether Mill was a rule or an act utilitarian. But even those who read him as an act utilitarian concede that rules play a crucial role in his account of morality. Mill argued that we are morally obliged to respect justified social rules as long as they do not conflict with each other.[12] These rules create the space for permissible actions. Some elements of the set of permissible actions, however, are better than others.[13] A permissible action α is better than a permissible action β if α tends to produce more happiness than β. If an agent decides to perform β instead of α, he does something laudable from a moral point of view; he aggrandizes the amount of good in the world more than he is morally required to do. This seems to be sufficiently close to what a Kantian would say about the 'over-fulfilment' of an imperfect duty. Arguably, something along these lines is what most people have in mind most of the time when they call an action altruistic. They express the view that an agent did more (good) than could be justifiably exacted from him and that this is commendable from a moral point of view.[14] In a nutshell, *Third stab* holds out against criticism (i).

beneficiary. Here, though, we shall assume that the benefactor's help is, indeed, helpful for the beneficiary.

[12] 'We must remember that only in these cases of conflict between secondary principles is it requisite that first principles should be appealed to. There is no case of moral obligation in which some secondary principle is not involved (. . .)' (Mill 1969/1861: 225–226).

[13] According to Mill, actions are 'right in proportion as they tend to promote happiness' (Mill 1969/1861: 210).

[14] 'Duty is a thing which may be *exacted* from a person, as one exacts a debt. Unless we think that it might be exacted from him, we do not call it his duty' (Mill 1969/1861: 246).

Criticism (ii): Let us consider a variation of an earlier example and assume that a father is on a sinking ship. He is confronted with the following three options:

– sacrificing his own life and save his child (option 1)
– sacrificing his own life and save five unrelated children (option 2)
– saving his own life (option 3).

In contrast to option (1) and (2), option (3) does not fulfil the criteria of psychological altruism. Following most moral theories, however, it would be permissible for the father to save his own life, for self-sacrifice cannot be justifiably exacted from the father. In contrast, preferring *either* option (1) to option (3) *or* option (2) to option (3), would be praiseworthy: it would be supererogatory to sacrifice oneself in order to save others. According to *Third stab*, both option (1) and option (2) are altruistic in the ethical sense. However, many people would deny that preferring option (2) to option (1) is praiseworthy. That is, our moral intuition is that even if saving five unrelated children and letting one's own child die is permissible, it is his *own* child the father should save. If one follows *Third stab*, people who share the aforementioned moral intuition would deem option (2) altruistic (in the ethical sense) when compared to option (3); but in view of all three options, they would claim that option (2) is not laudable and hence not supererogatory; option (2) could not therefore be a case of ethical altruism.

Does this challenge the plausibility of *Third stab*? One may argue that the status of an action (in this case, an altruistic one) should be *robust* in the sense that it is invariant to the introduction of additional options. According to *Second stab*, both option (1) and option (2) are altruistic (since they both fulfil the descriptive and the normative conditions) in view of all three options. If robustness is a desirable property of a definition of ethical altruism, *Second stab* has an edge over *Third*.

Criticism (iii): Consider now the case of a person who always helps other people (in a way that fulfils the conditions of *Second stab*) but is barred thereby from completing a novel which, were it completed, would produce more utility than the person's helpfulness. From a utilitarian point of view, completing the novel would be better than helping others. If we add the assumption that the novelist's own gain would be greater if he wrote the novel rather than helped others, we see, in the example, the germ of the thesis that the pursuit of self-interest leads to the greatest social good. A utilitarian in the tradition of Mill who adheres to *Second stab* could say of cases like this that, while it is permissible, it is not always commendable to be altruistic. *Third stab*, however, excludes such a statement by definition. If helping other people is not commendable (supererogatory), it cannot be altruistic for conceptual reasons. In view of the fact that there is a time-honoured tradition, which argues that altruistic

	Conditions for Altruism					
	Costliness	Benefit	REB	REB	GOAL	Normative component
Behavioural altruism	√	√				
Psychological altruism			√	√	√	
Ethical altruism			√	√	√	√

TABLE 2. Comparison of behavioural, psychological and ethical altruism.

actions do not promote as much good as selfish actions in many social arenas, this implication is unsatisfactory. Whether and when altruism must be more highly morally valued than egoism or vice versa seems to be a matter that must be settled theoretically and empirically, not by definition.

Criticisms (ii) and (iii) have arguably sufficient traction to decide the controversy between *Second stab* and *Third stab* in favour of the former.

CONCLUSION

We have analysed three definitions of altruism and have asked, for each, whether it includes or excludes particular cases as examples of altruism. When a definition's inclusion or exclusion of such cases has not corresponded to our engrained intuitions regarding what is (or is not) a case of altruism, we have refined the definition. In this way, we have reached reflective equilibrium and found that altruism should be defined according to a positive component (given by psychological altruism) and a normative component for which we have argued in favour of *Second stab*. Table 2 augments and completes Table 1 by presenting a comparison of the three types of altruism we have analysed according to the conditions they impose on altruism.

We are aware of the limits of our approach, for ordinary language has not exercised as firm a constraint on definitions of altruism as it does on other academic definitions. This leaves room for disagreement about our understanding of altruism. Furthermore, the existence of different moral views also allows for dissent about the normative component of altruism. Consequently, even those who agree with us that altruism has this evaluative facet might disagree with our conceptualization of it. For those who do disagree with us, we hope, nevertheless, to have addressed aspects of altruism which often go ignored in academic discussion and

thereby to have given stimulus to further debate on this enigmatic concept and phenomenon.

REFERENCES

Allison, H. E. 1996/1993. *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy*. Cambridge: Cambridge University Press.

Andreoni, J. 1995. Cooperation in public-goods experiments: Kindness or confusion? *American Economic Review* 85: 891–904.

Batson, D. 1991. *The Altruism Question: Toward a Social Psychological Answer*. Mahwah, NJ: Lawrence Erlbaum.

Batson, D. and L. Shaw 1991. Evidence for altruism: Towards a pluralism of prosocial motives. *Psychological Inquiry* 2: 107–122.

Brunero, J. 2002. Evolution, altruism and 'internal reward' explanations. *Philosophical Forum* 31: 413–424.

Butler, J. 1983/1726. *Five Sermons*. Indianapolis: Hackett.

Dawkins, R. 1976. *The Selfish Gene*. Oxford: Oxford University Press.

de Quervain, D., U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, and E. Fehr 2004. The neural basis of altruistic punishment. *Science* 305: 1254–1258.

Fehr, E. and S. Gächter 2002. Altruistic punishment in humans. *Nature* 415: 137–140.

Fehr, E. and U. Fischbacher 2003. The nature of human altruism. *Nature* 425: 785–791.

Hampton, J. 1993. Selflessness and the loss of self. *Social Philosophy & Policy* 10: 135–165.

Hill, T. 1992/1971. *Dignity and Practical Reason*. Ithaca: Cornell University Press.

Jacobs, R. 1985. Obligation, supererogation and self-sacrifice. *Philosophy* 62: 96–101.

Kirchgässner, G. 1992. Towards a theory of low-cost decisions. *European Journal of Political Economy* 8: 305–320.

Lewes, G. 1875/1853. *Comte's Philosophy of the Sciences*. London: G. Bell and Sons.

Mill, J. S. 1965/1871. *Principles of Political Economy with some of their Applications to Social Philosophy*. In *Collected Works*, Vol. II, ed. J. M. Robson. Toronto: University of Toronto Press.

Mill, J. S. 1969/1865. *Auguste Comte and Positivism*. In *Collected Works*, Vol. X, ed. J. M. Robson, 261–368. Toronto: University of Toronto Press.

Mill, J. S. 1969/1861. *Utilitarianism*. In *Collected Works*, Vol. X, ed. J. M. Robson, 203–259. Toronto: University of Toronto Press.

Mill, J. S. 1981/1873. *Autobiography*. In *Collected Works*, Vol. I, ed. J. M. Robson, 1–290. Toronto: University of Toronto Press.

Monroe, K., M. Barton and U. Klingelmann. 1990. Altruism and the theory of rational action: Rescuers of Jews in Nazi Europe. *Ethics* 101: 103–122.

Peacock, M., M. Schefczyk and P. Schaber 2005. Altruism and the indispensability of motives. *Analyse and Kritik* 27: 188–196.

Putnam, H. 2002. *The Collapse of the Fact/Value Dichotomy and other Essays*. Cambridge, MA: Harvard University Press.

Rawls, J. 1971/1999. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Raz, J. 1975. Permissions and supererogation. *American Philosophical Quarterly* 12: 161–168.

Sesardic, N. 1995. Recent work on human altruism and evolution. *Ethics* 106: 128–157.

Shine, R., L-X. Sun, M. Fitzgerald and M. Kearney 2002. Accidental altruism in insular pit-vipers (*Gloydius shedaoensis*, Viperidae). *Evolutionary Ecology* 16: 541–548.

Sober, E. and D. S. Wilson 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Sober, E. and D. S. Wilson 2000. Summary of: '*Unto others: the evolution and psychology of unselfish behavior*'. *Journal of Consciousness Studies* 7: 185–206.
Trivers, R. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology* 45: 35–57.
Urmson, J. O. 1958. Saints and heroes. In *Essays in Moral Philosophy*, ed. A. I. Melden, 198–216. Seattle: University of Washington Press.
Williams, B. 1985. *Ethics and the Limits of Philosophy*. London: Fontana.
Wilson, E. 1975 (2000). *Sociobiology: The New Synthesis*. Cambridge, MA: Belknap.
Wilson, J. 2002. The accidental altruist. *Biology and Philosophy* 17: 71–91.