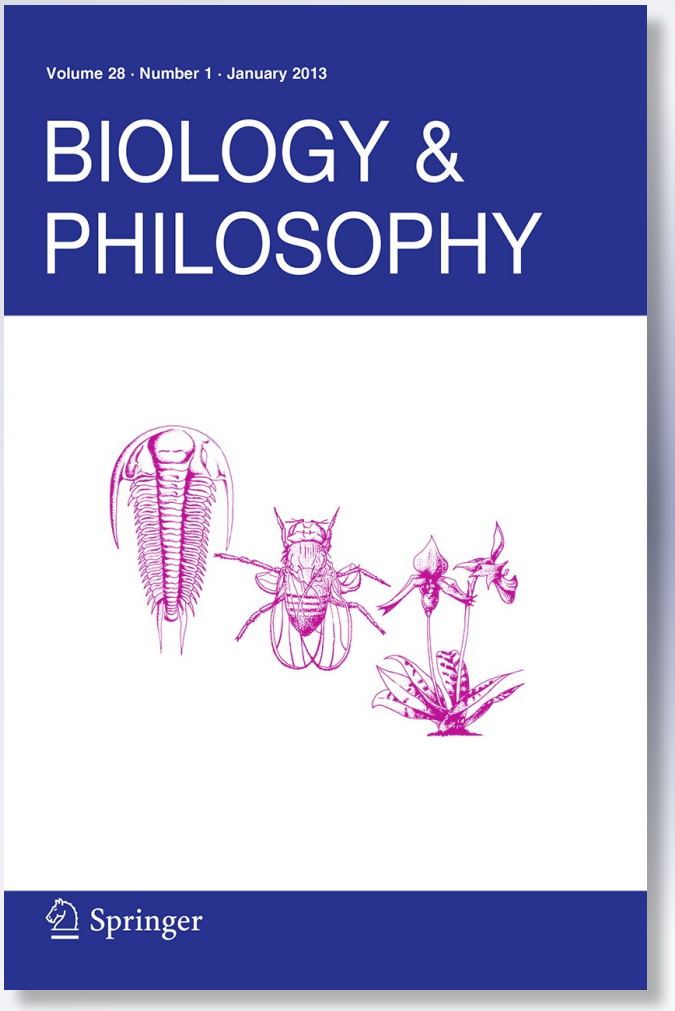
Altruism across disciplines: one word, multiple meanings

Christine Clavien & Michel Chapuisat

Biology & Philosophy

ISSN 0169-3867
Volume 28
Number 1

Biol Philos (2013) 28:125-140
DOI 10.1007/s10539-012-9317-3



Springer

Biol Philos (2013) 28:125–140
DOI 10.1007/s10539-012-9317-3

Altruism across disciplines: one word, multiple meanings

Christine Clavien · Michel Chapuisat

Received: 16 September 2011 / Accepted: 12 March 2012 / Published online: 1 April 2012

Abstract Altruism is a deep and complex phenomenon that is analysed by scholars of various disciplines, including psychology, philosophy, biology, evolutionary anthropology and experimental economics. Much confusion arises in current literature because the term altruism covers variable concepts and processes across disciplines. Here we investigate the sense given to altruism when used in different fields and argumentative contexts. We argue that four distinct but related concepts need to be distinguished: (a) *psychological altruism*, the genuine motivation to improve others' interests and welfare; (b) *reproductive altruism*, which involves increasing others' chances of survival and reproduction at the actor's expense; (c) *behavioural altruism*, which involves bearing some cost in the interest of others; and (d) *preference altruism*, which is a preference for others' interests. We show how this conceptual clarification permits the identification of overstated claims that stem from an imprecise use of terminology. Distinguishing these four types of altruism will help to solve rhetorical conflicts that currently undermine the interdisciplinary debate about human altruism.

Keywords Reproductive altruism · Psychological altruism · Behavioural altruism · Preference altruism · Experimental economics · Evolutionary anthropology

Introduction

In everyday language, altruism occurs when individuals are disposed to sacrifice part of their personal interest in favour of others; it is an honourable gift given

C. Clavien (✉) · M. Chapuisat
Department of Ecology and Evolution, University of Lausanne,
UNIL-Sorge Biophore, 1015 Lausanne, Switzerland
e-mail: christine.clavien@unil.ch

without any expectation of future personal reward. In this common usage, the type of altruism and the content of 'personal interest' are not specified, so that the term may apply to a range of phenomena. Altruism has been the topic of intense research in many academic disciplines, including biology, psychology, philosophy and economics. However, the term has been used in different ways in order to fit the particular research contexts and needs of each discipline. This has generated much confusion, because the differences are subtle and not always made explicit.

Multifarious and somewhat cryptic uses of the term altruism are frequent in emergent research fields on human social behaviour. For example, the experimental economists Fehr and Fischbacher (2003: 785, our emphases), after having explicitly stated that

> Throughout the paper we rely on a *behavioural* – in contrast to a *psychological* – definition of altruism as being costly acts that confer economic benefits on other individuals,

add on the same page that

> A combination of altruistic and selfish concerns *motivates* them [the altruists]. Their altruistic *motives* induce them to cooperate and punish in one-shot interactions and their selfish *motives* induce them to increase rewards and punishment in repeated interactions or when reputation-building is possible.

Here, we are at a loss to understand whether the authors are discussing the outcome or the motivational aspect of altruism, and what the behavioural and psychological categories exactly cover. In another paper, researchers from the same field note that

> The punishment of defectors is an altruistic act in the *biological* sense because, typically, it is costly for the punisher and induces the punished individual to defect less in future interactions with others (de Quervain et al. 2004: 1257, our emphasis).

However, altruism in biology typically refers to behaviours reducing personal reproduction, as observed for example in ants or termites. It is not obvious that punishment of defectors in humans meets this criterion. Thus, one can reasonably ask whether scholars in biology and economics are really discussing the same phenomenon, as suggested by experimental economists (e.g. the above quote; Bowles and Gintis 2011; Gintis et al. 2005).

Here, we propose to distinguish four notions of altruism that are commonly used in cross-disciplinary literature. Sober and Wilson (1998) have already shown that altruism is thematized very differently in psychology and philosophy, as opposed to biology. We will briefly review these uses in "Psychological versus reproductive altruism" section. We then argue that the recent focus on altruism in research fields such as evolutionary anthropology, evolutionary game theory and experimental economics calls for a distinction between two further notions of altruism ("Behavioural and preference altruism" section). To show why this new distinction is important, we outline the conceptual differences between the four forms of altruism ("Some crucial conceptual differences" section), provide some examples of

terminological confusions, and point to overstated claims that are based on an imprecise use of the altruism terminology ("Misleading uses of altruism terminology" section). We conclude that a consistent use of an explicit terminology, such as the one delineated in this article, is crucial for clarifying the numerous confusions that currently undermine cross-disciplinary debates about human altruism. This taxonomy of altruism will help to assess the actual contribution of various research fields to a better understanding of the general phenomenon of unidirectional helping behaviour.

Psychological versus reproductive altruism

Psychologists and philosophers have long been debating the possibility of altruism. The controversy began in the seventeenth and eighteenth centuries and involved British moralists (among them Butler 1991/1726; Hobbes 2005/1651; Hutcheson 2004/1725; Mandeville 1997/1714–1728; Smith 2002/1759).[1] The problem involved deciding whether human beings are exclusively motivated by self-interested concerns or whether they can be moved by genuine concerns for others' needs. A stance on this issue was considered fundamental for the elaboration of a realistic political system adapted to human nature.

Although less politically flavoured nowadays, this controversy is ongoing (Andreoni 1990; Batson 1991; Cabanac et al. 2002; Cialdini et al. 1987; Ghiselin 1974; Nagel 1970; Rand 1964; Sober 1992; Stich et al. 2010; Stocks et al. 2009). The main contemporary actors of the debate are philosophers and psychologists. They usually define altruism in terms of internal motivations responsible for helping actions (Batson 1991: 6). We refer to this as:

Psychological altruism An action is altruistic if it results only from motivations directed towards the goal of improving others' interests and welfare.

In other words, psychological altruism is more about *wanting* a beneficial outcome for others than about *achieving* this outcome. The most important condition for psychological altruism is that no self-directed consideration—such as a quest of pleasure, power or honour, or avoidance of pain—is causally responsible for the action.

The controversy around psychological altruism may never be resolved, because of the difficulty of finding arguments or empirical data showing that an other-directed action *cannot* be pervaded by—possibly unconscious—self-directed motivation (Stich et al. 2010; Stich 2007; Clavien 2012a). A possible way out might consist in reformulating the controversy (Clavien 2012a; Sober and Wilson 1998; Kitcher 2011). Alternatively, following many other unsettled issues, this debate might at some point be consigned to the history of philosophy.[2]

[1] The actual term 'altruism' started to be used around the mid-nineteenth century. Auguste Comte (1851–1854) defined it as the motivation to act benevolently—as opposed to 'egoistic' motives, which are directed towards the agent's self-interests.

[2] Behavioural and brain sciences have recently made important advancements in our understanding of human decision-making (Bargh et al. 2010). Philosophical questions originally formulated on dated views of the cognitive architecture of the mind—in our case, that action results from a single causal chain

Independently of the controversy about the possibility of psychological altruism, evolutionary biologists have also confronted a 'problem of altruism', but of a different kind. A pressing question for them was to explain how extreme forms of helping behaviour that decrease the actor's fitness could evolve (West et al. 2007; Foster 2008; Frank 1998; Grafen 1985; Clavien and Chapuisat 2012). We label this specific form of altruism:

Reproductive altruism A behaviour is altruistic if it increases other organisms' fitness and permanently decreases the actor's own fitness.

Fitness reflects individuals' rate of survival and reproduction. Typical examples of reproductive altruism are found in the social insects. For example, the vast majority of honeybee workers do not reproduce and spend their entire life rearing and defending their queen's offspring.

Since reproductively altruistic behaviours are, by definition, detrimental with respect to survival and reproduction, their persistence in the course of evolution is puzzling and calls for a special explanation. William Hamilton's (1964, 1970) kin selection theory provides such an explanation: genes responsible for altruistic behaviour can spread in a population if the altruistic behaviour they induce is more likely to benefit those who possess copies of the same genes, which is typically—but not exclusively—the case of close relatives, due to common ancestry (Hamilton 1970; Grafen 1985; Queller 1992; West et al. 2007, 2011; Frank 1998). An alternative way to formalize the evolution of reproductive altruism is group selection theory (Okasha 2007; Wilson 1975). Group selection models also show that reproductive altruism can evolve when group members are related. Thus, they make quantitative predictions identical to the kin selection theory.

As Sober and Wilson (1998) pointed out, reproductive altruism should not be confused with psychological altruism. Reproductive altruism is defined in relation with outcomes, independently of the actor's consciousness or subjective motivations. For example, individuals' intentions play a causal role in some species—e.g. humans—but not others—e.g. bacteria—but this difference does not preclude both types of species to show reproductive altruism. In contrast, psychological altruism refers to subjective motivation for individual actions. Interestingly, this partition largely fits with the classical distinction in biology between ultimate and proximate explanations (Mayr 1961; West et al. 2007; Tinbergen 1963). *Ultimate* explanations refer to the adaptive value and fitness consequences of a trait; they answer the question of why a trait has evolved. The Hamiltonian or group selection explanations of the evolution of reproductive altruism are of this kind. *Proximate* explanations refer to the mechanisms causing the behaviour. Individual altruistic *motivation* is such a causal mechanism; it is a proximate cause for helping behaviour. Ultimate and proximate explanations are complementary, but shed light

Footnote 2 continued
starting with one primary motive—might have simply become inadequate or confusing in view of current scientific understanding of the mind.

on different aspects of phenomena.[3] There are a number of implications, the first being that evolutionary explanations of reproductive altruism provide no direct insight into the psychological goals or preferences underlying these behaviours. For example, Hamilton's theory provides a powerful explanation for the reproductive altruistic behaviour of worker bees, irrespective of what these bees might be thinking of while devoting their lives for the good of the hive. Conversely, knowing that a human action is triggered by an altruistic—or selfish—motive does not mean that this behaviour is reproductively altruistic—or selfish.

It should also be noted that many behaviours that appear to be reproductively altruistic in the short term because they bear some immediate cost to the actor do in fact increase the fitness of the actor in the long run (Lehmann and Keller 2006). One example is the sexual cannibalism practised by some mantis and spiders, in which males are eaten by females during or just after mating (Andrade and Banta 2002). Environmental conditions in which food is scarce and males have little chance to mate with several females can favour the evolution of such interactions; males that do not try to escape after copulation are not reproductively altruistic if their submissive behaviour increases the number of offspring they will have. Other more common examples are cases of delayed fitness benefits that may arise through social prestige (Zahavi 1975) or future reciprocity (Trivers 1971). These types of social interactions are common and important in humans. However, when considering their long-term fitness consequences, such actions do not represent true cases of reproductive altruism (West et al. 2007).

Behavioural and preference altruism

More recently, the notion of altruism has been heavily used in a new kind of debate, mostly within the emerging fields of experimental economics and evolutionary anthropology (Gintis et al. 2005). The debate revolves around the question of whether ordinary people behave in the way predicted by a crude view of human agency, according to which humans have only self-regarding preferences. As it is usually described, this view presents human beings as optimal utility maximizers, and utility is reduced to hedonistic goods such as pleasure and money. It predicts that, humans act as selfish agents in all circumstances (Fehr and Camerer 2007; Henrich et al. 2005: 812). This view is often referred to as the "homo economicus" model.[4]

Although it is not easy to identify who really defends the "homo economicus" view,[5] there are many detractors of it in sociology and psychology (Bourdieu 2000;

[3] Interestingly, proximate mechanisms can themselves be accounted for with ultimate explanations (Clavien and Chapuisat 2012; West et al. 2007).

[4] We use quotation marks because the term *homo economicus* is also used for referring to economic theories that might not fit this description because they take no stance on which preferences are contained in human's utility function (Kirchgässner 2008).

[5] Some of Ken Binmore's big claims—e.g. he describes himself as a Hobbesian (2006)—might lead to think that he is an advocate of the "homo economicus" model. However, it should be noted that he does not deny the existence of sympathetic preferences—at least toward closely related individuals (2005: chap. 7).

Gigerenzer 2008/2007; Kahneman et al. 1982), evolutionary anthropology (Henrich et al. 2001), evolutionary game theory (Gintis 2000) as well as in economics (Simon 1996/1969), and the subfield experimental economics (Fehr and Fischbacher 2003; Fehr and Gächter 2002; Fehr and Schmidt 1999). Opponents of the "homo economicus" model maintain that humans are norm-abiding actors who do have preferences for the well-being of others. Such a stance is fully compatible with neoclassical economics, which states that humans' choices reflect the content of a utility function composed of long-standing and hierarchically ranked preferences. Scholars who endorse neoclassicism while rejecting the "homo economicus" view propose the inclusion of social and altruistic preferences in humans' utility functions.

Three major argumentative strategies used against the "homo economicus" model appeal to altruism. Below, we review these strategies. We argue that they rely on three different concepts of altruism: psychological altruism and two novel concepts that deserve their own labels.

The first strategy consists in showing that people do not act in the way predicted by the "homo economicus" model, but instead behave in a fair and altruistic manner. A large number of laboratory and field studies have used economic games to demonstrate that people are ready to anonymously give money to strangers, even while knowing that they have nothing to gain from their generosity—neither in terms of reputation nor of material benefit (Hoffman et al. 1996; Charness and Gneezy 2008; Fehr and Fischbacher 2004a; Fehr and Gächter 2002; Fehr and Fischbacher 2004b). Other studies showed that many people prefer to lose money by punishing free-riding rather than accept the inequality created by this behaviour (de Quervain et al. 2004; Fehr and Fischbacher 2003). This can even happen in an anonymous condition where the punisher is an external observer and not himself victim of the free-riding behaviour (Kurzban et al. 2007; Fehr and Fischbacher 2004b; Lewisch et al. 2011).

The nature of the altruism implicated in these studies is not fully clear. Two novel concepts seem to be used without being clearly disentangled. This is illustrated in the following citations discussing 'strong reciprocity', a paradigmatic example of altruism.

> (Cit. 1) Many of these experiments examine a nexus of *behaviours* that we term strong reciprocity. Strong reciprocity is a *predisposition* to cooperate with others, and to punish (at personal cost, if necessary) those who violate the norms of cooperation, even when it is implausible to expect that these costs will be recovered at a later date. (Gintis et al. 2005: 8, our emphases)
> (Cit. 2) Strong reciprocity is a combination of altruistic rewarding, which is a *predisposition* to reward others for cooperative, norm-abiding behaviours, and altruistic punishment, which is a *propensity* to impose sanctions on others for norm violations. Strong reciprocators bear the cost of rewarding or punishing even if they gain no individual economic benefit whatsoever from their acts. (Fehr and Fischbacher 2003: 785, our emphases)
> (Cit. 3) Altruistic cooperators are *willing* to cooperate, that is, to abide by the implicit agreement, although cheating would be economically beneficial for them (Fehr and Rockenbach 2003: 137, our emphasis)

(Cit. 4) We provide a model of team production in which the *motivation* to punish is strong reciprocity: the *willingness* of some altruistic team members to engage in the costly punishment of shirkers. (Carpenter et al. 2009: 221, our emphases)
(Cit. 5) By now we have substantial evidence suggesting that fairness *motives* affect the behaviour of many people. (Fehr and Schmidt 1999: 817, our emphasis)

Citations 1 and 2 refer to a particular type of 'behaviour' that can be observed in socio-economic contexts; a 'predisposition' or 'propensity' to reward, cooperate, and apply costly punishment to norm violators. This type of description calls for a new way of conceiving altruism, which we label:

Behavioural altruism 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.

This notion of altruism is reminiscent of reproductive altruism. However, it is much less restrictive because it applies to any type of cost and benefit. For example, in laboratory experiments the standard currencies used are money or material goods.

Rather than behavioural propensities, citations 3 to 5 refer to people's goals and motivations; their 'willingness' to cooperate and punish shirkers. This clearly refers to people's subjective preferences for the well-being of others and calls for a new definition of altruism, which we label:

Preference altruism An action is altruistic if it results from preferences for improving others' interests and welfare at some cost to oneself.

This notion is close to psychological altruism. It is however more explicit about the cost and less restrictive regarding the underlying psychological mechanism. For example, preference altruism occurs when an individual shows a reliable preference to help needy persons at some personal cost, although she could ignore their fate. The private and profound reasons for this preference however need not be specified; an empathic emotional reaction, a desire to enforce a social norm, or a demand for internal reward might underlie this preference.

Showing the existence of behavioural (Fehr and Fischbacher 2003) or preference altruism (Fehr and Rockenbach 2003) are both equally efficient ways of challenging the "homo economicus" model, because neither type of behaviour is predicted by the model.

The second argumentative strategy against the "homo economicus" view makes use of evolutionary game theory. An altruistic behaviour observed in the laboratory can be modelled as a behavioural strategy that competes with less cooperative strategies over a number of generations. These models are usually interpreted in terms of cultural evolution. After a finite series of social interactions involving material payoffs, the next generation copies the most successful strategies from the previous generation. Computer modelling shows that in reasonably realistic circumstances, altruistic strategies are favoured and prove evolutionary stable despite their incompatibility with the "homo economicus" model (Gintis et al. 2005

p. 10). Non-cooperative strategies often prove suboptimal in the long run; they die out after several generations or cause the collapse of the whole system (Bowles and Gintis 2004; Gintis et al. 2003). In this type of studies, the strategies are called 'altruistic' if they satisfy the definition of behavioural altruism.

The third—and more direct—way to challenge the relevance of the "homo economicus" model consists in investigating the motivations underlying social decisions. This is a common theme of many studies in classical behavioural psychology (e.g. Batson 1991; Cialdini et al. 1987), experimental economics and neuroeconomics (Glimcher et al. 2009; Sanfey 2007). These studies refer to different forms of altruism, depending on the particular aspect of human decision-making they investigate. In psychology, researchers are mostly interested in genuine motivation, thus they make use of psychological altruism which traces the motivational chain to its most fine-grained causal components. In economics, researchers mostly investigate ordinary people's preferences, and thus use the less restrictive notion of preference altruism. Some of them, however, examine the more detailed motivational structure of people's choices (Ellingsen et al. 2010; Glimcher et al. 2009; Harbaugh et al. 2007; Houser and Xiao 2010; Knoch et al. 2010; Mayr et al. 2009; Singer and Fehr 2005) and sometimes make explicit reference to psychological altruism (e.g. de Quervain et al. 2004).

Some crucial conceptual differences

The four forms of altruism described above appear in different scientific contexts, but they also diverge from one another at the conceptual level. Reproductive and behavioural altruism focus on the *outcomes of behaviours*, whereas psychological and preference altruism focus on *individuals' subjective motivations*—their goals or preferences—for actions. In order to understand the broad phenomenon of altruism, it is essential to disentangle behavioural outcomes from motivational aspects (Peacock et al. 2005). This distinction is closely connected to the one between ultimate and proximate explanations described in "Psychological versus reproductive altruism" section. Consider the example of 'altruistic punishment', a behaviour that has been repeatedly observed in laboratory experiments (de Quervain et al. 2004; Fehr and Gächter 2002). It consists in punishing at some cost an individual who has been unfair, although no future individual gain can be obtained by this punishment. Such action is behaviourally altruistic. However, at the proximate level, the behaviour seems to be driven by a desire for revenge: most individuals seem to seek the satisfaction of seeing free-riders pay for their actions (Clavien and Klein 2010). Hence, at least no psychological altruism is involved.

Reproductive altruism differs from behavioural altruism in three major ways. First, and most importantly, different currencies are at stake. While biologists measure reproductive altruism in terms of fitness—the genetic contribution to the next generation—scholars who argue against the "homo economicus" view are not interested in this measure. They calculate costs and benefits that are meant to reflect individuals' interests and welfare. In experimental economics, costs and benefits are usually translated into monetary units. Such currencies are not a good proxy for

fitness—especially in light of the current demographic trend observed in rich societies. Hence, the results of micro-economic experiments cannot be extrapolated to explain reproductive altruism in evolutionary biology.

Second, the time span considered for reproductive and behavioural altruism are generally different (Jensen 2010). The relevant timescale when measuring fitness is the reproductive output over the entire life of an organism.[6] In contrast, behavioural altruism is usually measured over short periods, with no or little assumption about its long-term consequences. In laboratory experiments, a behaviour is judged altruistic if it proves costly at the end of the game played by the participants. However, this behaviour may not be *costly in the long run:* the participants may be applying a strategy that usually proves beneficial in real life, long-term interactions. In particular, being generous is an excellent way to make friends and favour long-term cooperative interactions (West et al. 2011). Similarly, in theoretical models used in evolutionary game theory, the outcome of competing strategies can be calculated after a variable number of rounds. It is thus possible that strategies such as strong reciprocity are suboptimal in the short term—after one or a few interactions—while being efficient in the long run—after a number of interactions representative of the strategy-bearers' lifetime (Vromen 2012). As Sam Bowles and colleagues note, 'strong reciprocity may be a basic human behaviour that under conditions prevailing in recent history generally confers fitness advantages on its bearers in comparison with their self-interested compatriots' (Bowles et al. 2003: 5). Similarly, one can read in Gintis et al. that 'strong reciprocity must have promoted individual fitness, or it could not have evolved' (Gintis et al. 2008: 248).

The third difference is that reproductive and behavioural altruisms are used in distinct contexts. The notion of reproductive altruism is tightly associated with an evolutionary scenario. As for any selected trait, a predisposition for reproductive altruism can only be fixed in a population over many generations during which a large number of altruistic individuals have encountered a broad range of contexts (Maynard Smith 1989). This evolutionary view is usually neglected in discussions of behavioural altruism. A behaviour observed on one occasion in an artificial context may be described as altruistic, irrespective of whether the experimental setting is representative of humans' everyday experiences or past evolutionary conditions.

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.

A similar type of relationship links psychological and preference altruism. Both refer to people's motivation to act altruistically, but the definition of preference altruism is broader, as it makes no detailed stance on actual causal mechanisms underlying motivation; it only requires that people have robust other-regarding preferences and act upon them. Preference altruism may but does not necessarily

[6] In practical cases, the fitness consequences of a behaviour are often estimated over a shorter period, but with the assumption that they are representative of an effect on the final life-time fitness of the individuals.

call for genuine concern for others; it may be practised by individuals who indirectly seek to fulfil self-directed drives, such as the personal pleasure felt when being generous. For example, a person who refrains from buying luxury items and prefers to periodically give her extra-money to a charity displays preference altruism. However, if her other-regarding preference is itself motivated by the expectation to regularly feel a warm glow inside or by the hope to increase her chances to enter Paradise after death, it would not count as psychologically altruistic.

In sum, psychological altruism is a special case of preference altruism. It is characterized by genuine concern for others. The distinction is important in philosophical discussions on morality, because preference altruism may not reveal people's genuine motivations and may thus be seen as more other-regarding than it really is.

Notably, both psychological and preference altruism refer to proximate mechanisms for helping behaviour. Thus it remains interesting to provide ultimate accounts of the evolution of these preferences for others' interests and welfare. Kin selection and possibly individual advantages due to indirect or delayed reciprocity might be part of the explanation (Clavien and Chapuisat 2012; West et al. 2007).

Misleading uses of altruism terminology

When a single term covers different concepts in diverse disciplines, major difficulties are to be expected in interdisciplinary debates. Unfortunately, this applies to current discussions on human altruism. The notion of altruism has become so plastic that it is often hard to understand what is really meant by the authors using the term, and even harder to evaluate the degree to which results from one research field—e.g. experimental economics—may facilitate the resolution of debates in another research field—e.g. evolutionary biology or philosophy—(Vromen 2012; West et al. 2007; West et al. 2011).

Some of the cryptic conceptual moves are due to the different argumentative strategies applied against the "homo economicus" view, each of them calling for a different type of altruism (see Behavioural and preference altruism section). Terminological confusions also arise from the interdisciplinary nature of this research field. Here is an example:

> The experimental evidence [experimental economics' achievements] supporting the ubiquity of non-self-regarding motives, however, casts doubt on both the economist's and the biologist's model of the self-regarding human actor. (…) Evolutionary theory suggests that if a mutant gene promotes self-sacrifice on behalf of others (…) the mutant should die out. (…) Any model that suggests otherwise must involve selection on a level above that of the individual. Working with such models is natural in several social science disciplines but has been generally avoided by a generation of biologists weaned on the classic critiques of group selection by Williams (1966), Dawkins (1976), Maynard Smith (1976), Crow and Kimura (1970), and others,

> together with the plausible alternatives offered by Hamilton (1964) and Trivers (1971). But the evidence supporting strong reciprocity calls into question the ubiquity of these alternatives. (Gintis et al. 2005: 8-9)

The authors of this citation are researchers in experimental economics, evolutionary anthropology and evolutionary game theory. They state that experimental and theoretical results on strong reciprocity cast doubt on biological explanations for altruism. However, this claim is highly disputable, because it fails to distinguish dissimilar notions of altruism. Strong reciprocity can be understood in a motivational sense—preference altruism—or a behavioural sense—behavioural altruism ("Behavioural and preference altruism" section). The authors may be referring to preference altruism, as the formula 'non-self-regarding motives' suggests. However, preference altruism is logically distinct from reproductive altruism ("Some crucial conceptual differences" section); the first focuses on people's motives whereas the second refers to behavioural outcomes. Showing that people have other-regarding preferences provides no indication of why reproductive altruism has evolved. Here, the authors confuse proximate and ultimate causes (for more on this criticism, see West et al. 2011).[7] If the authors are referring to behavioural altruism, a similar difficulty arises. Behavioural and reproductive altruism are not identical. They differ with respect to currency, time span and contexts. Thus, studies on behavioural altruism cannot directly improve our understanding of reproductive altruism.

In a similar vein, a behavioural outcome—or a behavioural propensity—should not be confused with its proximate mechanisms or ultimate explanations, as when Gintis et al. (2005: 11) write that

> Primates form alliances, share food, care for one another's infants, and give alarm calls—all of which most likely can be explained in terms of long-term self-interest and kin altruism. Such forms of cooperation are no less important in human society, of course, and strong reciprocity can be seen as a generalization of the mechanisms of kin altruism to nonrelatives.

Thus, one should be particularly careful when reading authors who conflate various senses of altruism, as occurs here with the reproductive and behavioural forms of altruism.

> [We use] the term altruism for helping in situations where the helper would benefit in fitness *or other material ways* by withholding help. This *is the standard biological definition* adopted by Hamilton (1975), Grafen (1984). (Bowles and Gintis 2011: 8, our emphasis)

Here, the inclusion of benefits other than fitness deviates from the standard biological view adopted by Hamilton and Grafen, and opens up the way for confusions, and possibly, overstated claims.

Another upshot of an imprecise use of altruism terminology is that it makes unclear the relevance of theoretical or empirical results for an understanding of

[7] Other example of this type of confusion between proximate and ultimate explanation are to be found in (Gintis et al. 2008: 249; Chaudhuri 2011: 78).

human morality. The connection between altruism and morality is rarely thematized in the literature on human social decision-making, but lurks in the background. Book and article titles such as 'Moral Sentiments and Material Interests' (Gintis et al. 2005), 'Brain Responses to the Acquired Moral Status of Faces' (Singer et al. 2004), or 'Strong Reciprocity and the Roots of Human Morality' (Gintis et al. 2008) are evocative examples. However, moral terminology should not always be taken as face value, because the link between altruism and morality is not straightforward.

Reproductive altruism seems at best remotely linked to morality, since it can even occur in organisms that lack nervous system, such as slime moulds. By definition, behavioural altruism is disconnected from individuals' motivations, which makes it a poor indicator of morality. Indeed, common-sense as well as most philosophical understandings of morality emphasise individuals' motivations: an action is morally good if it is done with the right intention or emotional incentive—moral theories referring to reasons, values, emotions, or intuitions fall under this category. Even consequentialist views of morality would not attribute a privileged role to behavioural altruism; the relevant moral criterion in these models is whether the consequentialist principle—e.g. maximisation of utility—is fulfilled.[8] Preference altruism does not seem a good approximation for morality either, since it allows for self-directed instrumental motivations that no moral system puts forward. Psychological altruism seems to be the best candidate for drawing links with morality; genuine other-directed motivation is the mark of western moral ideals and is explicitly integrated in numerous philosophical accounts of morality (e.g. Nagel 1970; Ruse 1998; Sober 1993). However, psychological altruism—in the form defined in this paper—is hardly explicitly discussed in the economics and evolutionary anthropology literature.[9] This is understandable, as these research fields are mainly concerned with general descriptions or predictions of behaviour in socioeconomic contexts (Gintis et al. 2008: 247). Knowledge about the genuine motivations underlying individuals' choices of action seems unnecessary to understand the long-standing preferences that are central to decision-making. In this context, it makes more sense to use preference altruism which can easily be integrated into utility functions (Carpenter 2007).[10]

Conclusion

Altruism is a multifarious notion that requires specification. We have argued that the use of a clear taxonomy—independently of the details of the definitions—is

[8] We are of course not denying the fact that instances of behavioural altruism *can* be moral. Our point is that these behaviours unlikely qualify as moral *in virtue of* being behaviourally altruistic.

[9] One could debate on this point however because it is not always clear whether the authors defend a flexible or a more demanding form of other-regarding motivation (Vromen 2012). Moreover, it is in principle possible to formalize fine-grained other-regarding motivations in terms of utility functions (see Clavien 2012b).

[10] It is still a matter of debate however, to what extent this would help to improve the axiomatic theory used in economic theory (Binmore 2005).

essential for progress towards an integrative and multi-faceted understanding of human altruism.

The four definitions of altruism proposed here point to important distinctions between research fields and should help to avoid misunderstandings across disciplines and over-interpretations of findings relevant to specific aspects of altruism. Firstly, most of the research on human behaviour does not deal with reproductive altruism, and has little relevance for it—what is at stake is rarely survival and reproduction. Conversely, the powerful explanations for reproductive altruism that were proposed by evolutionary biologists long before experimental economists and evolutionary anthropologists joined the field are only indirectly linked to the three other notions of altruism. Secondly, psychological altruism is likely to be a useful concept in philosophy and psychology, but not so much in evolutionary biology, experimental economics, and evolutionary anthropology. Thirdly, preference and behavioural altruism are the most relevant concepts for economics; they help in challenging the "homo economicus" model and provide some hints on how to refine the content of human utility functions. However, these two forms of altruism only shed light on a tiny aspect of human social decision-making. Studying the conditions under which people are ready to help, collaborate, or punish others, and the intimate brain processes involved, goes a long way beyond studying preference and behavioural altruism. Indeed, altruism is one specific—may be only peripherally interesting—component of human social behaviour. Therefore, by confining the rhetoric of altruism to the space where it belongs and keeping to clear definitions, we will have more chance of grasping the true originality and complexity of human social behaviour.

Acknowledgments We thank the editor, an anonymous reviewer, Chloë FitzGerald, Conrad Heilmann, Laurent Keller, Laurent Lehmann, Jessica Purcell, and Jack Vromen for comments on the manuscript. We are grateful to Philip Kitcher who gave us the impulsion to work out the concept of "preference altruism". Our research is supported by the Swiss National Science Foundation and the Fondation du 450ème of the University of Lausanne.

References

Andrade MCB, Banta EM (2002) Value of male remating and functional sterility in redback spiders. Anim Behav 63:857–870. doi:10.1006/anbe.2002.003

Andreoni J (1990) Impure altruism and donations to public goods: a theory of warm-glow giving. Econ J 100(401):464–477

Bargh JA, Gollwitzer PM, Oettingen G (2010) Motivation. In: Fiske ST, Gilbert DT, Lindzey G (eds) Handbook of social psychology, 5th edn. Wiley, New York, pp 268–316

Batson CD (1991) The altruism question: toward a social psychological answer. Lawrence Erlbaum, Hillsdale

Binmore KG (2005) Natural justice. Oxford University Press, New York

Binmore KG (2006) Why do people cooperate? Politics Philos Econ 5(1):81–96

Bourdieu P (2000) Les structures sociales de l'économie. Seuil, Paris

Bowles S, Gintis H (2004) The evolution of strong reciprocity: cooperation in heterogeneous populations. Theor Popul Biol 65(1):17–28

Bowles S, Gintis H (2011) A cooperative species: human reciprocity and its evolution. Princeton University Press, Princeton

Bowles S, Fehr E, Gintis H (2003) Strong reciprocity may evolve with or without group selection. Working paper edn

Butler J (1991) Fifteen sermons. In: Raphael DD (ed) British moralists, 1650–1800: selected and edited with comparative notes and analytical index, vol 1. Clarendon Press, Oxford, pp 325–377
Cabanac M, Guillaume J, Balasko M, Fleury A (2002) Pleasure in decision-making situations. BMC Psychiatry 2(1):7
Carpenter JP (2007) The demand for punishment. J Econ Behav Organ 62(4):522–542
Carpenter JP, Bowles S, Gintis H, Hwang SH (2009) Strong reciprocity and team production: theory and evidence. J Econ Behav Organ 71(2):221–232. doi:10.1016/j.jebo.2009.03.011
Charness G, Gneezy U (2008) What's in a name? Anonymity and social distance in dictator and ultimatum games. J Econ Behav Organ 68(1):29–35
Chaudhuri A (2011) Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp Econ 14(1):47–83
Cialdini RB, Schaller M, Houlihan D, Arps K, Fultz J, Beaman AL (1987) Empathy-based helping: is it selflessly or selfishly motivated? J Pers Soc Psychol 52(4):749–758
Clavien C (2012a) Altruistic emotional motivation: an argument in favour of psychological altruism. In: Plaisance K, Reydon T (eds) Philosophy of behavioral biology. Boston Studies in Philosophy of Science. Springer, Dordrecht, pp 275–296
Clavien C (2012b), Kitcher's revolutionary reasoning inversion in ethics. Analyse & Kritik 1
Clavien C, Chapuisat M (2012) Altruism—a philosophical analysis, eLS. Wiley, Chichester. doi: 10.1002/9780470015902.a0003442.pub2
Clavien C, Klein RA (2010) Eager for fairness or for revenge? Psychological altruism in economics. Econ Philos 26:267–290
Comte A (1851–1854) Système de politique positive, ou, traité de sociologie instituant la religion de l'humanité. L. Mathias, Paris
Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper & Row, New York
Dawkins R (1976) The selfish gene. Oxford University Press, New York
de Quervain DJF, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E (2004) The neural basis of altruistic punishment. Science 305(5688):1254–1258. doi:10.1126/science.1100735
Ellingsen T, Johannesson M, Tjotta S, Torsvik G (2010) Testing guilt aversion. Game Econ Behav 68(1):95–107. doi:10.1016/j.geb.2009.04.021
Fehr E, Camerer C (2007) Social neuroeconornics: the neural circuitry of social preferences. Trends Cogn Sci 11(10):419–427. doi:10.1016/j.tics.2007.09.002
Fehr E, Fischbacher U (2003) The nature of human altruism. Nature 425(6960):785–791
Fehr E, Fischbacher U (2004a) Social norms and human cooperation. Trends Cogn Sci 8(4):185–190
Fehr E, Fischbacher U (2004b) Third-party punishment and social norms. Evol Human Behav 25(2):63–87
Fehr E, Gächter S (2002) Altruistic punishment in humans. Nature 415:137–140
Fehr E, Rockenbach B (2003) Detrimental effects of sanctions on human altruism. Nature 422(6928):137–140
Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. Quart J Econ 114(3):817–868. doi:10.1162/003355399556151
Foster KR (2008) Altruism. In: Jorgensen SE, Fath B (eds) Encyclopedia of ecology. pp 154–159
Frank SA (1998) Foundations of social evolution. Monographs in behavior and ecology. Princeton University Press, Princeton
Ghiselin MT (1974) The economy of nature and the evolution of sex. University of California Press, Berkeley
Gigerenzer G (2008) Gut feelings: short cuts to better decision making. Penguin Books, London
Gintis H (2000) Game theory evolving: a problem-centered introduction to modeling strategic behavior. Princeton University Press, Princeton
Gintis H, Bowles S, Boyd R, Fehr E (2003) Explaining altruistic behavior in humans. Evol Human Behav 24(3):153–172
Gintis H, Bowles S, Boyd R, Fehr E (eds) (2005) Moral sentiments and material interests: the foundations of cooperation in economic life. MIT Press, Cambridge
Gintis H, Henrich J, Bowles S, Boyd R, Fehr E (2008) Strong reciprocity and the roots of human morality. Social Justice Res 21(2):241–253
Glimcher PW, Camerer CF, Fehr E, Poldrack RA (eds) (2009) Neuroeconomics: decision making and the brain. Academic Press, London
Grafen A (1984) Natural selection kin selection and group selection. In: Krebs JR, Davies NB (eds) Behavioural ecology: an evolutionary approach. Sinauer, Sunderland, pp 62–84

Grafen A (1985) A geometric view of relatedness. In: Dawkins R, Ridley M (eds) Oxford surveys in evolutionary biology, vol 2. Oxford University Press, Oxford, pp 905–907
Hamilton WD (1964) The genetical evolution of social behaviour. J Theor Biol 7(1):1–52
Hamilton WD (1970) Selfish and spiteful behaviour in an evolutionary model. Nature 288(5277): 1218–1220
Hamilton WD (1975) Innate social aptitudes of man; an approach from evolutionary genetics. In: Fox R (ed) Biosocial anthropology. Malaby Press, London, pp 133–155
Harbaugh WT, Mayr U, Burghart DR (2007) Neural responses to taxation and voluntary giving reveal motives for charitable donations. Science 316(5831):1622–1625
Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R (2001) In search of homo economicus: behavioral experiments in 15 small-scale societies. Am Econ Rev 91(2):73–78
Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R, Alvard M, Barr A, Ensminger J, Henrich NS, Hill K, Gil-White F, Gurven M, Marlowe FW, Patton JQ, Tracer D (2005) "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. Behav Brain Sci 28(6):795–855
Hobbes T (2005) Leviathan. Broadview Press, Peterborough
Hoffman E, McCabe K, Smith V (1996) Social distance and other-regarding behavior in dictator games. Am Econ Rev 86(3):653–660
Houser D, Xiao E (2010) Inequality-seeking punishment. Econ Lett 109(1):20–23. doi: 10.1016/j.econlet.2010.07.008
Hutcheson F (2004) An inquiry into the original of our ideas of beauty and virtue. Liberty Fund, Indianapolis
Jensen K (2010) Punishment and spite, the dark side of cooperation. Philos T R Soc B 365(1553): 2635–2650. doi:10.1098/rstb.2010.0146
Kahneman D, Slovic P, Tversky A (1982) Judgment under uncertainty: heuristics and biases. Cambridge University Press, Cambridge, New York
Kirchgässner G (ed) (2008) Homo oeconomicus: the economic model of behaviour and its applications in economics and other social sciences. Springer, New York
Kitcher P (2011) The ethical project. Harvard University Press, Cambridge
Knoch D, Gianotti LRR, Baumgartner T, Fehr E (2010) A neural marker of costly punishment behavior. Psychol Sci 21(3):337–342. doi:10.1177/0956797609360750
Kurzban R, DeScioli P, O'Brien E (2007) Audience effects on moralistic punishment. Evol Human Behav 28(2):75–84
Lehmann L, Keller L (2006) The evolution of cooperation and altruism; a general framework and a classification of models. J Evol Biol 19(5):1365–1376
Lewisch PG, Ottone S, Ponzano F (2011) Free-riding on altruistic punishment? Experimental comparison of third-party-punishment in a stand-alone and in an in-group environment. Rev Law Econ 7(1):165–194. doi:10.2202/1555-5879.1460
Mandeville B (1997) The fable of the bees: and other writings. Hackett Publishing company, Indianapolis
Maynard Smith J (1976) Group selection. Quart Rev Bio 51: 277-283
Maynard Smith J (1989) Evolutionary genetics. Oxford University Press, Oxford
Mayr E (1961) Cause and effect in biology. Science 134:1501–1506
Mayr U, Harbaugh WT, Tankersley D (2009) Neuroeconomics of charitable giving and philantropy. In: Glimcher PW, Camerer CF, Fehr E, Poldrack RA (eds) Neuroeconomics: decision making and the brain, 1st edn. Elsevier, Amsterdam, pp 303–320
Nagel T (1970) The possibility of altruism. Clarendon Press, Oxford
Okasha S (2007) Evolution and the levels of selection. Oxford University Press, Oxford
Peacock MS, Schefczyk M, Schaber P (2005) Altruism and the indispensibility of motives. Anal Kritik 27:188–196
Queller DC (1992) A general-model for kin selection. Evolution 46(2):376–380
Rand A (1964) The virtue of selfishness: a new concept of egoism. New American Library, New York
Ruse M (1998) Taking darwin seriously: a naturalistic approach to philosophy. Prometheus Books, Buffalo
Sanfey AG (2007) Social decision-making: insights from game theory and neuroscience. Science 318:598–602
Simon HA (1996) The sciences of the artificial, 3rd edn. MIT Press, Cambridge
Singer T, Fehr E (2005) The neuroeconomics of mind reading and empathy. Am Econ Rev 95(2):340–345

Singer T, Kiebel SJ, Winston JS, Dolan RJ, Frith CD (2004) Brain responses to the acquired moral status of faces. Neuron 41(4):653–662. doi:10.1016/s0896-6273(04)00014-5
Smith A (2002) The theory of moral sentiments. Cambridge texts in the history of philosophy. Cambridge University Press, Cambridge
Sober E (1992) Hedonism and Butler's stone. Ethics 103(1):97–103
Sober E (1993) Evolutionary altruism, psychological egoism and morality; disentangling the phenotypes. In: Nitecki MH et al (eds) Evolutionary ethics. SUNY Press, Albany, pp 199–216
Sober E, Wilson DS (1998) Unto others: The evolution and psychology of unselfish behavior. Harvard University Press, Cambridge
Stich SP (2007) Evolution, altruism and cognitive architecture: a critique of sober and Wilson's argument for psychological altruism. Biol Philos 22(2):267–281
Stich SP, Doris JM, Roedder E (2010) Altruism. In: Doris JM (ed) The moral psychology handbook. Oxford University Press, Oxford, pp 147–205
Stocks EL, Lishner DA, Decker SK (2009) Altruism or psychological escape: why does empathy promote prosocial behavior? Eur J Soc Psychol 39(5):649–665
Tinbergen N (1963) On aims and methods of ethology. Zeitschrift für Tierpsychologie 20:410–433
Trivers RL (1971) The evolution of reciprocal altruism. Q Rev Biol 46(1):35–57
Vromen J (2012) Human cooperation and reciprocity. In Okasha S, Binmore K (eds) Evolution and rationality: decisions, cooperation and strategic behavior, Cambridge University Press, Cambridge
West SA, Griffin A, Gardner A (2007) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J Evol Biol 20:415–432
West SA, El Mouden C, Gardner A (2011) Sixteen common misconceptions about the evolution of cooperation in humans. Evol Human Behav 32(4):231–262
Williams GC (1966) Adaptation and natural selection; a critique of some current evolutionary thought. Princeton University Press, Princeton
Wilson DS (1975) A theory of group selection. P Natl Acad Sci USA 72(1):143–146
Zahavi A (1975) Mate selection—a selection for a handicap. J Theor Biol 53(1):205–214