

Altruistic emotional motivation: An argument in favour of psychological altruism

CHRISTINE CLAVIEN

UNIVERSITY OF LAUSANNE

christine.clavien@unil.ch

Abstract

In this paper, I reframe the long-standing controversy between ‘psychological egoism’, which argues that human beings never perform altruistic actions, and the opposing thesis of ‘psychological altruism’, which claims that human beings are, at least sometimes, capable of acting in an altruistic fashion. After a brief sketch of the controversy, I begin by presenting some representative arguments in favour of psychological altruism before showing that they can all be called into question by appealing to the idea of an unconscious self-directed motive. I will then point out that this argumentative strategy not only debunks the reasons for favouring psychological altruism, but also those for favouring psychological egoism; hence it is no use in settling the dispute between the two views. In the second part of the paper, I will try to break this deadlock by reframing the whole controversy, shifting it away from the concept of motive, towards the broader notion of motivation. As it turns out, this shift enables the debate to centre on altruistic emotions and their motivational power, thereby allowing evolutionary arguments to enter the debate and settle the dispute in favour of psychological altruism.

Keywords: affect, emotion, motivation, motive, psychological altruism, psychological egoism, reliability argument, Sober & Wilson, unconscious

The controversy

Human beings are capable of helping other human beings in need at their own expense and apparently without thinking of their own short or long term interest. Philosophers and psychologists label this sort of action ‘altruistic’, or ‘apparently altruistic’.

This notion of altruism should not be confused with ‘behavioural’ or ‘evolutionary altruism’, as it is understood in biology or economics. Behavioural altruism is defined in terms of its *outcomes* on individual fitness or well-being, whereas psychological altruism is about the internal *motives* responsible for helping actions. The notion of motive is not well defined in the literature. However, most authors would agree that the set of motives is a broad category that includes different things, such as desires, intentions or judgments. Motives underlie the whole procedure that eventually leads to a helping act. Moreover, they seem to have an articulated conceptual content – some would prefer to say that they are bound to beliefs – as well as an affective component, a bodily set of sensations that is felt as an urge to do something.

There is a traditional philosophical debate over the possibility of psychological altruism that divides philosophers into two categories: those who defend the possibility of genuine altruistic actions and those who think that all ‘apparently altruistic’ actions are in fact egoistically motivated. To make sense of these formulae, some definitions are needed. A key concept in the debate is the notion of the *primary*¹ – as opposed to *instrumental* – motive of helping actions. A primary motive is the *first* motive of a causal chain that leads towards action; it is also the driving force that lasts until the action has come about. If the action is set off by more than one cause – or causal chain – a primary motive must be at least a necessary condition for the action to come about.² Here is an example:

Raymond seeks pleasure [primary motive] → Raymond thinks that if he does x, he will obtain pleasure [instrumental practical reasoning] → Raymond desires to do x [instrumental motive in order to achieve pleasure] → Raymond does x

Motives are distinguishable in terms of their objects. If a primary motive is directed towards the needs and well-being of other individuals, it earns the label ‘altruistic’. If a primary motive aims at some personal benefit for oneself, it is considered ‘self-interested’.

¹ In the philosophical literature, primary motives are usually called ‘ultimate motives’. However, in order to avoid confusion with the notion of “ultimate cause” as described in biology, I will avoid this formula here.

² A complementary way to capture the distinction between *primary* and *instrumental* motive is to think of their ends, of what they aim at. A primary motive is directed towards an *end in itself* whereas an instrumental motive – which is situated in the centre of a motivational causal chain – is directed towards an *intermediate end*, an end that is supposed to help in reaching the ultimate end of the primary motive.

'Psychological altruism' (PA) is the view according to which *at least some* actions are motivated by altruistic primary motives. On the contrary, 'psychological egoism' (PE) denies the possibility of primary altruistic motives. According to this latter view, all human actions are motivated by the expectation of some personal benefit, usually conceived of in terms of pleasure and avoidance of pain – the hedonistic version – or such things as power, resources, or reputation.

It is worth noting that PE does not deny that actions motivated by self-interested motives can have positive effects for others. It is possible to seek one's own happiness without endangering others' well-being. PE does not deny the reality of non self-interested motives either, provided these motives are mediate objects of a primary self-directed motive. PE allows for sincere desires to help a person in need, but these desires can only be instrumental; they must be considered the best way to achieve a personal good – for example a fine reputation. In other words, others' well-being can be a mediate but not a primary object of one's motives.

Here, we can see that PE is a universal thesis about human motivation. It denies the reality of non self-interested motives; everything must be explained in terms of self-interest – for example, desire for applause, honour, pleasure, avoidance of pain. This universal aspect of PE makes it a very demanding claim because it is incompatible with *any* occurrence of a primary motive aiming at something other than one's own well-being; it must rule out any primary desire to help others, or to act in accordance with moral duties, even desires for self-destruction.

In favour of psychological altruism

Among philosophers, PA is usually favoured over PE. Rejection of the latter view is partly due to moral considerations. As such, PE is not a normative thesis; it does not take a stand on moral issues. However, when it is combined with the fairly widely-held thesis that an action is morally good only if it is caused by non self-interested motives, one cannot escape the conclusion that there is no moral action. This is a good reason to dislike PE, but it is not a knockdown argument. Firstly, one could question the very idea of defining moral action in terms of other-directed motives. Secondly, even if we accept this definition of moral action, PE might force us to admit that morality is only a matter of illusion. More is needed in order to convincingly reject PE. In this section, I will briefly present a representative panel of

arguments against this view. However, I will not elaborate on these arguments, my purpose being to give a glimpse of the sort of objections that can be made against PE, before showing in the next section that all these objections can be rejected by a single argumentative line.

We have seen that PE is a variety of motivational monism. This demanding aspect of the theory has led most advocates of PA to search for counterexamples, particular actions or types of action that cannot be convincingly explained in egoistic terms. Indeed, PE could be proven false by showing that *at least* one action has been performed that was motivated by a non self-interested primary motive. In a thought experiment, Francis Hutcheson (2004: treatise II, section II, p. 224) aimed to provide one particular example of a helping action that could not be explained in egoistic terms. His story is as follows: imagine God told you that you were going to be annihilated in a few seconds, but that you had a last choice to make; you could choose to make your families, friends and humanity in general either happy or miserable in the future. However, you would not be able to feel any pleasure or pain as a consequence of your choice. Under these circumstances, he argues, many of us would choose the first option, that is, to make others happy. Such a choice cannot be explained by self-interested motives. Therefore, PE is false.

Hutcheson's thought experiment was intended to provide *one particular example* of an action caused by an altruistic motive. In the literature, one can also find more general arguments, such as the attempt to show that some *types* of behaviours cannot be explained in egoistic terms. Take the 'argument from moral approbation' which is also to be found in Hutcheson's writings (2004: treatise II, section II, p. 102ff.). Its formalized version is as follows:

- P1: We do not morally approve of actions that have good effects, but are merely motivated by self-interested desires.
- P2: We morally approve of some actions.
- C: There must be some actions that are not motivated by self-interested desires, therefore PE is false.

Besides the thought experiments and formal arguments typically used by philosophers, psychologists have tried to prove the existence of actual cases where agents have no interest at all in helping others, but still choose to do so. For example, Daniel Batson famously demonstrated in a series of empirical studies that high levels of empathy cause people to help others, even when they are given the opportunity to escape, which he argues would amount to acting from an egoistic motive (Batson 1991). Batson takes these data as sufficient

counterexamples to PE; he cannot conceive of an egoistic interpretation for this sort of helping behaviour.

Finally, more general arguments have been proposed, which do not focus on particular actions or types of behaviour. For example, Joseph Butler famously defended the following line of argument (Butler 1991: § 415):

P1: We sometimes experience pleasure. For example, eating a piece of cake produces pleasure.

P2: Pleasure can *only* emerge as an epiphenomenon of actions caused by desires for external things. For example, pleasure from eating a piece of cake does not come from the desire to experience pleasure, but from the conjunction of a desire for a piece of cake *and* the satisfaction of this desire.

C: PE is false.

Plenty of other arguments are to be found in the literature. It is not my purpose to discuss them all. I only intended to propose a brief review of the sort of objections that can be made against PE in order to capture the efficacy of the unique type of counterargument to which we now come.

In favour of psychological egoism

Recently, Sober and Wilson (1998) have famously argued that philosophical arguments and empirical data stemming from social psychology cannot prove PA because an “internal reward” explanation can always be invented to explain human action. By this, they mean that we cannot rely on introspection to identify our primary motives. Implicitly, they state a fairly well-known argument, according to which it is always possible to be mistaken about our true motives; any apparent altruistic motive could be caused by an unconscious selfish one, such as the avoidance of painful memories or the attainment of a warm feeling of self-satisfaction.

Let us briefly return to each argument presented in the previous section and see how they can be rejected with help of the notion of the unconscious.

Hutcheson’s thought experiment does not allow for the fact that subjects may not be completely persuaded of the impossibility of being rewarded for their action. They might expect to collect ‘good marks’ for their afterlife. The point here is that introspection can be deceptive; we can be mistaken about our own motives.

Similarly, the argument from moral approbation can be rejected on the grounds that it is possible that humans who morally approve of moral actions are systematically mistaken about the true motives that have led to these actions: any apparently altruistic action can – consciously or unconsciously – be caused by a self-directed motive.

At the empirical level, even if it can be shown that empathy causes helping behaviour, actions out of empathic emotions can be interpreted in egoistic terms: empathising with a needy person might create a kind of sadness that subjects know can only be successfully assuaged by helping that person (for a more extensive discussion, see Batson 2000, Sober & Wilson 2000, Sober & Wilson 1998: chap. 8).

The unconscious motivation argument is even powerful against more general objections such as Butler's attack against PE. Besides the fact that the second premise of this argument is controversial,³ nothing precludes the possibility that the motives that lead us to seek consciously for x – where x is not pleasure but something else that can elicit pleasure once obtained – are unconsciously self-directed.

In brief, a defender of psychological egoism could accept all the premises of the arguments in favour of altruism, yet reject the conclusions on the grounds that conscious motives might be deceptive. In case of apparently altruistic actions, the following causal chains would hold:

Primary self-directed motive (conscious or not) → Instrumental practical reasoning (conscious or not) → Instrumental motive directed towards other's well-being → Action [→ If the action obtains, pleasure]

More precisely, depending on the circumstances, the primary self-directed motive could involve two possible scenarios: the subject finds himself in an uncomfortable state, for example he feels bad at the sight of somebody's suffering, and this situation motivates him – consciously or not – to rid himself of this state; alternatively, the subject anticipates – consciously or not – the fact that a particular action might be good for him, for example, helping a needy person will give him a pleasant feeling of self-satisfaction, and so he finds himself motivated to perform this action.

³ As Sober and Wilson argue, "satisfying the desire for an external thing is one way, among others, in which people obtain pleasure." (1998:, 279)

Most philosophers question the plausibility of these 'internal reward scenarios' on the grounds that some of them are contrived and counterintuitive. At least some examples of highly demanding conducts – such as whistle-blowing or testifying in Criminal Court after rape – can not plausibly be understood as motivated solely by concerns about personal self-being. In the light of these examples, the burden of proof is on the side of PE. Sticking to the logical possibility of an internal reward explanation reveals that PE is a dogma rather than an explanatory theory.

PA certainly has a point here but it does not seem sufficient to settle the debate. Such a line of argumentation will only convince readers who already accept PA. Despite philosopher's argumentative efforts, one should not overlook the fact that PE keeps cropping up, especially among psychologists (Cabanac, *et al.* 2002, Cialdini, *et al.* 1987) and economists (see Macpherson 1962). Ghislin's famous "scratch an 'altruist' and watch a 'hypocrite' bleed" (1974: 247) remains an evocative formula and economic thinking in terms of individuals' preferences renders the following reflexions appealing to many readers:

"Whenever a man systematically (i.e., as a general rule) continues to sacrifice primary reward x to other people, he does so only because he usually obtains thereby some primary reward y and because y ranks higher than x on the person's preference scorecard, as determined in situations where no considerations of other people's interests and thus of sacrifice to other people's interests were involved." (Slote 1964: 533)

It is true that advocates of PE mostly stick to their case-by-case argumentation and to the impossibility to *prove* that their interpretation is wrong. However, they can also claim that their hypothesis waits to be proven, much in the same way as the theory of illusion of colour perception has been waiting to be proven. In fact, there might be more in favour of egoism than is usually thought. Recent empirical findings seem to capture some elements linked to the unconscious and add some credence to PE. A wide range of studies in experimental economics has tested subjects pro-social versus self-regarding propensities to act. Subjects were asked to play social dilemma games with each other via anonymous computer platforms; games such as the prisoner's dilemma, the trust game, or the common good game have been extensively used. These studies show that people are often ready to invest their money for the sake of the common good (Fehr & Rockenbach 2003, Fischbacher, *et al.* 2001, Henrich 2004, Marwell & Ames 1981, Ostrom 1990) or in pro-social moves even when they know that it is at their own expense and that they cannot gain anything in return (Fehr & Fischbacher 2004a,

Fehr & Fischbacher 2004b, Fehr & Gächter 2002). At first glance, one might think that these empirical evidences could be used in favour of PA. However, it is important to distinguish between people's behaviour and motives. Worries concerning subjects real motives are nicely illustrated in the 'nobody's watching' experiment conducted by Halley and Fessler (2005). The experimenters showed that very subtle cues can have a drastic impact on cooperative and pro-social behaviour. For example, simple stylised eyespots on the computer's desktop background cause a dramatic increase in pro-social behaviour. These eyespots can best be interpreted as cues relating to the presence of observers, thereby as elicitors of psychological mechanisms linked to reputation. This experiment suggests that people think of their own benefit even under the usual condition of anonymity. In the light of these results, one cannot help thinking that past economic experiments might not have been sufficiently carefully designed to avoid similar cues.

Other reasons to favour PE are to be found in recent studies using brain-imaging as a research tool. In a study by Rilling and colleagues (2002) subjects' brains were scanned while they played an iterated prisoner's dilemma game, which is about choosing to cooperate or to defect. It was shown that choice to cooperate activates the actor's brain areas that are linked with reward processing – including the 'caudate nucleus', well-known to be associated with *anticipation* of reward. According to the experimenters, activation of these parts of the brain positively reinforces reciprocity and helps in resisting the temptation to defect.

Even if these studies do not directly address the question whether altruistic motivation exists, they indicate that behaviours that seem to be good candidates for altruistic explanation are in fact best explained in terms of self-interest.

The deadlock

We have seen that in order to respond to powerful arguments in favour of PA, PE needs to resort to the unconscious. This move is interesting for a defender of the latter view precisely because the unconscious cannot be sounded, therefore PE is not easily refutable. One can always appeal to an unconscious desire for internal rewards as an explanation for apparently altruistic actions. However, the unconscious has its drawbacks. It is a double-edged sword for the supporter of PE because, if the unconscious cannot be sounded, there is no reason for favouring egoism over altruism! In the end, it seems that this line of reasoning is not of any

use to either a defender of PE or an advocate of PA, precisely because it destroys any means of settling the dispute between them.

There is some hope of overcoming this deadlock with experimental data and more specifically with the help of the new brain imaging technology that is used extensively in young research fields, such as neuropsychology and neuroeconomics. In this respect, the aforementioned experiments seem to be of particular interest. Unfortunately, there are serious doubts about the real contribution of these studies to the particular philosophical debate over altruism. To begin with, the fact that people are highly sensitive to reputation cues (Haley & Fessler 2005) does not preclude the possibility that there is altruism in the absence of these cues. As for Rilling and colleague's study (2002), the experimental design is not fine-grained enough to discriminate between two concurrent interpretations. i) On an interpretation favourable to PE, the activation of the brain areas linked with reward processing represents both the anticipation of future reward and the direct cause of cooperation. ii) On an alternative interpretation favourable to PA, the activation of these brain areas is mainly a side effect of cooperation; even if there is some anticipation of reward, it is likely to be a minor motivating factor among other altruistic and more decisive factors. The feeling of reward experienced by the subject is hence mainly a side-effect of altruistic actions.

It is an open question whether brain-imaging studies could bring novel and crucial arguments to the altruism versus egoism debate. In principle, it should be possible, allowing for proper technology and well-designed experiments. However, I doubt that the current state of knowledge about the neural systems involved in motivation allows for this level of subtlety. The relationship between an observed behaviour and specific brain activation is difficult to spell out; correlated events might not be directly causally related. As with classical psychological experiments, we are faced with the difficulty of interpreting the results and modelling situations in which we can determine with certainty whether the subjects think – unconsciously or not – of a possible advantage for themselves or whether they are truly interested in others' well-being.

Overall, one gets the impression that the whole debate over altruism cancels out in a battle of a priori statements. In what follows, I will try to show that there are replies to PE but in order to give them real force, we need to reframe the debate: instead of focusing on primary *motives*, I suggest to concentrate on the more fundamental notion of *motivation*. As we shall see, such a reframing will make refutation of PE easier to obtain.

Two ways of conceiving the motivational causal chain

Besides the deadlock just mentioned, there is another puzzling fact about the altruism versus egoism debate. Until now, the causal chain underlying our choices of action has been explained in terms of primary and instrumental motives. Figure 1 depicts this view. The arrows describe the possible causal paths that lead a subject from the perception of a situation to the action.

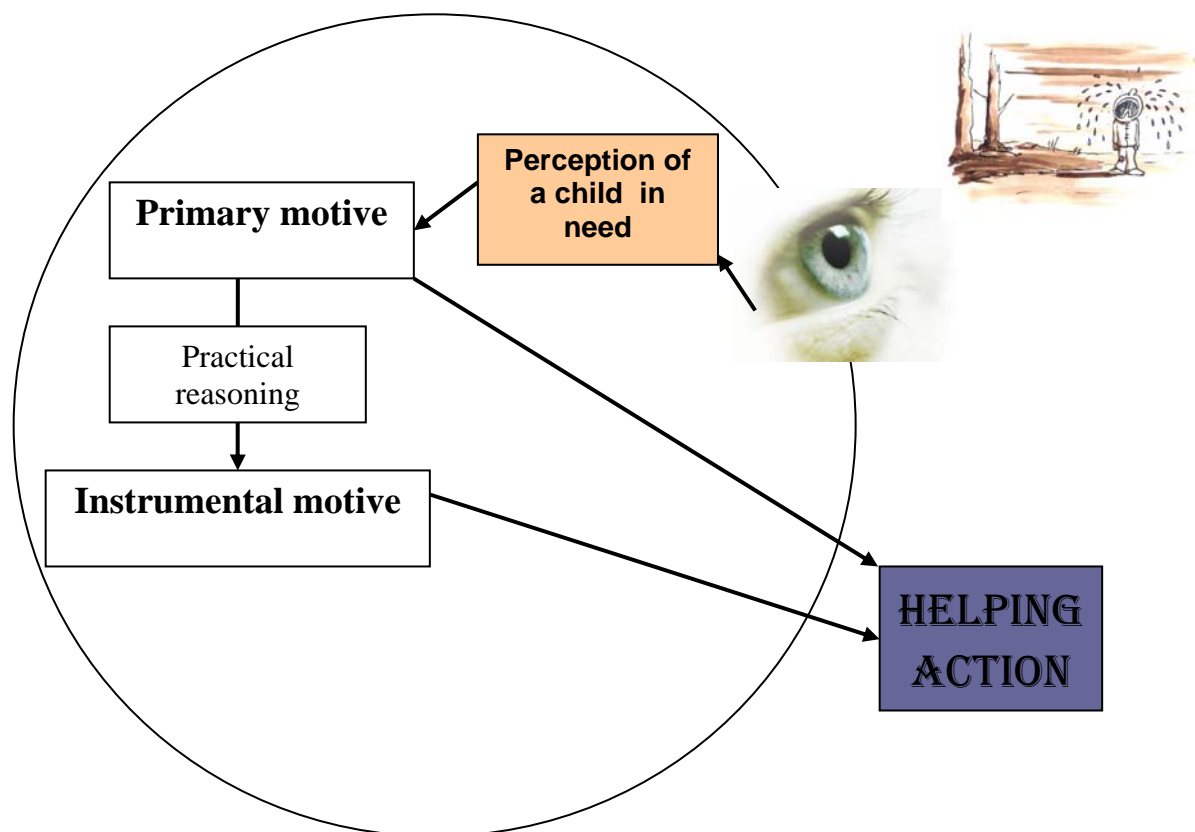


Figure 1

However, the classical way of defining the altruism versus egoism debate might prove too superficial as soon as one tries to grasp the starting point of the motivational process. Indeed, a motive is supposed to be the source of motivation. However, although usually assumed, it is not clear whether motivation always starts with a motive, that is, with a desire, an intention or a judgement. An analysis of the notion of motivation is needed here. Despite being widely used, the term ‘motivation’ is hardly ever defined in philosophical or scientific literature. This prompted Ronald de Sousa to write in an article on emotion that the motivational aspect of emotion “is infected by the obscurity of the notion of motivation” (de Sousa 2004: 65).

Motivation might be conceived of as a *relational property*: ‘D is motivated by x to do y ’; this relation holds between x and an action and takes a direction from x to y . In the context of the altruism debate, x is usually understood as a motive such as a desire, an intention or a judgment.⁴ However, x might also be an emotion. For example, being afraid usually leads to an escaping act: Charles can be motivated by his fear of the neighbour’s dog to take another path to get back home after his morning jogging. Hence, it seems that the relational property of ‘being motivated’ can stretch its arms beyond motives.

Alternatively, there is a more substantial understanding of motivation. One can point to the experiential aspect of motivation, which consists in the experience of being moved to do something. The dynamics of this experience implies that motivation does not simply refer to an abstract relational property but to ‘something’ that makes one move. The most sensible way to make sense of this ‘something’ is to consider it an *affect*, a bodily set of sensations that incites the subject to act.⁵ This affect can be embedded in different psychological states, such as an emotion, a desire, or possibly an affectively-laden judgment. This dynamic account of motivation helps to put flesh on the bones of what is often referred to as the “motivational aspect of” emotions or desires.

The dynamic account is particularly interesting because it reveals that motivation to act might not – at least, not always – come from decisions based in the will, as is often assumed. Consider a situation in which Denise is deeply touched by seeing a starving child. It seems fair to say that it is the affective part of Denise’s compassion that incites her to consider various possible helping actions, such as taking the child to her home, or giving money to his parents. The affective arousal – thus motivation – will cool down once Denise has realised one of these helping actions.⁶ Motivation is present during the whole process: it starts with an emotional reaction and is carried over from this basic state of mind to more complex states of mind such as the conscious desire to help the child.

⁴ A motivating judgement is typically considered to be the result of deliberation that is linked to an internalised norm or principle.

⁵ I argue in favour of this idea in my “An Affective Picture of Value Judgements” (submitted).

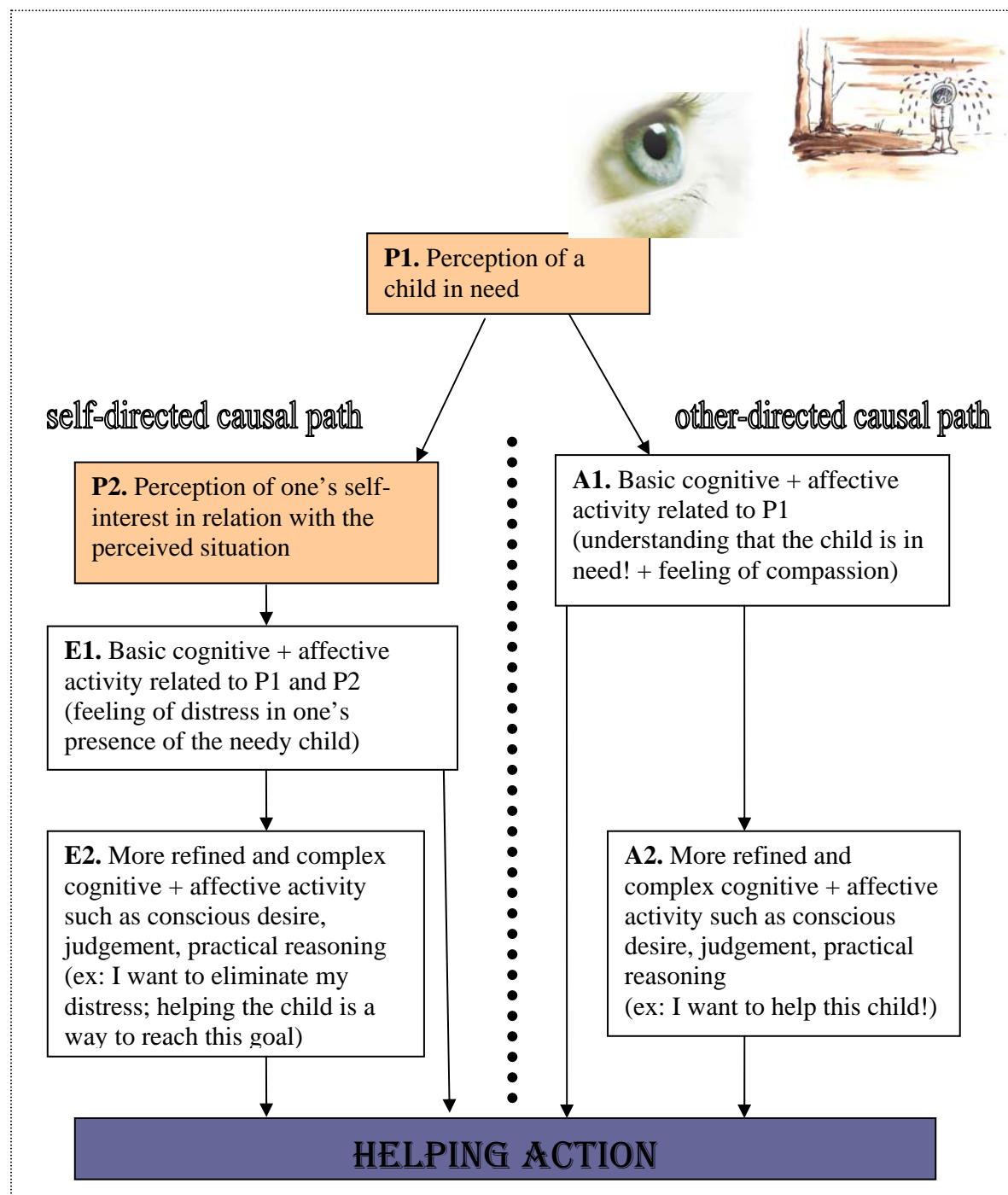
⁶ It is worth mentioning here that Denise’s situation might very well be re-described in terms of desires or judgements. However, these descriptions would not catch the fundamental source of motivation which seems to be the affective reaction itself.

The proposed distinction between motive and motivation is interesting in two respects. Firstly, it provides an alternative explanation for the causal relationship between the first input – a person in need – and the final output – the helping behaviour. According to this interpretation, motivation is not necessarily accompanied by full-fledged desires or intentions. For example, when motivation is embedded in a basic emotional reaction such as an empathic feeling, there might be no articulated conceptual content of the sort necessary for a desire or an intention. Emotions can be primitive fast and frugal reactions towards our environment.⁷ If we add some articulated conceptual contents to motivation, more complex states of minds such as desires, intentions or judgements can occur. Moreover, there is no need to set a clear line between conscious and unconscious motivation. People often become gradually aware of their tribes once they start to build cognitively around their affective reactions. Figure 2 depicts the possible causal paths that lead a subject from the perception of a situation to the action via an affective reaction. This picture can integrate classical notions such as desires, judgements,⁸ intentions or practical reasoning – boxes E2 and A2 – without being too explicit about the ways they are involved. These elements can appear down a causal path, or maybe not, depending on the situation described. More explanations of this schema will be given in the next section.

Secondly, the distinction between motive and motivation is of importance because, as we shall see in the next section, if one shifts one's attention from motives to motivation, the altruism versus egoism debate can quickly be settled.

⁷ For an detailed account of what an emotion is, see (Robinson 2005).

⁸ Some readers might want to consider emotions as a form of judgment – in the sense of appraisal. This could easily be integrated in my picture.



[Figure 2]

A proposal to reframe the debate

As we have seen, the classical debate over altruism focuses on motives – usually conceived in terms of intentions and desires – and reaches a deadlock once the unconscious argument comes into play. To resolve this deadlock my proposal is to make use of the above mentioned

distinction between motive and motivation. The strategy I propose is a shift of focus from motive to motivation. I take it to be legitimate to reframe the debate in this way for two main reasons. Firstly, we have seen that motives are not necessarily the primary cause of our actions; the causal chain goes back to the source of motivation, which seems to be an affective reaction. If motives are not – at least, not always – the original motivating source, it seems more interesting to focus the debate over altruism on the possibility of altruistic *motivation*, rather than altruistic *motives*. Secondly, the motivational component seems to have causal priority over the motive. The controversy is over whether any human action can be called altruistic. To resolve this controversy, one considers how actions are brought about. What is to be found at the beginning of the causal chain of action is often – if not always – a basic affective state, rather than a motive.

The reframed debate over altruism versus egoism would then focus on the question of whether truly altruistic motivation can exist. By this I mean the question of whether there exist motivational causal pathways triggered by the awareness of other's needs and well-being that do not include considerations upon one's own self-interest. Such a causal route would start with a basic affective state and follow the "other-directed path" – right side of the dotted line in figure 2.

Now, one very interesting aspect of the proposed shift of perspective is that it allows for introducing emotions in the analysis. Most – if not all – primary affective motivating elements are embedded in emotions.⁹ Therefore, an easy strategy for an advocate of altruism is to show the existence of 'altruistic emotions', which are capable of leading someone to act without the intervention of any further motivating factor. This sort of emotion would need to be directly elicited by the perception of another's needs and well-being and would diminish once the other's needs and well-being had been satisfied. In other words, the debate over altruism can be thought of in terms of the two following questions: Are there altruistic emotions? Is the affective component of these emotions sufficiently strong to bring about action?

Before responding to these questions, let us elaborate a bit more on the two ways emotional reactions can motivate one to act altruistically. Consider the example of parental care. Human beings are naturally inclined to feel caring emotions towards their children. Usually, when a

⁹ Although I will not argue for it here, it seems that most of our actions – possibly all – originate from emotional motivation, which amounts, more or less, to Hume's famous position. This is not to deny that emotional motivation can be monitored by conscious deliberation.

parent sees his child in need, a caring emotional reaction is elicited. The occurrence of this emotional reaction provides the first general instructions regarding the direction of the action that has to be taken. These general instructions can be followed in two ways.

In particular circumstances, the emotional reaction leads directly to a helping action. This direct motivating path is depicted in figure 2 with the arrow from box A1 to the action. Here, no particular desire, or practical reasoning is needed in addition to the emotion in order to move the subject to act. In this case, one can speak of 'actions out of emotions' (see Döring 2003). For example, if a mother suddenly sees that her child is in great danger – say, being attacked by a wild lion –, she might act spontaneously out of a caring emotion without forming any particular desire.

In most cases, however, the mental activity prior to action is more complicated. Emotional reactions can lead the subject to form complementary motives before acting – causal path A1–A2–Action. Recall Denise's example. Under emotional impulse, she builds cognitively both on her emotion and on her understanding of the child's critical condition. This mental activity leads her to form a proper motive such as a desire, an intention or a judgement which contains a more articulated conceptual content. There are also situations in which mental activity becomes even more highly complex; the agent might take time to employ practical reasoning before deciding to act.

The existence of altruistic emotions

Let us now come back to the question of whether there are altruistic emotions. At first glance, it seems that the question can quickly be settled. Who would deny the existence of emotions such as love, sympathy or compassion? It would be ridiculous to deny, for example, that human beings are naturally inclined to feel caring emotions towards their children. These emotions are clearly caused by the perception of others' needs and well-being.

Nevertheless, an advocate of PE might raise two doubts about the altruistic character of these emotions – which would amount to denying the causal links represented on the right side in figure 2.

Firstly, the supporter of PE could contend that, when examined more closely, the apparently altruistic emotional mechanisms prove to be self-directed – which amounts to saying that box A1 should be placed on the left side of the dotted line. Compassion, for example, could be described as a feeling of uneasiness that motivates the agent to engage in actions that will

eradicate this feeling. According to such an interpretation, motivation comes from the 'uneasiness' generated in compassion; the helping action is only performed because it enables the subject to rid himself of this uneasiness. On this account, it makes no sense to consider compassion an altruistic emotion.

The problem with this argument is that it distorts the notion of self-directed emotion by focusing on the phenomenological aspect of the emotion, rather than on its eliciting cause. The fact that compassion has a phenomenology of 'uneasiness', which vanishes once the input changes, does not make this emotion self-directed. Firstly, because the uneasiness itself expresses one's concern for others; secondly because the relief felt after acting to help others could be a side effect; and thirdly, because this argument overlooks the important fact that, by definition, what makes a motivational system 'altruistic' is the way it has been elicited and is maintained, whatever the physical processes and endocrine systems involved in the course of the motivational process. The only sensible way to speak of altruistic emotion is to say that it is an emotion triggered by an understanding of others' needs or well-being. Compassion clearly meets this criterion.

This leads us to the second objection, which questions the importance of the other-directed component of apparently altruistic emotions. Here the picture becomes a little more complicated. According to such a view, *apparently* altruistic emotions are in fact triggered by a combination of other-regarding perceptions reliably associated with self-regarding perceptions. Furthermore, only the latter are necessary ingredients for emotions to occur, therefore, only the latter ground motivation. Denise, for example, would only start to feel compassion once she had understood that the child was in need for help – box P1 – *and* that this situation was not advantageous for her – box P2. Both perceptions are needed for compassion – box E1 – to arise.

It is worth noting that these perceptions need neither be conscious nor conceptually well articulated: they can be simple apprehensions of relevant aspects of the situation observed. There is no need for inferential reasoning here; a simple mechanistic association of thoughts of the sort described by Damasio in his somatic marker hypothesis (1995) can trigger an emotion. Of course, if self-directed perceptions are needed for emotions to occur, there cannot be altruistic emotions – at least not of the sort wanted by an advocate of PA.

Moreover, another interesting aspect of such an account is that it is not always necessary to postulate the existence of self-directed *motives* – box E2 – in order to explain an action in

terms of self-interest. Consider the case of the mother who sees her child endangered by a lion. To make her act out of a caring emotion, it is sufficient for her to have two preliminary perceptions: 'my child is in danger' and 'my child being in danger is not good for me' – causal path P1– P2–E1– Action.

To recapitulate, according to PE, any emotion – including caring emotions – can only be elicited once the subject has taken her personal interests into consideration. More particularly, the necessary eliciting ingredients for *apparently* altruistic emotions are: a real situation in which an individual is in need, a corresponding perception about that individual and an additional self-directed perception of the sort 'this situation is against my interests'. Without the additional perception, a caring emotion simply cannot be experienced, which makes altruistic emotions impossible.

To respond to this 'egoistic' view, one can refer back to Sober and Wilson. According to these authors, the only way to ground psychological altruism is to use an evolutionary line of argument. Their strategy is to focus on the evolutionary *proximate mechanisms* that cause apparently altruistic behaviours. For them, there are good evolutionary reasons to think that highly social actions, such as human parental care, are set off by other-regarding proximate mechanisms instead of self-directed mechanisms. This assertion is based on what I will name the 'reliability argument' (1998: chap. 10).

A preliminary remark is needed here. Sober and Wilson's original argument was formulated in the context of considering the possibility of primary altruistic *desires*. So, one might think that it is not relevant in a reframed context that focuses on emotions instead of motives. However, when one looks more closely at the details of their argument, it appears that the notion of desire does not play a significant role after all. This will allow me to reformulate the reliability argument in a discussion about emotions.

In fact, I am convinced that Wilson and Sober's argument has even greater relevance in the reframed context proposed here. Their evolutionary argumentative strategy based on desires was not received with much enthusiasm and has encountered numerous objections (Brunero 2002, Dale 2002, Rottschaefer 2002, Stich 2007). Readers' scepticism comes partly from the fact that, in focusing on articulated desires, Sober and Wilson overlook other possible proximate mechanisms responsible for caring behaviour, such as simple encapsulated input-output systems. As Dale Jamieson points out, even non-psychological mechanisms could do the job: "Parental care behaviour is widely dispersed across species, and it is likely that it

occurs in many organisms that are not minded at all” (2002: 703). This said, the sort of mechanisms producing other-directed behaviour that have evolved in social species capable of feelings and complex mental activity are very likely to be *psychological mechanisms*. The real question is what sort of psychological mechanisms they are and what level of complexity they can reach. Sober and Wilson think that they are primary desires “produced by natural selection” (1998: 303) and the evolutionary explanation they provide is based on simple replicator-based models.

I am rather critical of the idea of altruistic motives as pure results of evolution. Of course, from an evolutionary perspective, a type of motive reveals a proximate mechanism, but not all proximate mechanisms can be given convincing evolutionary explanations – at least not with the present state of scientific knowledge. For me, the very idea of considering particular types of motives as direct results of replicator dynamics is questionable.¹⁰ To enter into this complex debate would push us too far, but it is worth keeping in mind how difficult it is to reliably account for cultural products – such as types of desires or intentions – with the mere use of evolutionary tools. It is less controversial to provide evolutionary explanations for basic and easily observable psychological mechanisms, such as simple emotions. To say that these evolved psychological mechanisms have a causal influence on people’s motives is uncontroversial. This is the less speculative line of reasoning that I propose to take in the revised ‘reliability argument’ that I shall now present.

We have seen that at least some ‘apparently’ altruistic actions seem to be mediated by emotional systems. Let us take for granted that such systems exist and are the results of evolution; basic parental love is a typical example of adaptive emotional system (Lazarus & Lazarus 1994). It is fairly easy to understand that caring behaviour’s biological function is to enhance the number of fit offspring who survive to adulthood. In an environment where competition is intense and resources are scarce or difficult to reach, parents need to develop capacities to respond quickly to the necessities of their offspring. In a species capable of feelings and minimal cognition, a quick motivational proximate mechanism such as parental

¹⁰ It might be possible to provide a more differentiated evolutionary explanation of motives by resorting to complex models of cultural evolution and the Baldwin effect (Ananth 2005). This is surely the most constructive way of trying to explain the evolution of fine-grained proximate mechanisms underlying particular types of desires. However, the task is not easy – if not impossible – because of the many intricate parameters that have to be considered; explanatory complexity often comes at the expense of clarity.

love – or a set of caring emotions – is an excellent response. We now need to ask whether it makes sense to expect that this system has evolved in a self-directed form rather than an other-directed form. Here, we have two competing emotional mechanisms, an altruistic and a self-directed one, and the question remains which of them is responsible for the occurrence of caring behaviour whenever a child is in need.

In principle, it is possible that both mechanisms have evolved; evolution does not always exclude redundancy. However, if two motivational mechanisms are both capable of generating the same type of behaviour, one of them might be more likely than the other to be selected. There are good reasons to think that this is precisely what has happened in the present case. As Sober and Wilson point out, one important selection criterion is the *reliability* of a system: among various mechanisms, the most reliable – that is the one that realises its function with the greatest probability – is much more likely to evolve.

Let us compare our two competing emotional mechanisms for reliability. Consider first the self-directed mechanism. Recall that, according to PE, in order for a subject to feel a caring emotion towards his children, three ingredients are needed: the children must be in need of help; the subject must have a corresponding perception of the sort “my children are in need of help”; the subject must have an additional self-interested perception of the sort “this situation is not good for my interests”. Unless these three conditions are met, the motivational mechanisms will not be put into motion and the subject will not engage in parental care at all – which is not desirable from the evolutionary point of view. In contrast, the altruistic motivational mechanism is much simpler. In order for a subject to be motivated to care for his children, only two ingredients are needed: the children must be in need of help and the subject must have a corresponding perception of the sort “my children are in need of help”.

There is evidence that the simplest or more direct of two competing strategies is likely to do a more reliable job than the more complex, indirect one. This remark is especially relevant in the present case because it is not clear at all why the thought association postulated by PE would have occurred in the first place. Moreover, it seems that the process underlying self-interested parental care is quite vulnerable to disruption. If an individual fails to have the self-directed associated perception, the link towards action is broken and he will fail to care for his children. If these cases of imperfect correlation happen regularly – a very plausible hypothesis – natural selection will be likely to opt for the alternative altruistic mechanism.

In brief, the altruistic emotional mechanism seems much more reliable than the self-directed one and is hence more likely to have evolved; in the light of evolutionary considerations, it

does not make sense to expect only self-directed emotions to result from natural selection processes.

The motivational power of altruistic emotions

We have seen that there are good evolutionary reasons to think that highly social actions, such as human parental care, are set off by purely other-regarding emotional proximate mechanisms, rather than with help of self-directed emotional mechanisms. Evolution has influenced motivational mechanisms in such a way that parents typically react altruistically towards their children via caring emotions. Since a single counterexample is sufficient to reject PE, PA seems to be the adequate way to explain altruistic action.

However, even if caring emotions are genuinely altruistic, the question remains whether they are strong enough to set off action. Indeed, one might still contend that altruistic emotional motivation always conflicts with other self-directed motivations and that the latter are stronger.

Again, a simple evolutionary argument enables a response to this objection. Emotions are proximate behavioural mechanisms. Without doubt, some altruistic emotions – such as compassion towards one's children – exist and are adaptive. The evolution of psychological mechanisms, such as basic emotions, is best explained in terms of the behavioural impact of these mechanisms; they have been selected *because* the behavioural propensities they induce are usually beneficial in terms of fitness to the subjects who possess them. Therefore, one can be sure that at least some altruistic emotions are causally efficacious. Simple altruistic emotional mechanisms would not have been selected if they did not have a behavioural impact. Besides, many empirical researchers in behavioural psychology have demonstrated the effect of empathic emotions on behaviour (see for example Batson 1991, Eisenberg 2006). This is enough to resolve the famous philosophical controversy in favour of psychological altruism.

Conclusion

I have argued that the altruism versus egoism controversy reaches a deadlock as soon as one makes use of the unconscious argument. In order to loosen this deadlock, I have proposed a shift away from an over-intellectualisation of the proximate motivational mechanisms responsible for altruistic action. Instead, I suggest a move towards an emotional account of

altruistic decision-making. In the context of the controversy over altruism, this move proves fruitful because it allows the debate to focus on self-directed versus altruistic emotions. This focus provides firm ground for a defence of PA; evolutionary arguments in favour of the existence of motivating altruistic emotions are sufficient to convincingly argue against PE. This conclusion depends on the acceptance of a shift of perspective from motive to motivation, which leads to a revised motivational causal chain beginning with simple affective reactions such as emotions, rather than motives. Incidentally, I also hope to have shown that Sober and Wilson's reliability argument assumes full relevance in the proposed reframed context.

Acknowledgements are due to Philip Kitcher who helped me improving my proposal to reframe the debate, Christian Maurer who drew my attention to Hutcheson's arguments in favour of altruism, and Rebekka Klein with whom I discussed the distinction between motive and motivation. Many thanks as well to Chloë FitzGerald for correction, advice, and comments on previous versions of this paper.

Bibliography

- Ananth, Mahesh (2005), "Psychological Altruism Vs. Biological Altruism: Narrowing the Gap with the Baldwin Effect", *Acta Biotheoretica*, 53, pp. 217-39.
- Batson, C. Daniel (1991), *The Altruism Question: Toward a Social Psychological Answer*. Hillsdale, N.J.: L. Erlbaum.
- Batson, C. Daniel (2000), "Unto Others: A Service... And a Disservice", *Journal of Consciousness Studies*, 7, pp. 207-10.
- Brunero, John, S. (2002), "Evolution, Altruism And "Internal Reward" Explanations", *The Philosophical Forum*, 33, pp. 413-24.
- Butler, Joseph (1991), "Fifteen Sermons", in D.D. Raphael (eds.), *British Moralists, 1650-1800 : Selected and Edited with Comparative Notes and Analytical Index*. Oxford: Clarendon Press, pp. 325-77.
- Cabanac, Michel, Guillaume, Jacqueline, Balasko, Marta & Fleury, Adriana (2002), "Pleasure in Decision-Making Situations", *BMC Psychiatry*, 2, p. 7.
- Cialdini, Robert B., Schaller, Mark, Houlihan, Donald, Arps, Kevin, Fultz, Jim & Beaman, Arthur L. (1987), "Empathy-Based Helping: Is It Selflessly or Selfishly Motivated?" *Journal of Personality and Social Psychology*, 52, pp. 749-58.
- Clavien, Christine (forthcoming), "Jugements Moraux Et Motivation À La Lumière Des Données Empiriques", *Studia philosophica*.
- Dale, Jamieson (2002), "Sober and Wilson on Psychological Altruism", *Philosophy and Phenomenological Research*, 65, pp. 702-10.
- Damasio, Antonio R. (1995), *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Avon Books.
- de Sousa, Ronald (2004), "Emotions: What I Know, What I'd Like to Think I Know, and What I'd Like to Think", in R.C. Solomon (eds.), *Thinking About Feeling:*

- Contemporary Philosophers on Emotions.* Oxford; New York: Oxford University Press, pp. 61-75.
- Döring, Sabine A. (2003), "Explaining Action by Emotion", *The Philosophical Quarterly*, 53, pp. 214-30.
- Eisenberg, Nancy (2006), "Empathy-Related Responding and Prosocial Behaviour", in G. Bock & J. Goode (eds.), *Empathy and Fairness*, pp. 71-88.
- Fehr, Ernst & Fischbacher, Urs (2004a), "Social Norms and Human Cooperation", *Trends in Cognitive Sciences*, 8, pp. 185-90.
- Fehr, Ernst & Fischbacher, Urs (2004b), "Third-Party Punishment and Social Norms", *Evolution and Human Behavior*, 25, pp. 63-87.
- Fehr, Ernst & Gächter, Simon (2002), "Altruistic Punishment in Humans", *Nature*, 415, pp. 137-40.
- Fehr, Ernst & Rockenbach, Bettina (2003), "Detrimental Effects of Sanctions on Human Altruism", *Nature*, 422, pp. 137-40.
- Fischbacher, Urs, Gächter, Simon & Fehr, Ernst (2001), "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", *Economics Letters*, 71, pp. 397-404.
- Ghiselin, Michael T. (1974), *The Economy of Nature and the Evolution of Sex*. Berkeley: University of California Press.
- Haley, Kevin J. & Fessler, Daniel M. T. (2005), "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game", *Evolution and Human Behavior*, 26, pp. 245-56.
- Henrich, Joseph Patrick (2004), *Foundations of Human Sociality : Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
- Hutcheson, Francis (2004), *An Inquiry into the Original of Our Ideas of Beauty and Virtue : In Two Treatises (1725)*. Natural Law and Enlightenment Classics. Indianapolis, Ind.: Liberty Fund.
- Jamieson, Dale (2002), *Morality's Progress: Essays on Humans, Other Animals, and the Rest of Nature*. Oxford, New York: Oxford University Press.
- Lazarus, Richard S. & Lazarus, Bernice N. (1994), *Passion and Reason: Making Sense of Our Emotions*. New York: Oxford University Press.
- Macpherson, Crawford B. (1962), *The Political Theory of Possessive Individualism: Hobbes to Locke*. Oxford: Clarendon Press.
- Marwell, Gerald & Ames, Ruth E. (1981), "Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods", *Journal of Public Economics*, 15, pp. 295-310.
- Ostrom, Elinor (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Rilling, James. K., Gutman, David A., Zeh, Thorsten R., Pagnoni, Giuseppe, Berns, Gregory S. & Kilts, Clinton D. (2002), "A Neural Basis for Social Cooperation", *Neuron*, 35, pp. 395-405.
- Robinson, Jenefer (2005), *Deeper Than Reason: Emotion and Its Role in Literature, Music, and Art*. Oxford: Oxford University Press.
- Rottschaefler, William A. (2002), "It's Been a Pleasure, but That's Not Why I Did It", in L.D. Katz (eds.), *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives*. Thorverton: Imprint Academic, pp. 239-43.
- Slote, Michael Anthony (1964), "An Empirical Basis for Psychological Egoism", *The Journal of Philosophy*, 61, pp. 530-37.

- Sober, Elliott & Wilson, David Sloan (2000), "Morality and `Unto Others. Response to Commentary Discussion", *Journal of Consciousness Studies*, 7, pp. 257-68.
- Sober, Elliott & Wilson, David Sloan (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, Mass.: Harvard University Press.
- Stich, Stephen P. (2007), "Evolution, Altruism and Cognitive Architecture: A Critique of Sober and Wilson's Argument for Psychological Altruism", *Biology and Philosophy*, 22, pp. 267-81.