

# Relational Desires and Empirical Evidence against Psychological Egoism

Joshua May

*European Journal of Philosophy* Vol. 19, No. 1 (2011), pp. 39–58.

[DOI: 10.1111/j.1468-0378.2009.00379.x](https://doi.org/10.1111/j.1468-0378.2009.00379.x)

This is the penultimate draft;  
any citations should be to the final, published version.

**Abstract:** Roughly, psychological egoism is the thesis that all of a person's intentional actions are ultimately self-interested in some sense; psychological altruism is the thesis that some people's intentional actions are ultimately other-regarding in some sense. C. Daniel Batson and other social psychologists have argued that there are experiments that provide support for a theory called the “empathy-altruism hypothesis,” which entails the falsity of psychological egoism. However, several critics claim that there are egoistic explanations of the data that are still not ruled out. One of the most potent criticisms of Batson comes from Elliott Sober and David Sloan Wilson. I argue for two main theses in this paper: (1) we can improve on Sober and Wilson's conception of psychological egoism and altruism, and (2) this improvement shows that one of the strongest of Sober and Wilson's purportedly egoistic explanations is not tenable. A defense of these two theses goes some way toward defending Batson's claim that the evidence from social psychology provides sufficient reason to reject psychological egoism.

## 1. Introduction

It's uncontroversial that some people sometimes help others. So, one can hardly deny that there are what one might call “helpful actions.” But one can still help another without being altruistic. One could have an ulterior motive: one's ultimate goal could be to benefit oneself, not the other person. Thus, we wonder: When someone helps another person, is it always ultimately motivated by self-interest? Fairly recently there has been a resurgence of interest in this psychological egoism-altruism debate. There are two opposing positions on the issue. Roughly, *psychological egoists* claim that all of a person's intentional actions are ultimately self-interested or selfish in some sense; *psychological altruists* claim that some people's intentional actions are not ultimately self-interested, since some are ultimately other-regarding.

The (psychological) egoism-altruism debate is a difficult one to settle. Consider a hypothetical case. Suppose Omar helps Kima, a complete stranger, by

informing her that the floor is slippery up ahead. Was the ultimate motivation for Omar's action simply to help Kima, or was it to achieve some benefit for himself? The psychological egoist claims that Omar's ultimate motivation is really self-serving; he helped only to achieve some self-interested goal. Perhaps his ultimate goal was to avoid the feelings of guilt that would surely follow if he didn't tell her about the slippery floor and she fell and hurt herself as a result. Or perhaps his ultimate goal was to gain a good feeling that he figured he could gain by helping her out. On the other hand, the psychological altruist claims that at least sometimes the ultimate motivation for performing such actions is *not* for one's own benefit. On this view, we needn't assume that Omar's ultimate motive was self-interested; it's at least psychologically possible for his ultimate goal to be to help Kima, a goal that is not just a means to helping himself in some way.

Over the past several decades much attention has been drawn in various fields to experimental research in social psychology that has been conducted on this issue. The key figure, C. Daniel Batson, argues that the data from the experiments provide sufficient evidence against psychological egoism by supporting an anti-egoistic theory called the "empathy-altruism hypothesis." However, several critics—including philosophers, psychologists, and biologists—have argued that Batson's conclusions are not justified. Two prime critics are Elliott Sober and David Sloan Wilson, who argue in their book *Unto Others* (1998) that there are egoistic alternatives that can equally explain the data. While Sober and Wilson do ultimately argue that psychological egoism is false, they think evolutionary theory provides the evidence for this, not traditional philosophical arguments or the psychological arguments based on the empirical evidence.

I have two main theses I want to defend in this paper. First, Sober and Wilson's conception of the egoism-altruism debate is not quite correct. Second, improving on the conception of the debate shows that one of Sober and Wilson's purportedly egoistic explanations of the data does not support psychological egoism. A defense of these two theses goes some way toward defending Batson's claim that the evidence from social psychology provides sufficient reason to reject psychological egoism. I begin (in §2) by summarizing the conceptual framework for the egoism-altruism debate that has been developed primarily by Sober and Wilson. In defense of my first thesis, I subsequently argue (in §3) that we must incorporate what I call "relational desires" and "essential beneficiaries" into the debate. I then turn to defending my second thesis. To this end, I review the relevant experiments (in §4) and Sober and Wilson's criticisms of the argument that Batson bases on them (in §5). Equipped with the improved conception of the debate, I show that Sober and Wilson's supposedly egoistic explanation of the data—what I will call the "modified negative-state relief hypothesis"—is incompatible with psychological egoism.

## 2. A Conceptual Framework for Egoism and Altruism

In order to understand the claims of the psychological egoist versus those of the psychological altruist, some conceptual groundwork must be laid. Pre-theoretically,

a person acts altruistically if she acts with an irreducible concern for the welfare of another. Thus, the ordinary concept of altruism applies to actions. But whether a given action is altruistic depends on whether one performs the action with a certain goal—namely of benefiting another—that is not reducible or instrumental to any other goal. Thus, as Batson and Sober and Wilson make clear, altruism is fundamentally a property of the motive, goal, or motivational state that underlies the action. Furthermore, following Thomas Nagel (1970), they recognize that, while we ordinarily only apply the term “altruism” and its cognates to especially heroic or self-sacrificing acts that are fairly atypical, the egoism-altruism debate concerns the underlying motives of all helpful actions.

With these basic constraints in mind, Sober and Wilson (1998) define *psychological egoism* as the view that “people desire their own well-being, and nothing else, as an end in itself” (p. 224). They then define *psychological altruism* as the view that “people sometimes care about the welfare of others as an end in itself” (p. 228). However, to make their characterizations more precise, they introduce the terms “self-directed” and “other-directed.” They then define the theses in these terms. They state that *psychological egoism* “maintains that the only ultimate goals an individual has are self-directed” (p. 224) while *psychological altruism* maintains that “we [sometimes] have other-directed ultimate desires” (p. 229).

While Sober and Wilson focus on defining the positions on the debate, Batson (1991) focuses on defining the motivational states at issue. He writes that *egoism* “is a motivational state with the ultimate goal of increasing one’s own welfare” (p. 7) while *altruism* “is a motivational state with the ultimate goal of increasing another’s welfare” (p. 6). Presumably Batson would agree that the *thesis* of psychological egoism is that all motivational states are a form of what he calls “egoism” and the *thesis* of psychological altruism is that there are some motivational states that are a form of what he calls “altruism.” My main concern is which thesis is true, but it will be easiest and most natural to provide non-substantive definitions of the theses in terms of the states and give substance to these definitions by providing a substantive characterization of the relevant states that figure in the theses. However, to avoid confusing the theses with the types of state, I will depart from Batson’s terminology here and say that a motivational state is “egoistic” or “altruistic.” I will work with the following non-substantive definitions of the theses, which should be uncontroversial:

Psychological Egoism: All of a person’s ultimate desires are *egoistic*.

Psychological Altruism: Some of a person’s ultimate desires are *altruistic*.<sup>1</sup>

---

<sup>1</sup> Even though it is the motivational state that ultimately determines whether a person’s action is altruistic, one might think that only a motivational state that does in fact move one to act is either a form of altruism or not. (Consider someone who has an altruistic motive and an egoistic motive, but acts on the egoistic motive. Was there at all an instance of altruism in that case? Perhaps not.) If this is so, then we can alter the relevant definitions accordingly. Such modifications won’t affect any of the arguments in this paper.

Substance will be added to these definitions by characterizing what it is to be an egoistic or altruistic desire.

But before that, there are a few things to note about using the term “desire” here. The relevant mental item at issue in the egoism-altruism debate is what one might ordinarily call a “motive,” “goal,” or perhaps “intention.” Many, however, use more technical terms. Batson, for example, tends to use “motivational state,” while Sober and Wilson (and many other philosophers) use the term “desire.” I will, for the most part, follow Sober and Wilson in using the term “desire,” primarily because it is a common term to use among contemporary philosophers and it allows for making explicit that the relevant mental items have content (by using such technical locutions as “*S* desires that *p*”). However, the use of the term “desire” in this debate should not be taken as a commitment to the so-called “Humean theory of motivation.” We needn’t take a stance on such issues. For our purposes, we simply need mental states with certain features, such as motivating (at least in part) an agent to act, having intentional content, being goal-directed, and so on. Throughout this paper I will use the term “desire” to designate what is at issue in the egoism-altruism debate (namely, a motive, goal, etc. of an individual).

So which desires are egoistic and which are altruistic? While few have provided a precise answer to this question, Sober and Wilson do. They do this by characterizing what it is to be an altruistic or egoistic desire in terms of self-directed and other-directed desires. Roughly speaking, one’s *self-directed desires* are those desires that only concern oneself, whereas one’s *other-directed desires* are those desires that only concern others. The two can be defined more technically as follows:<sup>2</sup>

#### Self-Directed Desires

*S*’s desire that *p* is *self-directed* if and only if: (a) *p* contains a representation of oneself as such, and (b) *p* does *not* contain a representation of someone as distinct from oneself.

#### Other-Directed Desires

*S*’s desire that *p* is *other-directed* if and only if: (a) *p* does *not* contain a representation of oneself as such, and (b) *p* contains a representation of someone as distinct from oneself.

For example, Reginald’s desire that his business partner, Sherad, learns arithmetic is other-directed, because the content of Reginald’s desire contains a representation of someone and that someone is represented as distinct from himself. But Reginald’s desire to no longer be addicted to heroin (i.e. his desire that *he* is no longer addicted to heroin) is self-directed, since the content of his desire represents someone and that someone is represented as himself (i.e. represented first-personally).<sup>3</sup>

---

<sup>2</sup> These definitions are based on Sober and Wilson (1998, p. 225), though I have tried to make them more precise. They have been greatly improved by discussions with Aaron Zimmerman.

<sup>3</sup> On the importance of first-personal representations of oneself here, see Sober and Wilson (1998, p. 214).

Notice that the above definitions take desires to be mental states with objects or contents. Following Sober and Wilson, I will take beliefs and desires to have propositional content. So, for example, Thomas's desire to eat some coffee ice cream is at least roughly equivalent to his desiring *that he has some coffee ice cream*, where the desire is in some sense *satisfied* when the relevant proposition (i.e. that he eats some coffee ice cream) is true. As Sober and Wilson make clear, the sense of "satisfaction" here is not in the phenomenological sense of feeling satisfied. Rather, it is in the sense of the desire being fulfilled—that is, the content of the desire being true. We can follow Scanlon (1998, p. 41) in calling this the "logical sense" of satisfaction or fulfillment.

We must lastly include in an appropriate way the concept of one's well-being, benefit, or welfare. As Sober and Wilson (1998, p. 229) point out, a person who has an ultimate desire for another's death is not altruistic, even though the content of the ultimate desire is other-directed. Likewise, consider an ultimate desire to snap one's fingers from time to time (where the agent doesn't conceive of this as benefitting anyone). Such desires are neither altruistic nor egoistic. So, we must recognize that the psychological egoist claims that all of our ultimate desires concern one's own *benefit* while the psychological altruist claims that some of our ultimate desires concern the benefit of another.<sup>4</sup>

We are now in a position to understand the two types of motivational state at issue, at least as far as Sober and Wilson are concerned. Recall that psychological egoism is the thesis that all of one's ultimate desires are *egoistic*; psychological altruism is the thesis that some of one's ultimate desires are *altruistic*. Sober and Wilson supply a way to add substance to our non-substantive definitions by providing an account of what makes a desire egoistic or altruistic. Their account centers primarily on the distinction between self-directed and other-directed desires and the notion of a person's benefit:

Egoistic Desires (v.1)

*S*'s desire that *p* is *egoistic* if and only if: (a) *p* concerns the benefit of some person, (b) *S*'s desire is self-directed, and (c) *S*'s desire is ultimate.

Altruistic Desires (v. 1)

*S*'s desire that *p* is *altruistic* if and only if: (a) *p* concerns the benefit of some person, (b) *S*'s desire is other-directed, and (c) *S*'s desire is ultimate.

I take these definitions of the relevant states combined with the relevant theses to accurately represent how Sober and Wilson, if not many others as well, characterize the egoism-altruism debate. While I think these definitions are close to an accurate representation of the debate, there is a problem once we consider ultimate desires

---

<sup>4</sup> Sober and Wilson are not always careful on this point. Sometimes they don't include the crucial notion of well-being in their definitions of psychological egoism and altruism. For example, they sometimes simply say: "Egoism claims that all of our ultimate desires are self-directed; altruism, that some are other-directed" (2000a, p. 197). See also their "more precise" statements I quoted toward the beginning of this section.

with contents that contain representations of both oneself and another—what I will call “relational desires.”

### 3. Improving the Conception of the Debate

#### 3.1 Relational Desires

Sober and Wilson (1998) recognize the issue of relational desires, calling them “desires that mention both self and other” (p. 225), but they refrain from construing psychological egoism or altruism as making any claims about them. They contend that such a construal is “difficult to justify” and may “bias the case in favor of one position over the other” (p. 226). They propose to simply admit a third view into the taxonomy, separate from both psychological egoism and psychological altruism, called *relationism*—“the view that people sometimes have ultimate desires that certain relational propositions (connecting self and specific others) be true” (p. 226). Fortunately, they invite anyone who “thinks that some cases of relationism are properly viewed as subspecies of altruism or of egoism... to adjust the conceptual taxonomy” since their “assessment of these theories will not be affected by such amendments” (p. 226). I intend to make the case that we should adjust the conceptual taxonomy and that this does affect Sober and Wilson’s discussion of the psychological experiments.<sup>5</sup>

Sober and Wilson (1998, pp. 225-6) in fact provide brief examples of ultimate relational desires, yet reject them too quickly as ultimately inessential to the debate. Building on one of their examples, consider the following case.

#### Stolen Kiss

Milhouse is enamored with Lisa. He thinks about her constantly and finds himself ultimately desiring to kiss her. (It’s not that he just wants to feel good and thinks kissing her is a means to this ultimate end. Rather, he simply has the relevant ultimate desire.) However, he doesn’t at all care how Lisa feels about it. He’d like just to “steal” a kiss from her. One day he throws himself upon Lisa and kisses her promptly on the lips. Lisa, who is shocked and a bit annoyed by the whole incident, pushes Milhouse away. Though Lisa is rather upset, Milhouse is pleased; he got just what he wanted.

Milhouse’s ultimate desire will be satisfied only if *he* benefits from the kiss, but whether Lisa benefits is not one of the conditions required for the satisfaction of his desire—what one might call the “*satisfaction conditions*” of the desire.<sup>6</sup> Given that

---

<sup>5</sup> The idea that Sober and Wilson cannot remain neutral on whether certain relational desires are altruistic was brought to my attention by Aaron Zimmerman. My discussion of this issue has been greatly aided by valuable discussions with him.

<sup>6</sup> Stipulating that Milhouse’s benefit is a satisfaction condition of his ultimate desire probably requires characterizing the propositional content of his desire in a more specific way, such as the following: *I have an enjoyable kiss with Lisa*. However, for convenience I often will spell these details out in the form of a more detailed description of what the satisfaction conditions of the desire in the case are, rather than in the form of a long and complicated proposition that constitutes the content of the desire.

Milhouse is ultimately concerned with his own benefit, not Lisa's, his desire—though relational—should be classified as egoistic. It's the kind of desire whose existence is perfectly compatible with the thesis of psychological egoism. Such examples suggest that we must broaden our characterization of egoistic desires to capture certain relational desires.

A similar case can be made that some relational ultimate desires are altruistic. Consider another scenario.

### Single Savior

While driving, Nate sees a car veer off the road up ahead, hit a bump, and flip in the air. As he pulls up to the scene, he sees that a young woman, Brenda, is trapped under the vehicle. Without thinking, he immediately rushes to her aid, motivated by an ultimate desire to help her (i.e. an ultimate desire that *he* help her). (He doesn't first ultimately desire that she is helped by someone or other and then form the instrumental desire that *he* help her because he sees that no one else is around to do it—others are around as well. Rather, he is just immediately and ultimately motivated to run over and save her himself.) In the end, Nate successfully helps Brenda out from under the car, and he is relieved.

Although Nate's ultimate desire is clearly relational, I submit that it is altruistic. After all, Nate is not ultimately concerned with his own well-being at all; he's concerned with someone else's. In other words, the well-being of someone else is a satisfaction condition for his ultimate desire. That Nate is represented in the content of his own desire seems rather irrelevant to whether the desire is altruistic. Thus, it is incorrect to claim that only other-directed ultimate desires are altruistic.

Some relational desires, then, are egoistic and some are altruistic. And this is what we should expect. Sober and Wilson's characterization of such desires was in terms of who is represented in the contents of the agent's desire. But who is represented is rather unimportant for whether the desire is selfish in the sense relevant to the egoism-altruism debate. The significance of psychological egoism is that, if true, it yields that we're all "deep down" out to benefit ourselves—that all of our ultimate desires concern our own personal gain. Yet, on Sober and Wilson's characterization of the debate, psychological egoism wouldn't be true even if psychologists discovered that all our ultimate desires were relational yet selfish in the way Milhouse's desire to kiss Lisa is. This is because they think such relational desires are neither egoistic nor altruistic. Likewise, on Sober and Wilson's view, if it's possible to have ultimate relational desires like Nate's, then psychological egoism wouldn't be defeated—even though Nate's desire involves a non-instrumental concern for Brenda's benefit. This gets the debate wrong. In order to get it right, we must broaden the definitions of both psychological egoism and psychological altruism so that they make the appropriate claims about relational desires.

### 3.2 Essential Beneficiaries

Now, of course, the crucial question is: Which relational desires are compatible with psychological egoism versus psychological altruism? The answer is already becoming clear. As we have seen, the claim that the contents of all of one's ultimate desires represent oneself as the person who benefits is crucial to psychological egoism. That is, psychological egoism claims that we are always represented as what I will call an "essential beneficiary" in the contents of our ultimate desires. We can more technically define an essential beneficiary in the following way:

#### Essential Beneficiary

A subject  $S^*$  is an *essential beneficiary* of  $S$ 's desire that  $p$  if and only if: it is a necessary condition for the satisfaction of  $S$ 's desire that  $S^*$  benefits.

One way, perhaps, to understand the notion of a satisfaction condition here is the following counterfactual: if  $S$  were to believe that  $S^*$  would not benefit from  $p$ 's being true, then  $S$  would not desire that  $p$ . However, we must be careful here. One way to read this counterfactual (or the word "necessary" in the definition provided) is such that the other's benefit is necessary but not somehow represented in the content of the agent's ultimate desire. This, however, is not the intended reading. Suppose, for example, that Jimmy has an ultimate desire with the following content: I have an enjoyable kiss with Beatie. But suppose that it just so happens that the only way for him to benefit, given the circumstances, is if she enjoys the kiss as well. On my account, Beatie is *not* an essential beneficiary, because Jimmy does not represent her benefit as essential, even though in some sense her benefit is essential to his. So, someone is an essential beneficiary only if that person's benefit is represented as essential in the content of the agent's desire.

Whether a desire is egoistic or altruistic seems to turn on whether the agent is represented as an essential beneficiary. In *Stolen Kiss*, Milhouse's desire is clearly egoistic because he is the only essential beneficiary in the content of his ultimate desire. And in *Single Savior*, Nate's ultimate desire is clearly altruistic because someone other than Nate himself is an essential beneficiary. But what about cases in which both self and other are represented as essential beneficiaries? Consider a third case:

#### Tennis Game

Liz is preparing to play a game of tennis with her friend, Tracy, and ultimately desires to have a mutually enjoyable game with him. (Since her desire is ultimate, she doesn't want to have a mutually enjoyable game simply as a *means* to achieving good feelings or some other benefit.) Throughout the game, her behavior is guided by her ultimate goal: she tries to not embarrass her opponent with her superior skills but also to be challenging enough so that the game is enjoyable for them both. They both have a good time, and Liz is pleased.



In this case, both the agent (Liz) and another (Tracy) are represented as essential beneficiaries. Is such an ultimate desire egoistic? I think not. The benefit of someone other than the agent is represented as essential in this ultimate desire. Since the existence of such a desire entails that someone can ultimately desire the welfare of another, it is incompatible with psychological egoism. But is such a desire *altruistic*? This is certainly a more difficult question to answer. Since we are using “egoistic” and “altruistic” as technical terms here, we cannot appeal to linguistic intuition. We must consider whether such a desire is compatible with the thesis of psychological altruism at issue in the debate. While I will assume that such desires are altruistic, for the purposes of this paper, we need only hold the weaker claim that they are *not egoistic*.<sup>7</sup> We’ll return to this issue later.

Egoistic and altruistic desires can now be redefined in light of these considerations. Roughly speaking, *egoistic desires* are ultimate desires in which oneself is represented as the only essential beneficiary; *altruistic desires* are ultimate desires in which another person is an essential beneficiary. More technically:

Egoistic Desires (v. 2)

*S*’s desire that *p* is *egoistic* if and only if: (a) *S*’s desire is ultimate, (b) *p* contains a single essential beneficiary *S*\*, and (c) *S*\* is represented as *S*’s self.

Altruistic Desires (v. 2)

*S*’s desire that *p* is *altruistic* if and only if: (a) *S*’s desire is ultimate, (b) *p* contains at least one essential beneficiary *S*\*, and (c) *S*\* is represented as distinct from *S*’s self.

This account of altruistic desires admittedly relies on the stronger claim that the desire in Tennis Game is altruistic. However, for the second thesis of this paper, we only need an account of egoistic desires. And the account I’ve provided here only relies on the weaker claim that the desire in Tennis Game is not egoistic. If the stronger claim is false, we can simply adjust the account of altruistic desires by adding the following clause (d): *S* is not an essential beneficiary.

This construal of the relevant states yields the correct judgments about all the sorts of cases we’ve discussed. It captures all the previous cases, because all self-directed ultimate desires for the benefit of a person will be ones in which oneself is the only individual represented as an essential beneficiary; thus they are egoistic. Likewise, all other-directed ultimate desires for the benefit of a person are altruistic because another individual is represented as an essential beneficiary. Furthermore,

---

<sup>7</sup> Here, briefly, is a defense of the stronger claim: It may be true that other-directed ultimate desires for the benefit of another are in some sense *more* altruistic than certain relational ultimate desires for the benefit of another. But it is nevertheless true that certain relational ultimate desires are altruistic. Here we might want to say that other-directed ultimate desires for the benefit of another are *purely altruistic*, while the corresponding relational desires are not. So, Liz’s ultimate desire that she and Tracy have a mutually enjoyable game of tennis may not be *purely* altruistic, but it is nevertheless an altruistic one. Likewise, we can say that Milhouse’s ultimate desire that he have an enjoyable kiss with Lisa is not *purely egoistic*, but it is nonetheless egoistic.

the new conception of psychological egoism accords with Stolen Kiss, in which Milhouse ultimately desires that he have an enjoyable kiss with Lisa. While Lisa and Milhouse are both essential *constituents* of the content of his desire, the essential *beneficiary* represented is only Milhouse. Similarly, this construal of psychological altruism accords with Single Savior in which Nate ultimately desires that he save Brenda. While both Nate and Brenda are featured in the content of the desire, Nate himself is not an essential beneficiary.

## 4. Psychological Experiments

I hope it is now clear that, contra Sober and Wilson, relationism is neither a sub-species of psychological egoism nor of psychological altruism. Relational desires should be incorporated into both theses. But is this problematic for Sober and Wilson's overall view? I will now attempt to show that it is by applying this new conception of the egoism-altruism debate to one of Sober and Wilson's explanations of some of the data gathered by psychologists. Before evaluating their proposal, however, we must look at the relevant experiments.

### 4.1 An Empirical Approach

The egoism-altruism debate is largely empirical, but not entirely. The competing positions make empirical claims, since they concern how people are actually motivated to benefit others. However, the claims are about the nature of that motivation, which involves more conceptual issues about motivational mental states and their contents. So the issue is subject to both empirical and conceptual investigation. But a problem arises with empirical investigation of the issue: we do not obtain indubitable demonstration of a person's ultimate desires by mere observation. If the issue is pursued empirically, experiments must be devised in which subjects are put into a situation with the opportunity to help another perceived to be in need while certain variables are manipulated to allow experimenters to make reasonable inferences about the underlying motivations. Thus, the empirical investigation will involve testing hypotheses, theories, or explanations of data obtained from such experiments that either support the thesis of psychological egoism or the thesis of psychological altruism. The goal of providing experimental data for the debate, then, only provides confirmation or disconfirmation of explanations of data that are either "egoistic" (in the sense of supporting psychological egoism) or "altruistic" (in the sense of supporting psychological altruism).

Recognizing these points, social psychologists have designed a number of experiments in an attempt to test various versions of psychological egoism and altruism. A key figure in this literature is C. Daniel Batson. Batson (1991) synthesizes the evidence and argues that the experiments provide support for an

altruistic theory: the *empathy-altruism hypothesis* (EAH).<sup>8</sup> As Batson et al. (1988) put it, the hypothesis is that “empathy evokes motivation directed toward the ultimate goal of benefiting the person for whom empathy is felt, not toward some subtle form of self-benefit” (p. 52). Batson proposes to substantiate this hypothesis by showing that the best explanation of the data is the EAH. If Batson is correct, this provides justification for believing that psychological egoism is false since the EAH entails that psychological altruism is true: people do sometimes have ultimate desires for the welfare of others (and this is induced by empathic feelings).

To reach this conclusion, Batson reviews findings from experiments since around the 1960s as well as more recent experiments on the matter, including his own. Around the 60s and 70s, psychologists learned much about the emotion of empathy, including how to effectively induce it in subjects.<sup>9</sup> The term “empathy” here may not line up precisely with how the term is ordinarily used. Batson (1991) states that by “empathy” he means “an other-oriented emotional reaction to seeing someone suffer” (p. 58) that “includes feeling sympathetic, compassionate, warm, softhearted, tender, and the like” (p. 86). Later, psychologists discovered what Batson calls the “*empathy-helping relationship*” (EHR)—experience of relatively high empathy for another perceived to be in need causes people to help that person more than relatively low empathy. But the EHR doesn’t settle the egoism-altruism debate, for the underlying motivation to benefit the other that’s induced by empathic feelings could be either ultimately altruistic or egoistic. Well aware of this, Batson and others have devised experiments to test various egoistic versus altruistic explanations of the EHR.

Given some basic assumptions about how to infer people’s motives from their behavior, psychologists have formed an empirical strategy for addressing the egoism-altruism debate.<sup>10</sup> While there is no doubt that addressing the egoism-altruism debate experimentally is extremely difficult, the empirical approach developed over the past several decades using the EHR as a springboard has proven quite fruitful. By controlling the situations in which people intentionally help or offer to help another person (often called “the needy other”), one can hypothesize about why the subjects ultimately chose to perform that action—what underlying ultimate desires and means-end beliefs contributed to motivating the person to help. But which variables are manipulated in the experiments depends on which egoistic explanation of the data is being tested against the altruistic explanation (i.e. the EAH).

---

<sup>8</sup> For a more concise review of the psychological literature and the argument that it poses a threat to psychological egoism, see Batson & Shaw (1991).

<sup>9</sup> There are various methods for inducing empathy. Two main methods that have been demonstrated to be quite effective are: (a) the Krebs method of describing the needy other as similar to the subject (similar background, sex, life experiences, etc.), and (b) the Stotland perspective-taking method of instructing the subject to imagine what it is like for the needy other (how the other feels, etc.). See Batson (1991, pp. 92-6) for further discussion.

<sup>10</sup> For a more detailed discussion of the background assumptions involved here, see Batson (1991, pp. 64-67), Sober and Wilson (1998, Ch. 6), and Stich, Doris and Roedder (forthcoming).

## 4.2 Testing Egoistic Hypotheses

Competing with the EAH are three main categories of egoistic explanations of the EHR: (i) aversive-arousal reduction, (ii) empathy-specific punishment, and (iii) empathy-specific reward. Each hypothesis claims that empathic arousal (i.e. experience of relatively high empathy) causes people to help only because it induces an egoistic ultimate desire; the desire to help the other is solely instrumental to the ultimate desire to benefit oneself. The *aversive-arousal reduction hypothesis* (AARH) states that empathically aroused individuals help because they ultimately seek to eliminate their empathic arousal, since it is an experience to which people are averse. The second species of egoistic explanation, the *empathy-specific punishment hypothesis* (ESPH), claims that people help more when they are empathically aroused only to avoid negative evaluation from others or themselves. The third species of egoistic explanation, the *empathy-specific reward hypothesis* (ESRH), claims that people help more when they are empathically aroused only to gain benefits or rewards, such as praise (from oneself or others) or a boost in mood.

Each species of egoistic explanation has been tested in multiple experiments against the EAH by various psychologists. Batson argues that the predictions of all these competing egoistic hypotheses are not commensurate with the data and so concludes: “Pending new evidence or a plausible new egoistic explanation for the existing evidence, the empathy-altruism hypothesis, however improbable, seems to be true.” (1991, p. 174). Several commentators have taken issue with this claim regarding a number of the experiments, arguing that egoistic explanations are still in the offing. But here I want to focus on an especially poignant criticism from Sober and Wilson. For our purposes, then, we need only discuss experiments involving the ESRH, since the explanation of Sober and Wilson to be discussed is a version of this third category of egoistic hypothesis.

Testing of the ESRH was fairly complex because there are several versions of the hypothesis that make different predictions. However, for our purposes, we need only discuss the second version of the ESRH: the *negative-state relief hypothesis* (NSRH), initially proposed by Robert Cialdini and colleagues. According to Cialdini et al. (1987), the NSRH states that “an empathic orientation causes individuals viewing a suffering victim to feel enhanced sadness” and “these saddened subjects help for egoistic reasons: to relieve the sadness in themselves rather than to relieve the victim’s suffering” (p. 750). The hypothesis predicts, in other words, that the ultimate desire of empathically aroused individuals (so-called “high-empathy subjects”) is to restore themselves to a normal state. Furthermore, it is assumed that subjects believe the negative-state can be relieved by some distinct varieties of mood-enhancing experience (such as payment or praise) that are comparable to helping the needy other. Psychologists reasoned that this egoistic hypothesis predicts (contrary to its altruistic competitor) that empathically aroused individuals will no longer help in higher frequencies if their mood is restored before they are given the chance to offer help or if they believe their mood will be enhanced in an alternative way if they choose not to help. The results of the experiments were that leading people to believe their mood would be enhanced even if they chose not to

help did not affect how often they volunteered to help, which supports the EAH (for review, see Batson, 1991, pp. 163-74).

## 5. Sober and Wilson's Criticism of Batson

Sober and Wilson argue that a modified version of the NSRH eludes Batson's purported refutation of psychological egoism. Recall that the studies conducted on the NSRH tested whether leading people to believe they will receive a mood-enhancing experience if they choose not to help mitigated the frequency of offers to help. The experiments provide evidence that this is not the case: whether high-empathy subjects believe their mood will soon be enhanced does not diminish the EHR. In response to the studies, Sober and Wilson claim that it is "not surprising that the pain we experience in empathizing with the suffering of others is not completely assuaged by any old pleasant experience.... When we are sad, we usually are sad about something in particular..." (1998, p. 271). So, Sober and Wilson agree that the NSRH is undermined by the experiments, but only because that hypothesis predicts that empathically aroused individuals think that a mood-enhancing experience other than helping the needy other will restore them to a normal state. According to their *modified* negative state relief hypothesis, however, "empathizing with a needy other creates a kind of sadness that subjects know [or at least believe] cannot be assuaged by any old mood-enhancing experience" (2000b, p. 267). Largely because they believe Batson's experiments do not rule out this sort of egoistic hypothesis, Sober and Wilson conclude that "the psychological literature has not established that the egoism hypothesis is false" (1998, p. 272).<sup>11</sup> However, they do not contend that psychological research cannot *ever* adjudicate the debate (although, they do not assert the opposite). They officially "take no stand" on that issue (1998, p. 272).

In response to Sober and Wilson's critique, Batson (2000) insists that this egoistic hypothesis is "of course simply a restatement of the aversive-arousal-reduction explanation that... has already been laid to rest" (p. 210). Batson here is referring to the experiments conducted on whether ease of escape eliminates the EHR. The key difference between the NSRH and the AARH is a minor one having to do with the predictions they make regarding such experiments. The AARH holds that empathically aroused individuals ultimately desire to get rid of the aversive feeling of empathy. Most importantly, the view holds that empathizing with someone perceived to be suffering is like being in the vicinity of an annoying machine—once one leaves the vicinity, the aversive experience subsides. Thus, the AARH makes the disconfirmed prediction that high-empathy subjects will no longer

---

<sup>11</sup> I qualify with "largely" because Sober and Wilson do have one other major criticism of the empirical evidence against psychological egoism. They argue that the so-called "empathic joy hypothesis" (a version of the ESRH) is not ruled out by the experiments conducted on it (1998, p. 268). However, I will not address their criticism here. Interested readers may consult Batson's reply (2000, p. 210) and their rebuttal (Sober and Wilson, 2000b, p. 266).

help in higher frequencies if they believe they can easily exit or escape the situation. The NSRH is quite similar but holds that empathic arousal causes subjects to feel an enhanced sadness that lingers with them for at least awhile so that they figure they will still be sad even if they can no longer directly perceive the needy other. Thus, it doesn't predict that easy escape will reduce helping. Yet Batson doesn't provide any reasons for thinking Sober and Wilson's proposal makes this prediction.<sup>12</sup>

Contra Batson, Sober and Wilson intend their explanation to be a version of the NSRH, and there is a way to read it as such. They're proposing something like the following. As a product of feeling higher levels of empathy, individuals are especially sad. Unlike the AARH, the ability to exit the situation will not significantly diminish the rate at which such individuals offer to help, for they believe that if they simply leave, they will not be restored to a normal state—the sadness will linger. However, the proposal is a modification of the NSRH in holding that not just any mood-enhancing experience will satisfy the ultimate desire of the participants. So the data gathered thus far on the NSRH do not disconfirm this hypothesis. And nothing about this modification forces the claim that easy escape will mitigate helping, which was not borne out in experiments on the AARH. Thus, Batson's response here is deficient.<sup>13</sup>

However, Sober and Wilson's proposal is not without its problems. And this is revealed once we attempt to cash out the details. Recall that, in order to explain the data, Sober and Wilson's proposal hinges on the claim that not just any mood-enhancing experience will mitigate helping. But what are the details of the proposal that underwrite this claim? Will the mood-enhancing experiences not satisfy the desire because the desire is for something more specific or because subjects believe

---

<sup>12</sup> Batson is here assuming, as I will, that only one of the types of egoistic hypothesis may be employed to explain the relationship between empathy and helping. At the very least, he is assuming that the experiments conducted so far are similar enough in setup that we should expect a single, unified explanation—see Batson (2000, p. 210). This is why Batson thinks Sober and Wilson cannot propose an explanation that makes predictions disconfirmed in other experiments, even though they were designed to test other egoistic explanations of the EHR against the EAH. This is an assumption behind the entire research method of Batson and his associates. However, one might worry about, as Robert Cialdini (1991) has, Batson's "one-at-a-time pattern of assault on egoistic alternatives" (p. 125). Even if Batson is right that the EAH is a more plausible explanation of the data than the *particular* egoistic alternative in a *particular* experiment, the psychological egoist might be more pluralistic in her account of what kinds of egoistic motives people have. Perhaps any one of the many egoistic motives posited could be operative in the various experimental settings. However, since Sober and Wilson don't express any qualms with this aspect of Batson's strategy, and since my aim in this paper is only to raise a worry for their view, I needn't defend Batson's assumption here. But interested readers may want to consult Stich et al. (forthcoming) for a brief discussion of doubts about the strength of this sort of objection to Batson.

<sup>13</sup> The labels "NSRH" and "AARH" don't much matter here. Batson takes Sober and Wilson's proposal to be a version of the AARH because of the prediction he thinks it makes regarding easy escape. One might concede this but still think the NSRH is itself a version of the AARH because it involves an aversive feeling. But according to the NSRH, empathically aroused people ultimately desire to have their mood boosted. And while negative feelings play a role in bringing about the ultimate desire for a boost in mood, a boost in mood is a reward in the somewhat technical sense of a personal gain that is often used among psychologists. But, again, the labels aren't important. To rebut Batson's reply, one need only show that Sober and Wilson's proposal doesn't make the disconfirmed prediction about easy of escape.

it can only be satisfied in a certain way? The proposal must modify the NSRH in order to explain the data. But it is unclear how exactly Sober and Wilson intend to modify it. There are several options, but all of them seem to fail.

### 5.1 The First Option: Modify the Belief

Sober and Wilson may want to follow the original NSRH and say that empathically aroused individuals possess an ultimate desire with the following content: *I am restored to a normal state*. Here the strategy is to retain the ultimate desire posited by the NSRH, but modify the relevant means-end belief. Sober and Wilson sometimes seem to indicate that their proposal is to be interpreted along such lines. For example, they write that high-empathy subjects believe the mood-enhancing experience offered to them “won’t do a very good job of removing the bad feelings they have when they empathize with a needy other” (2000b, p. 267). There are two sorts of means-end belief they can be attributing to people here.<sup>14</sup>

*Option 1a:* On this proposal, high-empathy subjects ultimately desire to be restored to a normal state and believe that helping is a better means to this than the *particular* mood-enhancing experience offered by the experimenters. On this proposal, the subjects either believe that what they are being offered won’t work as well or that it will but it’s a more costly option. Sober and Wilson seem to suggest the former when they ask rhetorically: “why expect the subject to think that listening to music will be a completely satisfactory mood corrective?” (1998, p. 271). Here they are referring to the particular mood-enhancing experience Cialdini et al. (1987) offered to participants in one of their experiments. I agree that it’s questionable whether subjects would take this to be a good way of boosting their mood (or at least whether it would be a better means than helping). However, Sober and Wilson fail to mention that Batson provided an alternative which it is quite plausible to think participants took to be mood-enhancing.

Batson et al. (1989) told subjects they would watch a short video. Some were told it is in a category which previous research has shown to “cause strong feelings of happiness and pleasure” while others were told the video is in a category that would cause “moderate feelings of depression and sadness” (p. 924). Furthermore, participants were asked to “give an example of a movie or TV show they had seen that might fit into the category to which they had been assigned” after which the experimenter “assured them that the material they identified ‘was just the sort of thing in [that category]’” (p. 924). Various checks were done as well which verified the effectiveness of this manipulation of anticipated mood-enhancement. So there is no reason to think participants thought either that this method would be ineffective or more costly than helping. Yet the results were still problematic for the NSRH: empathically aroused subjects who were led to believe they could have their mood

---

<sup>14</sup> The relevant belief need only be attributed to empathically aroused people in these situations, not all people. But such a proposal requires holding that empathy induces such a belief. A more plausible proposal, however, is that most people have this belief (when empathically aroused or not), but the belief only becomes relevant to people’s practical reasoning when their empathic arousal induces the relevant ultimate desire.

boosted in this way (rather than helping) still tended to help. Besides, the more general problem with this option is that it's viciously *ad hoc* in the absence of a principled reason to think the particular mood-enhancing experience offered would be deemed by subjects to be insufficient or otherwise inferior.

*Option 1b:* Perhaps, then, Sober and Wilson are proposing that empathically aroused subjects believe that the *only* way to satisfy their ultimate desire is to help (i.e. no other possible mood-enhancing experience will be sufficient). The idea is that these subjects *falsely* believe that it is a satisfaction condition of their ultimate desire that the other is helped. While they only ultimately want their mood to be boosted, empathically aroused individuals for some reason think this desire can only be satisfied by the person getting help. But attributing such a belief is unwarranted when other, more plausible explanations are available. To better understand this, consider an analogy. Suppose we run an experiment in which participants are put in a room with a television and told they can watch whatever they want for an hour. Presumably most people in such a situation would have a desire to watch something entertaining. But suppose we told a joke to half of the subjects beforehand and tracked (for all subjects) what type of show they chose to watch. What should we say about their mental states if we find that telling the joke has a significant effect on watching a comedy? Should we conclude that the subjects in the joke condition maintain the more general desire to watch something entertaining but for some odd reason falsely believe that the *only sufficient means* is to watch a comedy? Or is it more plausible that telling participants a joke induces in them the more specific desire to watch a comedy? Surely the latter is the more reasonable conclusion of the two.

Option 1b is like the first, implausible explanation in our imaginary experiment. It holds that high-empathy subjects merely ultimately desire to have their mood boosted (because they're sad) and think that the *only* way to boost their mood is by helping. But if these subjects just ultimately wanted their mood to be restored, why wouldn't they simply take the less costly option of watching a video that they have every reason to believe causes strong feelings of happiness and pleasure? After all, one's mood can undoubtedly be boosted in a multitude of ways. Most participants would surely recognize this. Attributing to them a belief to the contrary is unwarranted when a better explanation is available (e.g. that they ultimately desire that the other is helped).<sup>15</sup> At the very least, more would need to be said (or found in further experiments) to justify Option 1b. Of course, this option would be much more plausible if the idea is that subjects *correctly* think the needy other must be helped in order for their ultimate desire to be satisfied. But this isn't Option 1b. In modifying the ultimate desire, it's a version of the next option.

---

<sup>15</sup> Sober and Wilson's alternative proposal cannot be a mere possibility in the sense of merely compatible with the data however implausible. Their egoistic hypothesis must be at least *as plausible* as the altruistic alternative if it is to be taken as a live possibility that diffuses Batson's argument. Showing that an alternative is not ruled out by the data involves showing more than its mere compatibility with the data. Otherwise, Batson would be wrong simply because an egoistic evil demon hypothesis is commensurate with the data.



## 5.2 The Second Option: Modify the Desire

So, how can we interpret Sober and Wilson's modified NSRH such that it predicts that high-empathy subjects' ultimate desires cannot be satisfied by any old mood-enhancing experience? The most plausible option is to modify the ultimate desire of the empathically aroused. The proposal must be that the ultimate desire is more specific than the original NSRH claimed such that the contents of the ultimate desires of these individuals connect them specifically with the other's plight. In fact, there is evidence that Sober and Wilson have thought of their proposal along such lines. Alejandro Rosas (2002) reports, according to personal communication with Sober, that "what Sober and Wilson are postulating is an egoistic desire for relief from the specific sadness *that* someone else is suffering, where this relief requires, *necessarily*, an improvement in the situation of the person who is the object of empathy" (p. 105). On this proposal, the ultimate desires of empathically aroused subjects have something like the following, more specific, content: *I am relieved from being sad about the suffering of this person by this person being helped*. This, unlike the ultimate desire postulated by the original NSRH, makes sense of why the satisfaction of the relevant ultimate desires requires the specific mood-enhancing experience of the other person being helped. Now it is explicitly a satisfaction condition of the ultimate desire that the other is helped. Their ultimate desire is not simply to be restored to a normal state; it is to be relieved from being sad about this person's suffering *by the other being helped*. After all, we sometimes not only want certain things, but want them to be a certain way. And this seems especially true of empathically-motivated helping: when we empathize with someone, it seems we're not just sad in general, we're sad in particular about this person's suffering.<sup>16</sup>

But this interpretation of the modified NSRH also has its problems. As Rosas points out, "under Sober and Wilson's 'egoistic' hypothesis, the specific propositional object of the sadness in question is the same state of affairs that would be the object of an ultimate altruistic concern, namely the welfare, or lack of it, of some other person" (2002, p. 105). The implication seems to be that such a hypothesis is not really egoistic, given that the ultimate desire will only be satisfied if the needy other is helped. Elaborating on Rosas's complaint, one might argue that the ultimate desire on Sober and Wilson's hypothesis is *other-directed*, since the needy other is part of the content of the ultimate desire. So construed, then, the modified NSRH is a version of psychological *altruism*.

Such an objection to Sober and Wilson, however, goes a bit too fast if we are only equipped with Sober and Wilson's conception of the debate. On our interpretation of their modified NSRH, the ultimate desire attributed to empathically aroused individuals is relational. But Sober and Wilson's conception of the debate doesn't say anything about relational desires and whether they are compatible with psychological egoism or altruism.

---

<sup>16</sup> Note that such a proposal isn't committed to the claim that the desire is satisfied only if the subject is the one to help, which may conflict with data gathered on other versions of the ESRH. One can be relieved from being sad about the suffering of someone without being the one to help. Being the one to help is simply the preferred means to achieve their end in such situations because they think they are the only one available to help.

But now, with the inclusion of relational desires into both psychological egoism and altruism, we can properly evaluate the modified NSRH of Sober and Wilson. We can now see that the charge that it fails to be an egoistic hypothesis poses a substantial threat. First, recall that to account for all the current data the proposal must construe a subject's empathically aroused ultimate desire as relational, connecting the subject to the needy other and her plight. On this supposedly egoistic explanation, the ultimate desire of high-empathy individuals is to be relieved from feeling sad about the suffering of this person by this person being helped. This explains the data by predicting that not just any old mood-enhancing experience can satisfy the ultimate desires of the empathically aroused. However, the question then arises: Can the psychological egoist legitimately help herself to such an explanation of the data? We are now in a position to answer this question in the negative. Sober and Wilson's hypothesis is construed in such a way that the needy other is an essential beneficiary of the desire: it is a necessary condition for the satisfaction of the desire that the other perceived to be in need is helped. But then the ultimate desire is no longer egoistic. Indeed, it may even be altruistic (recall Tennis Game). And this is the case whether the person represents herself as an essential beneficiary or not. Insofar as someone other than the agent herself is an essential beneficiary in the content of the ultimate desire, it ceases to be the kind of desire allowed by psychological egoism.

## 6. Conclusion

Two things, I hope, are now clear. *First*, relational desires must be incorporated into our conception of psychological egoism and altruism, and this can be done by making use of the concept of essential beneficiaries. *Second*, Sober and Wilson's modified NSRH is not available for the psychological egoist. Indeed, our improved conception of the debate shows that any explanation of that type attributes a non-egoistic desire to empathically aroused individuals. While ultimate relational desires can be egoistic (as in Stolen Kiss), relational desires whose contents relate the individual to another's *benefit* cannot. So, while the hypothesis is an explanation of the data, it's not an *egoistic* one.

These two theses I've tried to establish have two immediate consequences that I'd like to mention. First, this could spell some bad news for Batson. Batson can take comfort in the fact that Sober and Wilson's modified NSRH cannot be used to support their claim that the empirical evidence against psychological egoism is inconclusive. However, if I am right that Batson intends the EAH to be positing ultimate desires that are *other-directed*, then there is another non-egoistic explanation on the table now in competition with Batson's EAH. This competing hypothesis is that empathy induces ultimate desires that are *relational*, rather than other-directed. It could of course be that Batson conceives of the EAH in a way that is neutral on whether the ultimate desire is relational or not. If so, then the truth of my theses would pose no problem for Batson's hypothesis, so long as such relational desires are in fact altruistic. I think they are, but I need not argue as much here. In

any event, the data gathered so far do not provide enough evidence for us to say whether the ultimate desires of empathically aroused individuals are relational or other-directed.

The second consequence is that this does go some way toward defending the sort of non-deductive argument against psychological egoism that Batson has attempted to provide. While the data may not favor the standard EAH over a modified NSRH (or something similar), one might still argue that psychological egoism is false because the data rule out all the plausible hypotheses that are properly *egoistic*. If this sort of Batsonian argument is sound, then it's still reasonable for us to believe that psychological egoism is false based on the empirical evidence.<sup>17</sup>

## References

- Batson, C. D. (1987). Prosocial Motivation: Is It Ever Truly Altruistic? In L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, 20: 65–122.
- Batson, C. D. (1991). *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Batson, C. D. (2000). Unto Others: A Service... And a Disservice. *Journal of Consciousness Studies*, 7 (1-2): 207–10.
- Batson, C.D., Batson, J.G., Griffitt, C.A., Barrientos, S., Brandt, J.R., P. Sprengelmeyer, & Bayly, M.J. (1989). Negative-State Relief and the Empathy-Altruism Hypothesis. *Journal of Personality and Social Psychology* 56 (6): 922-933.
- Batson, C. D., Dyck, J., Brandt, J., Batson, J., Powell, A., McMaster, M., & Griffitt, C. (1988). Five Studies Testing Two New Egoistic Alternatives to the Empathy-Altruism Hypothesis. *Journal of Personality and Social Psychology*, 61: 52–77.
- Batson, C. D & Shaw, L. L. (1991). Evidence for Altruism: Toward a Pluralism of Prosocial Motives. *Psychological Inquiry*, 2 (2): 107–122.
- Cialdini, R. B., Schaller, M., Houlihan, D., Arps, K., Fultz, J., & Beaman, A. (1987). Empathy-based Helping: Is it Selflessly or Selfishly Motivated? *Journal of Personality and Social Psychology*, 52 (4): 749–758.
- Cialdini, R. B. (1991). Altruism or Egoism? That Is (Still) the Question. *Psychological Inquiry*, 2: 124-126.
- Nagel, T. (1970). *The Possibility of Altruism*. Princeton University Press.
- Rosas, A. (2002). Psychological and Evolutionary Evidence for Altruism. *Biology and Philosophy*, 17: 93–107.
- Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.

---

<sup>17</sup> I would like to thank the following individuals for helpful comments on or discussions about drafts of this paper: Aaron Zimmerman, Michael Rescorla, Matthew Hanser, Alex Bundy, Ian Nance, Jonathan Way, Luke Manning, Stephanie May, Stephen Stich, and an anonymous referee.

- Sober, E. & Wilson, D. S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
- Sober, E. & Wilson, D. S. (2000a). Summary of: 'Unto Others. The Evolution and Psychology of Unselfish Behavior'. *Journal of Consciousness Studies*, 7 (1-2): 185–206.
- Sober, E. & Wilson, D. S. (2000b). Morality and 'Unto Others'. Replies to Commentary Discussion. *Journal of Consciousness Studies*, 7 (1-2): 257–268.
- Stich, S., Doris, J.M., & Roedder, E. (forthcoming). Altruism. *The Oxford Handbook of Moral Psychology*, ed. by the Moral Psychology Research Group, Oxford University Press.