



Why there might not be an evolutionary explanation for psychological altruism



Stephen Stich

Department of Philosophy & Center for Cognitive Science, Rutgers University, United States

ARTICLE INFO

Article history:
Available online 21 November 2015

Keywords:
Altruism
Psychological altruism
Daniel Batson
Norm acquisition
Cultural evolution

ABSTRACT

The existence of psychological altruism is hotly debated in the psychological and philosophical literature. In this paper I argue that even if psychological altruism does exist in some (or all) human groups, there may be no purely evolutionary explanation for existence of psychological altruism.

© 2015 Elsevier Ltd. All rights reserved.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Biological and Biomedical Sciences*

1. Introduction

From Hobbes to the present, philosophers have debated whether humans are capable of psychological altruism, and as psychology, economics and other social sciences developed, scholars working in those disciplines have joined the fray. For much of the four centuries since Hobbes, the dominant view among both philosophers and social scientists was that people are *not* capable of psychological altruism. But in recent decades, the prevailing opinion seems to have shifted. While there are still voices to be heard on both sides, many philosophers and social scientists now think that humans can and do sometimes act altruistically.¹ This has led to an increased interest in the evolutionary explanation of psychological altruism.² If humans have the capacity for psychological altruism, it seems important to ask how this capacity evolved? While I applaud the quest for an “ultimate” explanation of psychological altruism, I think its focus may be too narrow. For it may be the case that while psychological altruism exists, it does not have an evolutionary explanation, at least as such explanations are traditionally conceived. My goal, in this paper, is to explain why it

might be the case that psychological altruism has no purely evolutionary explanation, and to draw attention to the sort of empirical work that will help us decide the issue. But before getting to any of this, a fair amount of scaffolding will be needed. The first job is to say what I *mean* by “psychological altruism.”

2. Psychological altruism and evolutionary altruism

As Sober and Wilson rightly stressed in *Unto Others* (1998), contemporary discussions of evolution and altruism are all too often impeded by unclear or ambiguous terminology. One crucial distinction to note at the outset is between two very different conceptions of altruism, which, following Sober and Wilson, I'll label *psychological altruism* and *evolutionary altruism*. Here is how I will be using these terms:

A behavior (or a behavioral disposition) is *evolutionarily altruistic* if and only if it decreases the inclusive fitness of the organism exhibiting the behavior and increases the inclusive fitness of some other organism. To a rough first approximation, inclusive fitness is a measure of how many copies of an organism's genes exist in future generations.³ “Behavior,” in this definition, is to be interpreted very

E-mail address: stich.steve@gmail.com.

¹ For valuable discussions of the history of the debate, see Batson (1991), ch. 1–3; MacIntyre (1967); Sober & Wilson (1998), ch. 9.

² See, for example, Kitcher (2006), Schulz (2011), Sober & Wilson (1998) ch. 10, Stich (2007).

³ Some authors, including Sober and Wilson, prefer an account of evolutionary altruism that invokes individual fitness rather than inclusive fitness, where individual fitness is, roughly, a measure of how many descendants an organism has.

broadly—so broadly that amoeba and even plants can behave. A behavior or an action is *psychologically altruistic* if and only if it is motivated by an ultimate desire for the well-being of others. This account needs further elaboration since it invokes some philosophically freighted jargon. But even without further clarification, it should be clear that psychological altruism and evolutionary altruism are quite distinct phenomena. It is, after all (logically) possible for an organism like a paramecium or a petunia to exhibit evolutionary altruism even though it has no mind at all, and thus can't have any ultimate desires. It is also (logically) possible for an organism's behavior to be psychologically altruistic but not evolutionarily altruistic. If, for example, some of a mother's actions are motivated by an ultimate desire for the well-being of her child, those actions are psychologically altruistic, but they are not evolutionarily altruistic, since they are likely to increase the mother's inclusive fitness.

Psychological altruism is the sort of altruism that moral philosophers have traditionally been concerned with. And since it will be center-stage in this paper, we'd do well to elaborate on the definition offered above. A first question is how we are to understand the notion of a desire *for the well-being of others*. This is a hotly debated topic that raises some deep and challenging issues about the nature of well-being.⁴ But for our purposes, I think a few examples will suffice. Desires to save someone else's life, to alleviate someone else's pain, to cure someone else's illness, or to make someone else happy all count as desires for the well-being of others. The claim that people sometimes have desires of this sort is not controversial. What is controversial is whether these desires are ever *ultimate* desires. If they are, then Hobbes and others, who deny the existence of psychological altruism, are mistaken. Since the notion of an *ultimate desire* looms large in the debate, some clarification of that notion is in order.

The intuitive idea is that a desire is *ultimate* if the object of the desire is desired for its own sake, rather than because the agent thinks that satisfying the desire will lead to the satisfaction of some other desire. This can be made more precise by appeal to one interpretation of the traditional notion of practical reasoning. On this interpretation, practical reasoning is a causal process via which a desire and a belief give rise to or sustain a new desire. That new desire, along with a second belief can give rise to yet another desire, and so on. If this causal chain produces a desire for a "basic" action—the kind of desire that causes bodily movement without the mediation of further desires, the result is an action.⁵ Fig. 1 is a depiction of the process that will prove useful in what follows.⁶

To make this account of practical reasoning a bit more intuitive, let's consider an example. Suppose I want an espresso. Let that be DESIRE 2 in Fig. 1. Suppose further that I believe that the best place to get an espresso is the Starbucks on George St.—that will be BELIEF 2. Together, these cause the formation of DESIRE 3, the desire to go to the Starbucks on George St. If I believe that the best way to get to the Starbucks on George St. is to take the elevator to the street level (BELIEF 3), I will form the desire to take the elevator to the street level (DESIRE 4). Though that is not a desire for a basic action, it should be clear how the process can continue until we reach such a desire—the desire to touch the Ground Floor button in

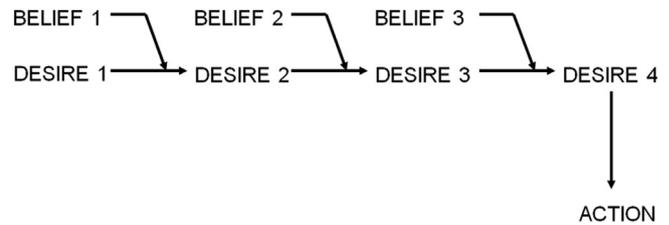


Fig. 1. The process of practical reasoning.

the elevator, perhaps—and I'm on my way to Starbucks to get my espresso.

Desires that are formed or sustained by practical reasoning are *instrumental desires*. But obviously, if we are to avoid circularity or infinite regress, there must be some desires that are not instrumental. Those desires, which are *not* the product of practical reasoning are *ultimate desires*. So, in the very short practical reasoning chain in Fig. 1, DESIRES 2, 3 and 4 are instrumental; DESIRE 1 is ultimate. As noted earlier, those on both sides of the debate over psychological altruism agree that people often have desires for the well-being of others. But those who deny the existence of psychological altruism insist that *all* such desires are instrumental. And while those who defend psychological altruism grant that *some* desires for the well-being of others are instrumental, they maintain that there are also some *ultimate* desires for the well-being of others. Thus both those who deny the existence of psychological altruism and those who defend it would find the pattern of practical reasoning displayed in Fig. 2 to be unproblematic. But the two sides disagree about the pattern of practical reasoning displayed in Fig. 3. Those who defend the existence of psychological altruism insist that episodes of practical reasoning like this occur, while those who deny the existence of psychological altruism maintain that they do not.

3. Should we seek an evolutionary explanation of psychological altruism?

The critics of psychological altruism maintain that people do not have ultimate desires for the well-being of others, and thus that patterns of practical reasoning like the one illustrated in Fig. 3 never actually occur. If the critics are right, then obviously constructing and evaluating evolutionary explanations of psychological altruism would be a fool's errand, on a par with seeking an evolutionary explanation of mental telepathy or X-ray vision. Over the last three decades, however, psychologists have assembled an increasingly impressive body of evidence suggesting that the critics

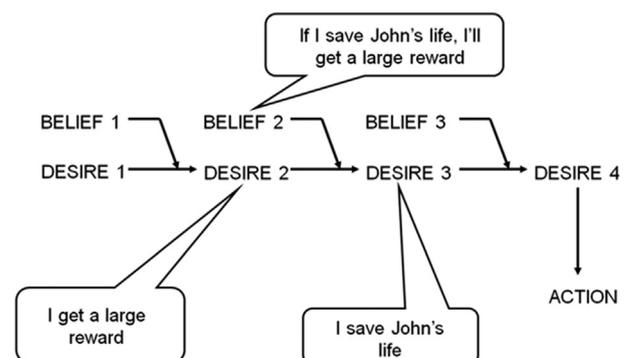


Fig. 2. A pattern of practical reasoning that both critics and defenders of psychological altruism find unproblematic.

⁴ Kahneman, Diener & Schwarz (1999) is a wide-ranging collection of papers on the issue. For an insightful discussion of the literature, see Haybron (2008).

⁵ For a classic elaboration of this picture of practical reasoning, see Goldman (1970).

⁶ In Fig. 1, and throughout this paper, I adopt the simplifying assumption that practical reasoning is a process in which *one* belief and *one* desire lead to the formation of a new desire. Though this assumption makes exposition easier, it is clearly unrealistic. Often several beliefs and/or several desires will jointly cause the formation of the new desire.

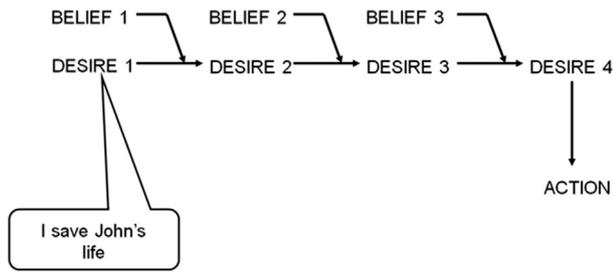


Fig. 3. Defenders of psychological altruism claim that episodes of practical reasoning like this occur, while those who deny the existence of psychological altruism maintain that they do not.

are mistaken. Many investigators have contributed to this work, though in my opinion (which I think is widely shared) Daniel Batson has set out the most careful and systematic case for the existence of psychological altruism (Batson, 1991, 1998, 2011). In dozens of experiments, Batson and his collaborators have set up situations in which people exhibit helping behavior, and then systematically attempted to rule out non-altruistic explanations of that behavior. Though Batson thinks that, in some cases at least, he has ruled out all the non-altruistic alternatives, I am not entirely convinced. The reasons for my skepticism make for a very long story indeed, which John Doris, Erica Roedder and I have set out elsewhere (Stich, Doris, & Roedder, 2010). In this article, however, I propose to put those concerns aside and assume, for argument's sake, that Batson and his colleagues have indeed shown that psychological altruism exists in many of their experimental participants. But even if we make this assumption, I am not convinced that we should be seeking an evolutionary explanation for psychological altruism. My reason can be stated very briefly: *All of Batson's participants were WEIRD*. Obviously, this needs some unpacking.

In an important and widely discussed article Joseph Henrich, Steven Heine and Ara Norenzayan (2010) used the term WEIRD as an acronym for cultures that are Western, Educated, Industrialized, Rich and Democratic. They went on to document two claims, one of which was much more surprising than the other. The unsurprising claim is that the vast majority of experimental participants in psychology are people from WEIRD cultures. The surprising fact is that in those cases where reliable cross-cultural data are available, it is often (though not always) the case the WEIRD people are outliers. Moreover, a large majority of WEIRD experimental participants are American, and Americans are often outliers among the outliers. College students are vastly over-represented in this sample, and, yes, they are often outliers among Americans. The conclusion that Henrich and colleagues draw from all this is that while we now know a great deal about the psychology of WEIRD people (particularly American college students), we may know much less than we think about the psychology of *Homo sapiens*. Batson's experimental participants fit the pattern that Henrich et al. describe. All of them are WEIRD, all of them are American, and just about all of them are college students. None of this detracts from the importance of Batson's work. If, as he thinks, and as I am conceding, these participants exhibit psychologically altruistic behavior, then a centuries old philosophical dispute has been resolved. *Some people, sometimes* exhibit psychological altruism, and that is exactly what Hobbes and his many fellow skeptics denied. But since Batson's participants are WEIRD and WEIRD people are often outliers, his work leaves open whether psychological altruism is a feature of the psychology of *H. sapiens*. Perhaps psychological altruism is like skill at throwing an American football, or like a taste for highly spiced food, or like a belief that Jesus Christ is one's personal savior. Or, to mention possibilities that will loom large shortly, perhaps it is like the conviction that one is

morally obligated to circumcise one's sons (or one's daughters!), or like a desire to obey the 613 commandments in the Torah. Each of these phenomena is widespread in some cultures and rare or non-existent in others. If psychological altruism, too, is common in WEIRD cultures but rare or non-existent in others, then while evolution surely has *some* role to play in explaining the phenomenon, just as it does for any psychological phenomena, evolutionary theorizing will only tell part of the story. The local history, local culture and local environment of groups in which psychological altruism is found will also play an important role. So there will be no *purely* evolutionary explanation of psychological altruism—no explanation that does not appeal to culturally local facts about groups of people who exhibit psychological altruism.

4. How could psychological altruism be a culturally local phenomenon?

At this point, I think it would be reasonable for the reader to raise the following challenge: Yes, it is true that even if Batson's work definitively establishes that WEIRD humans sometimes exhibit psychological altruism, his work does not show that humans in other cultures behave altruistically; and if they don't then perhaps there is no purely evolutionary explanation of psychological altruism. But how could it come about that psychological altruism is present in some cultures and absent in others? That's a fair question, and without a plausible answer I think the reader would be entirely justified in viewing the idea that psychological altruism is culturally local as a possibility that needn't be taken seriously. So in this section, I'll offer a sketch of how it might come about that psychological altruism is culturally local.

The first step is to note that while (by definition) ultimate desires can't be caused by practical reasoning, they can be caused by a variety of other processes. Almost certainly, some ultimate desires, like the desire to avoid pain, perhaps, are innate and pan-cultural. However it is also very likely that many ultimate desires are acquired. Moreover, it is likely that some desires that start out as instrumental desires *become* ultimate desires as a person matures. Here is a hypothetical example:

As a teenager, Tom very much wanted to make the U.S. Olympic track team. He believed that if he went for a long and strenuous run every day, he would make the team. So, via practical reasoning, he formed the desire to go for a long, strenuous run every day. When he first formed the desire, it was a clear case of an instrumental desire. But as time went by, Tom realized that he would never be fast enough to make the Olympic team, even if he went for a long, strenuous run every day. Nonetheless, Tom continued to want to go for a long, strenuous run every day. He still does. Moreover, he does not believe that running every day is a good means to anything else he desires. The desire to go for a long run every day has become one of Tom's ultimate desires. It is something he desires for its own sake.

A psychological hedonist—someone who maintains that the only ultimate desires are the desire to experience pleasure and the desire to avoid pain—might deny that this is psychologically possible and insist that Tom must believe that running will lead to the satisfaction of one or both of those ultimate desires. But since we have *assumed* that psychological altruism exists, this is a challenge we can safely ignore, since if psychological altruism exists, psychological hedonism is false. Hypothetical examples like this are hardly conclusive. But unless one thinks that our cognitive system keeps careful track of the beliefs that originally led to the formation of each instrumental desire, and always deletes a desire when the belief changes, there are likely to be lots of cases in which

instrumental desires are retained despite the loss of the belief that helped cause them. When this happens, the instrumental desire has become ultimate.⁷ Here is another example that is more directly relevant to our current concerns:

As a child, Emma wanted to obey the moral norms that prevailed in her community because she believed that God would punish her if she did not. But later in life, Emma retains the desire to obey some of these norms, even though she no longer believes that God, or anyone else, will punish her for violating them. Obeying these norms has become something Emma wants to do for its own sake. The desires to obey specific norms, which started out as instrumental desires, have become ultimate desires.

If it is agreed that cases like Emma's are common, the next question to ask is why we have psychological mechanisms that lead us to acquire ultimate desires to perform the actions that are required by local norms. There are a number of answers to be found in the literature. One of the most interesting and elegant of these is proposed by Chandra Sripada (2007). The central idea in Sripada's account is that norms are an important tool for coordinating behavior. Thus, while different sets of norms may be equally adaptive, it is important for people to behave in accordance with the norms that prevail in *their* community. Under these circumstances, Sripada argues, natural selection may well have led to the evolution of a psychological mechanism that generates ultimate desires to adhere to locally prevailing norms. In Emma's case, we might imagine that the desires to adhere to various local norms start out as instrumental and become ultimate when the mechanism Sripada posits kicks in. But it might well be the case that the intermediate instrumental desire stage is often unnecessary. The local norm acquisition mechanism might simply identify locally prevailing norms and generate ultimate desires to behave in the ways that these norms require.

What does all this have to do with psychological altruism? Well, if evolution has indeed endowed us with a local norm acquisition system of the sort that Sripada proposes, and if there are, in some communities, norms that require specific sorts of helping behavior in specific circumstances, then people in those communities will acquire ultimate desires to help others in the circumstances specified by the local norms. And when they act on those desires, their behavior will be psychologically altruistic. But people in other communities may well have quite different norms that require helping behavior in very different circumstances. And in still other communities, there may be no norms at all requiring helping behavior. Moreover, even if it turns out that all communities have norms requiring helping behavior of one sort or another, or that there are a few specific norms requiring helping behavior that are present in every culture, biological evolution may provide only part of the explanation for these facts. If Sripada's theory is correct, it is biological evolution that produced a psychology designed to acquire local norms. But local history, local ecology, and cultural evolution—including cultural group selection—will be needed to

explain which helping norms exist in a given community and why all cultures have norms requiring helping behavior, if indeed they do.

What I have been sketching is an account of how some psychologically altruistic behavior might have no purely evolutionary explanation. But there is no reason to assume that all psychologically altruistic behavior is caused in the same way. Perhaps some ultimate desires for the well-being of others are acquired by a Sripada-style norm acquisition device, while others, like Sober and Wilson's example of an ultimate desire for the well-being of one's own children, are "hardwired" and can be straightforwardly explained by appeal to biological selection. If that's the case, then some sorts of psychologically altruistic behavior will have a purely evolutionary explanation and others will not.

5. Summary and conclusions

My goal in this paper was *not* to argue that there is no purely evolutionary explanation of psychological altruism. Rather, my goal was to argue that, even if we grant that psychological altruism exists, there *might* not be any explanation for its existence that appeals only to biological evolution. For it might be the case that biological evolution endowed humans with a norm psychology which identifies important norms that prevail in their culture, and leads them to form ultimate desires to act in a way that those norms require. If that is the case, then cultural evolution, local ecology and local history would be an important part of the explanation of why some people have ultimate desires for the well-being of others. However, it might also be the case that some or all psychologically altruistic behavior is motivated by innate ultimate desires for the well-being of others, that the existence of those innate desires can be explained entirely by appeal to biological evolution, and that cultural evolution, local history and local ecology play no significant role. In order to decide between these options and to move the inquiry forward, we need to know a lot more about the helping norms that prevail in cultures, large and small, that are not WEIRD, and about the behavior that people in these cultures would exhibit in culturally appropriate versions of Batson's now classic experiments.

References

- Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Batson, C. D. (1998). Altruism and helping behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.) (4th ed.) *The handbook of social psychology* (Vol. 2, pp. 282–316). Boston: McGraw-Hill.
- Batson, C. D. (2011). *Altruism in humans*. Oxford: Oxford University Press.
- Goldman, A. (1970). *A theory of human action*. Englewood-Cliffs, NJ: Prentice-Hall.
- Haybron, D. (2008). *The pursuit of unhappiness: The elusive psychology of well-being*. Oxford: Oxford University Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavior and Brain Sciences*, 33(2/3), 1–23.
- Kahneman, D., Diener, E., & Schwarz, N. (Eds.). (1999). *Well-being: The foundations of hedonic psychology*. New York: Russell Sage Foundation.
- Kitcher, P. (2006). Between fragile altruism and morality: Evolution and the emergence of normative guidance. In G. Boniolo, & G. De Anna (Eds.), *Evolutionary ethics and contemporary biology* (pp. 159–177).
- MacIntyre, A. (1967). Egoism and altruism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 2, pp. 462–466). New York: Macmillan.
- Schulz, A. (2011). Sober & Wilson's evolutionary arguments for psychological altruism: A reassessment. *Biology and Philosophy*, 26, 251–260.
- Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.
- Sripada, C. (2007). Adaptationism, culture, and the malleability of human nature. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *Foundations and the future: Vol. 3. The innate mind* (pp. 311–329). Oxford: Oxford University Press.
- Stich, S. (2007). Evolution, altruism and cognitive architecture: A critique of Sober and Wilson's argument for psychological altruism. *Biology and Philosophy*, 22, 267–281.
- Stich, S., Doris, J., & Roedder, E. (2010). Altruism. In J. Doris, & The Moral Psychology Research Group (Eds.), *The moral psychology handbook* (pp. 147–205). Oxford: Oxford University Press.

⁷ There are some delicate issues lurking here. A critic might insist that if a desire was initially caused by practical reasoning, then it is *always* an instrumental desire, even if the beliefs that helped to cause the desire are no longer present. Thus a desire that began as instrumental can never become ultimate. I am inclined to think that this purely historical account of instrumental desires, and the very restrictive account of ultimate desires that it entails, does not do justice to the traditional debate over psychological altruism, since both friends and foes of psychological altruism would agree that if a person has a desire for the well-being of others, and she does not now believe that satisfying that desire will lead to the satisfaction of any other desire, then that desire is ultimate. But even if I am wrong about this, Sripada's theory, discussed in the following paragraphs, explains how lots of new ultimate desires can be acquired.