

# 1

## Did evolution make us psychological egoists?

### 1. TWO CONCEPTS

The concept of altruism has led a double life. In ordinary discourse, as well as in psychology and the social sciences, *altruism* refers to behaviors that are produced because people have certain sorts of motives. In evolutionary biology, on the other hand, the concept is applied to behaviors that enhance the fitness of others at expense to self.

A behavior can be altruistic in the evolutionary sense without being an example of psychological altruism. A plant that leeches insecticide into the soil may be an altruist, if the insecticide benefits its neighbors and imposes an energetic cost on the producer. In saying this, I am not attributing a mind to the plant. Evolutionary altruism has to do with the fitness consequences of the behavior, not with the mechanisms inside the plant (mental or otherwise) that cause the plant to behave as it does.

Symmetrically, a behavior can be altruistic in the psychological sense without being an example of evolutionary altruism. If I give you a volume of Beethoven piano sonatas (or a package of contraceptives) out of the goodness of my heart, my behavior may be psychologically altruistic. However, the gift giving will not be an example of evolutionary altruism, if the present fails to augment your prospects for survival and reproductive success.

Although the concepts are different, they have a few things in common. Both point to causal explanations of the behaviors so labeled. If I say that a behavior is an example of psychological altruism, I am making a claim about the motives that produced the behavior. If I say

I am grateful to Daniel Batson, Don Moskowit, and the members of the philosophy departments at University of Pittsburgh, Rutgers University (New Brunswick), and at University of Wisconsin (Madison) for useful discussion.

## *Did evolution make us psychological egoists?*

that a behavior is an example of evolutionary altruism, I am suggesting a certain sort of explanation, which I'll describe in the next section, for why the behavior evolved.<sup>1</sup>

The second common feature of the psychological and the evolutionary concepts is that both have been controversial and even unpopular in much scientific investigation. Psychological egoism, which claims that all of our ultimate motives are selfish, has viewed psychological altruism as a comforting illusion. Egoism has been the dominant position in all major schools of twentieth-century psychology (Batson 1991). And within evolutionary biology, the theory of the selfish gene has been hostile to the idea that evolution produces behaviors that help the group at the expense of the individual (Williams 1966; Dawkins 1976).

My goal in this paper is to clarify these concepts and to further discuss why they are logically independent of each other. Then, having separated them, I will attempt to bring them back into contact. I will explore the question of whether there are evolutionary considerations that help us decide whether we are ever psychological altruists.

## 2. EVOLUTIONARY ALTRUISM

Altruism has been an important subject for evolutionary theorizing ever since Darwin. I will not describe the history of how this subject has developed, nor will I discuss intricacies that are internal to various theories of current interest.<sup>2</sup> My modest goal in this section is to describe with more care what altruism and selfishness mean in an evolutionary context and to show how each trait is connected with its own picture of how natural selection has operated.

For the most part, Darwin viewed natural selection as a process in which organisms within the same breeding population compete with each other to survive and reproduce. His picture of competition was not the lion versus the lamb, but lions competing with lions and lambs with lambs. In this process, the traits that evolve are the ones that benefit the individual organism. Although Darwinism is sometimes described by saying that characteristics evolve "for the good of the species," this is a major distortion of how Darwin *usually* thought about natural selection.

*Usually*, but not *always*. There were a small number of occasions on which Darwin took seriously the idea that natural selection involves competition among objects other than individual organisms. One of

### *From a biological point of view*

the clearest expressions of this alternative occurs in his discussion of human morality in *The Descent of Man*. Here is Darwin's statement of the problem:

It is extremely doubtful whether the offspring of the more sympathetic and benevolent parents, or of those which were the most faithful to their comrades, would be reared in greater number than the children of selfish and treacherous parents of the same tribe. He who was ready to sacrifice his life, as many a savage has been, rather than betray his comrades, would often leave no offspring to inherit his noble nature. The bravest men, who were always willing to come to the front in war, and who freely risked their lives for others would on average perish in larger numbers than other men. (Darwin 1871, p. 163)

Darwin's point is that if we consider a single tribe that contains both altruistic and selfish individuals, altruists will do *worse* than selfish individuals. If natural selection is the main cause of evolutionary change (as Darwin thought), and if natural selection causes fitter traits to increase in frequency and less fit traits to decline, why hasn't altruism altogether disappeared from human conduct? Here is the answer that Darwin suggests:

It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an advancement in the standard of morality and an increase in the number of well-endowed men will certainly give an immense advantage to one tribe over another. (Darwin 1871, p. 166).

Although altruistic *individuals* do worse than selfish *individuals* in the same tribe, altruistic *groups* do better than selfish *groups*. Here Darwin was imagining a process of group selection, in which groups compete against each other. This picture of the process of natural selection differs markedly from his more customary formulation, in which organisms within a single population engage in a struggle for existence.

To make sense of the idea of evolutionary altruism, and of the process of group selection that is associated with it, one must be able to think simultaneously about the fitnesses of organisms and the fitnesses of groups of organisms. How are these two levels related? And since altruism is a behavior produced by an individual organism, how do these two kinds of fitness make it possible for altruism to evolve when there is group selection?

Figure 1.1 depicts some of the main conceptual ingredients. It shows that the fitness of an individual depends on two factors. Whether the

## *Did evolution make us psychological egoists?*

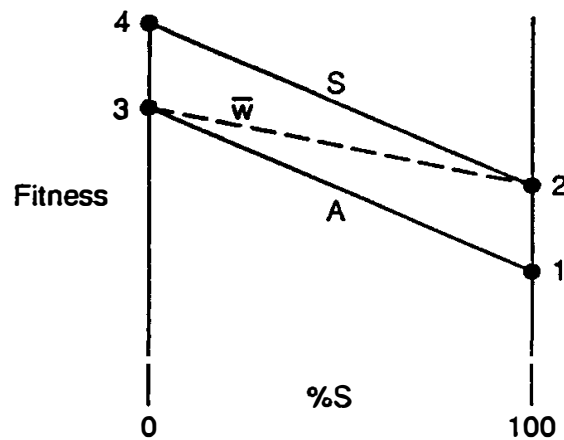


Figure 1.1

organism is altruistic (A) or selfish (S) influences its fitness; in addition, an individual's fitness also is influenced by the frequency of altruism in the group in which the individual lives.

Besides describing the fitnesses of the two traits, the figure also represents a third quantity, labeled  $\bar{w}$  ("w-bar"). This is the average fitness of the organisms in the group. It is a plausible measure of what is meant by *the fitness of the group*. Interpreted in this way, the figure represents the idea that groups of altruists are fitter than groups of selfish individuals.

This figure expresses two ideas that are crucial to the definition of evolutionary altruism: (i) within each group, selfish individuals are fitter than altruists; (ii) groups of altruists are fitter than groups of selfish individuals. The question we now need to consider is how these two facts combine to determine whether altruism will evolve. We appear to have here a "conflict of interest" between what is good for the individual organism (selfishness) and what is good for the group (altruism). Which of these influences will be stronger?

If natural selection occurs within the confines of a single breeding population, then altruism cannot evolve. Even if the population somehow manages to be 100 percent altruistic, sooner or later a mutant or migrant selfish individual will appear. That selfish individual will be more reproductively successful than the average altruist, so selfishness will increase in frequency. In the next generation, selfishness still will be the fitter trait, and so the process eventually will carry the population all the way to 100 percent selfishness. This is the process that Dawkins (1976) called "subversion from within."

Matters change if we consider a more complicated scenario. We need

*From a biological point of view*

to consider an ensemble of several populations, each with its own mix of altruism and selfishness. There is selection within each of these populations as well as selection between populations. The within-group process favors selfishness, but the between-group process favors altruism. What will be the net effect of these two conflicting processes?

The criterion for altruism to evolve is simply that altruism must be fitter than selfishness. True, within each group, altruism is less fit than selfishness. But this does not ensure that altruism is less fit than selfishness when we average over the groups. That is, to see how it is possible for altruism to evolve, we must see that the following argument is fallacious:

Within each group, altruists have a lower fitness than selfish individuals.

Hence, altruists have a lower fitness than selfish individuals in the ensemble of groups.

The premiss is part of the definition of evolutionary altruism. However, the conclusion does not follow.

To explain why this argument is fallacious, I will consider a simple example. Suppose that there are two groups. The first is 1 percent selfish and 99 percent altruistic; the second is 99 percent selfish and 1 percent altruistic. Let there be 100 individuals in each group. The fitnesses of the two traits, both within each group and averaging over the two groups, can be extracted from Figure 1.1 as follows:

<i>Group 1</i>	<i>Group 2</i>	<i>Global ensemble</i>
1(S): 4	99(S): 2	100(S): 2
99(A): 3	1(A): 1	100(A): 3

Notice that within each group, altruists are less fit than selfish individuals ( $3 < 4$  and  $1 < 2$ ). However, the global average is such that altruists are fitter than selfish individuals ( $3 > 2$ ).

This decoupling of what is true within each group and what is true in the ensemble of groups is puzzling. It seems even more paradoxical if we consider what these numbers imply about the frequencies of the two traits in the next generation. What will happen is that altruism will decline in frequency within each group, but will increase in frequency in the two-group ensemble. The evolution of altruism requires that the fitnesses of the traits fit a pattern that statisticians call Simpson's Paradox (Sober 1984, 1988b).

### *Did evolution make us psychological egoists?*

Although altruism increases in frequency during the one generation time slice I have just described, it will not do so in the long term, if the two populations remain intact. Subversion from within will eliminate altruism from each subpopulation. And if altruism is absent in each part, it must be absent from the whole.

What is required for altruism to evolve by group selection is that groups go extinct and found colonies at different rates. In addition, these extinction and colonization events must occur often enough to offset the process that occurs within each group wherein selfishness replaces altruism. When these various assumptions are satisfied, an altruistic trait can evolve despite the fact that it is disadvantageous to the organisms possessing it.

The ideas just outlined do not show that altruism has evolved, either in our species or in others. Rather, I have simply sketched the biological assumptions that must be true if group selection is to have this result. It is a matter of continuing controversy in evolutionary biology how often this type of selection process actually occurs.

### 3. PSYCHOLOGICAL ALTRUISM

Let us turn now from the evolutionary issue to the psychological one. Is psychological altruism ever a part of human motivation or are our motives thoroughly egoistic? Before this question can be answered, I must clarify what the two psychological concepts mean.

It is quite clear that we sometimes help others. It also is clear that we sometimes want to do this. Consider the example of parental care. Human beings take care of their children (very often, though, unfortunately, not always). Moreover, this is something that parents want to do; the helping behavior stems from a desire to help.

It does not follow from the fact that we help our children, nor from the fact that we want to provide such help, that we are altruistically motivated. To tell whether parental care is an example of psychological altruism, we must ask *why* we help our children and *why* we want to provide this help.

The thesis of psychological egoism maintains that parents want to take care of their children only because parental care provides some sort of benefit to the parents. For example, it might be argued that parents who take care of their children experience various pleasures and avoid feeling guilty. For them, the welfare of their children is not an end in itself, but is merely a means to some more ultimate selfish

### *From a biological point of view*

goal.<sup>3</sup> The contrary position, which says that the behavior in question is at least sometimes altruistic, claims that parents have an irreducible interest in having their children do better rather than worse. The welfare of one's children is an end in itself, not merely a means to some selfish goal.

To clarify the difference between these two theories about human motivation, we need to understand what it means to say that *we want X only because X is a means to attaining Y*. Egoism affirms, and altruism denies, a claim of this form. I'll need to distinguish what I call the *self-directed* and the *other-directed* preferences an agent might have in some situation. A self-directed preference describes the situation of the agent, but not the situation of anyone else. An example would be my preferring that I have more money rather than less. An other-directed preference describes the situation of someone else, but not of the agent. An example would be my preferring that you have more money rather than less. In addition to these two "pure" sorts of preference, there are "mixed" preferences as well, which describe both the situation of self and the situation of other. An example would be my desire that you and I have the same income. Mixed preferences, though ubiquitous, can be left to one side.

I will suppose that agents decide which action to perform in a situation by seeing which of the available actions (by their lights) maximizes the satisfaction of their preferences.<sup>4</sup> For example, suppose you are contemplating whether to send a check for \$25 to a charity. This action, you believe, will benefit malnourished children. It will cost you a modest amount of money. And there are the psychic consequences as well; you will think well of yourself and avoid feeling guilty. Suppose you send the check to the charity. The action you performed was the upshot of the various preferences you had. Perhaps you preferred that you feel good about yourself rather than bad; perhaps you also preferred that the children be better off rather than worse. For simplicity, I'll ignore additional preferences you may have had (including the one about money), and focus on just these two possibilities. Notice that one of them is self-directed and the other is other-directed.

There are four relationships that might obtain among these two possible preferences. They are depicted in Table 1.1. The numbers indicate a preference order; their absolute values have no significance.

Individuals with the preference structure I call *extreme egoism* care only about themselves; the welfare of others does not matter to them at all. Symmetrically, people with the preference structure I call *extreme altruism* care only about others, not about themselves. Each of these

## *Did evolution make us psychological egoists?*

Table 1.1

<i>Extreme altruism</i>		<i>Moderate altruism</i>		
	Other-directed preference		Other-directed preference	
	+    -		+    -	
Self-directed preference	$\left\{ \begin{array}{l} + \\ - \end{array} \right.$	$\left[ \begin{array}{cc} 4 & 1 \\ 4 & 1 \end{array} \right]$	$\left\{ \begin{array}{l} + \\ - \end{array} \right.$	$\left[ \begin{array}{cc} 4 & 2 \\ 3 & 1 \end{array} \right]$
	<i>Moderate Egoism</i>		<i>Extreme Egoism</i>	
	Other-directed preference		Other-directed preference	
	+    -		+    -	
Self-directed preference	$\left\{ \begin{array}{l} + \\ - \end{array} \right.$	$\left[ \begin{array}{cc} 4 & 3 \\ 2 & 1 \end{array} \right]$	$\left\{ \begin{array}{l} + \\ - \end{array} \right.$	$\left[ \begin{array}{cc} 4 & 4 \\ 1 & 1 \end{array} \right]$

structures says that people are motivated by a single kind of preference. Note that when people face choices in which self-interest and the welfare of others *coincide*, they will perform the same action, regardless of whether they are extreme egoists or extreme altruists.<sup>5</sup> When the choice is between upper-left and lower-right, both extreme altruists and extreme egoists select the former (since  $4 > 1$ ).

The other two structures depicted in Table 1.1 describe people who are motivated by two sorts of preference. *Moderate egoists* and *moderate altruists*, as I call them, care both about themselves and about others. However, they differ when they face a choice in which self-interest and the welfare of others *conflict*. When the choice is between upper-right and lower-left, a moderate egoist will place self-interest ahead of the welfare of others. A moderate altruist, in the same situation, will sacrifice self-interest for the sake of the other.

Moderate and extreme egoists are not willing to sacrifice self for the sake of another. Moderate and extreme altruists are willing to do so. The difference between altruism and egoism, when defined in this way, does not concern whether people *have* other-directed preferences, but whether self or other *matters more* in the choice situation at hand (Sober 1989).<sup>6</sup>



### *From a biological point of view*

In describing these four arrangements, I do not assume that a person will have the same preference structure from one choice situation to another. Presumably, a person might be an altruist in one situation but an egoist in another. Even if we are prepared to sacrifice our own happiness for the sake of our children, few of us are prepared to die to make a stranger smile.

I believe that much of human behavior involves choice situations in which self-interest and the welfare of others coincide. We decide to give the money to charity, and this action benefits both self and others. I hope it is clear that behavior in such choice situations is entirely ambiguous; the behavior does not tell us what our motives are. All four of the motivational structures just described prefer upper-left over lower-right. The fact that we give money to charity does not show that we are altruists; and the fact that we feel good when we donate to charity does not show that we are egoists. Only when there is *conflict* between self-interest and the welfare of others does behavior reveal whether we are altruists or egoists.

According to this way of seeing things, psychological egoism and altruism are behavioral dispositions. These dispositions should not be confused with the behaviors they sometimes occasion. Altruism is not the same as helping. In addition, the dispositions thus defined should not be confused with the individual motives that may underlie them. Altruists are people disposed to sacrifice self-interest for the sake of others, if the two conflict. Yet it is perfectly possible for altruists to care about themselves; they may have irreducibly self-directed preferences. Symmetrically, egoists are people who are unwilling to sacrifice self-interest for the sake of others, if those should come into conflict. Yet it is perfectly possible for egoists to care about others; they may have irreducibly other-directed preferences. A consequence is that we should not conflate the following two questions:

- (i) Do people ever have preferences concerning the welfare of others in which the welfare of others is not a means for advancing self-interest?
- (ii) Are people ever willing to sacrifice self-interest for the sake of others when these conflict?

An affirmative answer to (ii) entails an affirmative answer to (i), but not conversely. If people are willing to sacrifice self-interest, not only do they care about others; in addition, they sometimes give more weight to the welfare of others than they do to self-interest.

## *Did evolution make us psychological egoists?*

It follows that we must keep separate the following two egoistic theses:

(E1) People never have irreducible preferences about the welfare of others.

(E2) People never are willing to sacrifice their own interests for the sake of others when these come into conflict.

(E1) entails (E2), but not conversely. (E1) rules out three of the preference structures I have described; (E2) rules out only two.

Although I presented the four preference structures in order to clarify what (E2) asserts, the same ideas can be used to explain what the word “irreducible” means in (E1). Altruists have irreducibly other-directed preferences; the same is true of moderate egoists. Extreme egoists are the only people for whom the welfare of others is a mere instrument for promoting the satisfaction of self-directed preferences.<sup>7</sup> All four of these characters will donate the \$25 to charity, thereby causing both self and other to benefit. What is distinctive about extreme egoists is that they help *only because* they want to feel good about themselves.

### 4. PROXIMATE AND ULTIMATE EXPLANATION

Whenever natural selection causes a behavior to evolve, it must equip the organism with a mechanism that triggers the behavior in the appropriate circumstances (Mayr 1961). For example, if parental care is to evolve, the organism needs to have a device that causes it to dole out care to some organisms rather than to others.

For the organism to produce the selected behavior, it must possess both a *detector* and an *effector* (Williams 1966). For example, an organism may provide care to its offspring if it is programmed to drop food into gaping mouths. This will be a feasible solution to the design problem, if the parent can tell when gaping mouths are present and if the gaping mouths it sees are the mouths of its offspring.<sup>8</sup> If selection favors performing behavior *B* when and only when circumstance *C* obtains, the organism must have a detector of *C* and a way of producing *B* when *C* obtains.

Most organisms manage to evolve behavior without having minds as proximate mechanisms. Bacteria swim toward nutrients, but they do not form beliefs about where the nutrients are nor do they have desires concerning what is good to eat. We human beings, on the other hand, are guided in much of what we do by the beliefs and desires we have. It

### *From a biological point of view*

follows that if natural selection has caused human beings to behave in certain ways, it is plausible to suspect that natural selection has done this by equipping us with beliefs and desires that produce the selected behaviors in the appropriate circumstances.

If natural selection has worked strictly by way of individual selection, then it will have produced behaviors that are evolutionarily selfish. On the other hand, if group selection has played a substantial role, then some of the behaviors that result will be evolutionarily altruistic. In each case, natural selection must provide the organism with a proximate mechanism that causes the organism to produce the selected behavior. If our behaviors are evolutionarily selfish, should we expect them to be implemented by a mind that is psychologically selfish? And if our behaviors are evolutionarily altruistic, should we expect them to be implemented by a mind that is psychologically altruistic? In short, which proximate psychological mechanisms should we expect to find associated with strictly individual selection on the one hand and with group selection on the other?

In principle, evolutionary altruism can be implemented by psychological altruism *and* by psychological egoism. And evolutionary selfishness likewise can be implemented by the two psychological motivational structures as well. Suppose that natural selection has favored a behavior because the behavior benefits the group (and in spite of the fact that it harms the actor who performs the behavior). One psychological setup that will cause the organism to perform the behavior is for the organism to care only about the welfare of the group. But a quite different arrangement also can do the trick. Let the organism care only about feeling pleasure and avoiding pain, and let the organism be so constituted that how it feels is correlated with how well the group is doing.

The same point applies if we consider a behavior that is evolutionarily selfish. Consider, for example, the behavior of parental care. Parents enhance their own fitness by helping their offspring. Fitness, don't forget, reflects both the organism's chance of surviving *and* its prospects for reproductive success. In our species, as well as in many others, organisms that care for their offspring are fitter than organisms that do not. If individual selection has favored this behavior, and if the proximate mechanism for producing the behavior is to be a mind equipped with beliefs and desires, the question remains of what sorts of preferences that mind will contain. One possibility is that parents are made to care enormously and irreducibly about the welfare of their children. This paramount other-directed preference would mean that parents are

## *Did evolution make us psychological egoists?*

altruistically motivated. On the other hand, parents might care only about feeling pleasure and avoiding pain, and be so constituted that their good and bad feelings are correlated with how well their offspring fare. Under this scenario, parents dole out parental care for purely egoistic reasons. I conclude that there is no *a priori* reason why evolutionary altruism must be paired with psychological altruism, and evolutionary selfishness with psychological egoism (Sober, 1993a; Wilson, 1991).

Having separated the evolutionary and the psychological concepts, I now want to consider how they might be connected. Are there evolutionary considerations that help decide whether we are psychological altruists or egoists? The example behavior I will consider is parental care. Human beings provide enormous amounts of parental care, compared with many other species. I'm going to assume that this difference has an adaptive explanation. In saying this, I am not denying that there is variation within our species for the behavior. Child abuse and child neglect are all too common phenomena. I certainly do not deny their existence, although I will say nothing about why they occur. My interest is in between-species differences, not in within-species variation.

If we assume that natural selection played a role in causing human beings, on average, to provide considerable amounts of parental care, the question is whether we should expect natural selection to have done this by providing us with an egoistic or an altruistic motivational structure. That is, we are trying to judge the plausibility of the following two hypotheses about proximate mechanism:

- (A) Parents care about their children, not as means to the parents' own happiness, but as ends in themselves.
- (E) Parents care only about their own well-being, but they are so constituted that their own well-being is correlated with the well-being of their children.

Our problem is to discover whether natural selection would have favored one of these proximate mechanisms over the other.

In addressing the psychological problem in this way, I am ignoring a rather different avenue of inquiry. It might be suggested that we are sometimes psychological altruists, but that this does not have an adaptive explanation. That is, one might argue that this aspect of our psychology is *evolutionary spin-off* (on which see Sober 1988b). It is like our ability to do calculus or play with hoola hoops; though useless or even deleterious in itself, it was correlated with traits that were advantageous. This is an interesting suggestion, but one that I won't attempt to evaluate here.<sup>9</sup>

## *From a biological point of view*

### 5. DIRECT AND INDIRECT STRATEGIES

Once we consider altruism and egoism as alternative mechanisms for implementing a selected behavior, we can situate this problem within a broader evolutionary context. Consider the fact that fruitflies need to live in habitats that are humid. How might natural selection have caused *Drosophila* to move away from arid areas and toward areas of higher humidity? We can identify a *direct* and an *indirect* solution to this design problem (Williams 1966):

(D) Fruitflies have humidity detectors.

(I) Fruitflies have light detectors (not humidity detectors), but darkness and humidity are correlated in the flies' usual habitat.

Of course, (D) and (I) are incompletely specified; we must imagine that each detector is wired to a behavior effector in the right way.

I hope it is clear why I say that (D) is direct and (I) is indirect. If the organism needs to be in a humid locale, the direct solution is to give the organism a detector of that very property. It is more indirect to get the organism what it needs by giving it a detector of some other property that happens to be correlated with the property of interest. (I) is more in the nature of a Rube Goldberg device.

As it happens, (D) is false and (I) is true. Fruitflies have eyes, not humidity detectors. They fly toward darkness; that is how they manage to avoid drying out. So it is false that natural selection always provides direct solutions to design problems. But suppose you didn't already know which of (I) or (D) is correct (and were not allowed to just look and see). What considerations would be relevant to predicting which sort of mechanism the flies will evolve, given that they need a device that steers them away from dryness and toward humidity?

One consideration that is relevant to this problem is the relative *effectiveness* of the two strategies. If darkness and humidity are perfectly correlated, then (D) and (I) will be equally effective. But if the correlation is less than perfect, (D) will be more effective.

Since evolution does not always produce direct solutions, it seems clear that effectiveness is not the only factor influencing which proximate mechanism will evolve. In addition, there is the question of *availability*. We must consider what variants actually were present in the ancestral population in which the trait evolved. Perhaps (D) would have been more effective than (I), but (D) was not available. Natural selection favors the best trait *available*; there is no guarantee that the available traits include all traits that are *conceivable*.

## *Did evolution make us psychological egoists?*

A third consideration, beyond effectiveness and availability, also is relevant. Even if the ancestral population contains both light detectors and humidity detectors, and even if humidity detectors are more effective, it does not follow that natural selection will favor humidity detectors. The complication is that the two traits may have further consequences for the fitnesses of the organisms in which they occur. For example, suppose that humidity detectors impose huge energetic costs on the organism; suppose they require many more calories to build and maintain than do light detectors. This third consideration, which may be lumped into a category called *side constraints*, cannot be ignored. The effectiveness of a device in regulating some particular behavior of interest does not settle how fit the organisms are that have it (Sober 1981).

### 6. PSYCHOLOGICAL ALTRUISM IS A DIRECT STRATEGY

I hope the bearing of the problem about fruitflies on the problem about human motivation is starting to become clear. Psychological altruism, as described in (A), is a direct solution to the design problem of getting organisms to take care of their offspring. Psychological egoism, as described in (E), is an indirect solution. For fruitflies, we can simply look and see how they manage to find humid spots. However, when it comes to human motivation, the proximate mechanisms underlying parental care are harder to discern. Let us consider (A) and (E) as hypotheses about proximate mechanisms, whose plausibility we can judge in the same way we evaluated (D) and (I).

If a parent's well-being were perfectly correlated with the well-being of his or her children, then (A) and (E) would be equally effective. However, it is perhaps too much of an idealization to imagine that the correlation will be perfect. People occasionally find themselves in bad moods despite the fact that their children are doing well; and sometimes they find themselves in good moods despite the fact that their children are doing badly. If either of these two types of "error" occurs, (A) will be more effective than (E).

What about the availability of the two strategies? Here we may have a *disanalogy* with the fruitfly example. I suspect that *if (E) was available evolutionarily, so was (A)*. To see why this conjecture is plausible, let us consider in more detail how (E) would be implemented. According to (E), parents care only about their own well-being, but their well-being is

### *From a biological point of view*

correlated with the well-being of their children. How is this correlation established? Since we are talking about individuals who produce behaviors because of the beliefs and desires they have, we may suppose that these egoists reliably form beliefs about the welfare of their offspring. These beliefs cause the parents to feel good or bad. Since egoists care only about how they feel, they produce actions intended to help their children.

Notice that individuals implementing (E) must be able to form beliefs with the contents “my children are doing well” and “my children are doing badly.”<sup>10</sup> This suggests that if (E) is an available mechanism for producing parental care, so is (A). My reason for saying this is based on the following conjecture about how beliefs and preferences are connected: *If people can form beliefs about whether some proposition  $p$  is true, they also can form preferences about whether  $p$  should be true.* I am conjecturing that belief formation and desire formation are not so rigidly modularized that propositions available to the one are inaccessible to the other.

My suggestion is that the mental equipment needed for someone to be a psychological altruist does not differ in kind from the equipment needed for a person to be an egoist. Granted, the egoist desires instrumentally what the altruist desires as an end in itself; however, the basic belief/desire structure is the same.

These considerations, of course, do not absolutely guarantee that if (E) describes some of the people in an ancestral population, then (A) describes other people in that same population. An accident of history may have excluded (A) but have allowed (E) to make an appearance. What I am suggesting should be understood as a burden of proof argument. I do not see why (A) should have been unavailable evolutionarily. (E) and (A) look similar enough in terms of the basic equipment they require that some special consideration is needed to think that (E), but not (A), was present.

The final question, about side constraints, brings us to the frontier, beyond which there is only *terra incognita*. We know too little about the human mind, and about its evolutionary past, to say much about the side effects that (A) and (E) might have had on the organisms who provided parental care because of them. Within evolutionary theory, this confession of ignorance has different meanings for different biologists. Some biologists may incline to think that it is an acceptable simplification to assume that the trait of interest evolved independently of other traits. Others may regard this as radically implausible. This is an instance of the debate about *adaptationism*. I'm afraid that space does

## *Did evolution make us psychological egoists?*

not permit me to discuss the general issue or assess its seriousness for the problem at hand.<sup>11</sup>

In summary, (A) appears to be at least as effective as (E) in delivering parental care. And with respect to the issues of availability and side constraints, there appears to be no reason to think that (A) is less probable than (E). A bold conclusion to draw would be that (A) is more plausible than (E). A more circumspect conclusion would be that there is no evidence favoring (E) over (A).

### 7. A PUZZLE

In evolution, old characteristics often take on new functions. Sea turtles use their front fins to dig nests in the sand, but the fins did not evolve because they enabled turtles to do this. The fins were in place long before sea turtles emerged from the ocean to lay their eggs. The evolution of psychological motives for parental care is the mirror image of what happened to sea turtles. In the psychological case, a new structure took on an old function. Parental care is a much older characteristic than the cognitive and affective apparatus that provides its proximate implementation in our species. Beliefs and desires were grafted onto a set of behaviors that were at least partially in place before the advent of mind. The point is not that the human mind engendered no behavioral novelties. That is obviously absurd. Rather, my point is that parents were taking care of their offspring long before minds made their appearance.

To understand how mental features were grafted onto a preexisting behavior, it may be useful to compare the evolution of belief with the evolution of desire. No one suggests that our beliefs are solely about our own psychological states. Throw a rock at some people and they will believe *that a rock is approaching*; their belief will not be limited to a description of how they feel. This familiar fact presents an interesting puzzle. Why do beliefs reach out into the world? From the point of view of survival and reproduction, perhaps a belief mechanism would work just fine if beliefs were solely about internal states of the organism that happened to be correlated with external conditions. Solipsism may be *false*, but it is hard to see why solipsism should be *selectively disadvantageous*.

Regardless of how this puzzle is to be solved, it seems clear that we *do* have beliefs about events outside our own minds. The question I now want to pose is this: Why should desire be different? Why should



### *From a biological point of view*

our ultimate preferences be limited to our own psychological states? If beliefs reach out into the world, why should desires not do the same? Perhaps if we could understand why beliefs are about the external world, this might help resolve the puzzle about desire. This is a line of inquiry that I pursue in Essay 11.

## 8. CONCLUSION

I have argued that there is no evolutionary reason for preferring the hypothesis of psychological egoism. Neither effectiveness, nor availability, nor side constraints offers us a reason to think that human beings take care of their children for purely egoistic reasons. In the light of this conclusion, it is a curious fact that psychological egoism continues to dominate thinking in psychology and the social sciences. The very idea that people have irreducibly other-directed preferences, and that these sometimes are stronger than whatever self-directed preferences they have, is often viewed as naive. The implicit attitude of many theorists seems to be that if a range of behaviors *can* be explained by the hypothesis of egoism, then it *should* be so explained. The fact that the hypothesis of altruism *also* could explain the behavior is dismissed or ignored. However, it is worth asking why egoism should be accepted as the explanation of a behavior if the behavior also is consistent with the hypothesis of altruism. In the background of this bias in favor of egoism is the idea that egoism is the more parsimonious hypothesis (Batson 1991). Since we already have reason to think that *some* motives are egoistic, it is more parsimonious to suppose that *all* are ultimately egoistic.<sup>12</sup> I believe that there is less to this parsimony argument than might at first appear. If both hypotheses can explain the behavior, I don't see why parsimony provides much of a reason for favoring egoism.<sup>13</sup> In my opinion, psychological egoism does not deserve to be the default hypothesis that we blithely assume until we are forced to think otherwise.

Reinforcing the inclination of many social scientists to accept a selfish view of human nature is the belief that evolutionary theory leads to this very conclusion. But, in fact, evolutionary biology entails no such psychological consequence. Even if various aspects of our behavior (e.g., parental care) are *evolutionarily* selfish (i.e., were molded by individual selection), it is not immediately obvious why *psychological* egoism should have been the proximate mechanism that evolution pro-

## *Did evolution make us psychological egoists?*

vided to implement the behavior. Quite the contrary; it is rather puzzling why creatures whose fitness depends on the welfare of others and who are capable of forming beliefs and preferences about a wide range of propositions should be so constituted that they care only about themselves. Psychological egoism is not at all what one would expect from the point of view of evolutionary biology.

### NOTES

1. There is a fine point of disanalogy here: If a behavior is an example of psychological altruism, then it *must* proceed from motives of a certain sort. However, if a behavior is an example of evolutionary altruism, it may simply have arisen yesterday by mutation; it *need not* have been the result of group selection. Nonetheless, if further assumptions are adopted, the hypothesis of group selection is *de facto* a commitment that follows from the label of evolutionary altruism.
2. For example, I won't discuss whether kin selection and game theoretic interactions are properly regarded as varieties of individual selection or of group selection. On this, see Sober (1992a, 1993b).
3. Many philosophers seem to believe that Joseph Butler refuted egoism, at least in its hedonistic form. I do not. My objections to Butler's argument are given in Sober (1992b).
4. As used here, "satisfaction" does not name a sensation, but denotes a logical relation between an outcome and a preference. If I want it to rain tomorrow and rain occurs, my desire has been satisfied, even if I never learn that it has rained; in this case, the desire is satisfied even though I receive no pleasure. My thanks to Daniel Hausman for impressing on me the importance of emphasizing this point.
5. Here and in what follows my description of what agents will do in a given situation should be understood as shorthand for what they will do, given their *beliefs* about the situation.
6. Notice that the taxonomy I have proposed entails that an egoist can care about the welfare of others. Moderate egoists prefer that others be better off rather than worse, quite apart from the effect this might have on their own well-being. For them, an interest in the welfare of others is irreducible. I nonetheless classify this motivational structure as an instance of egoism because such individuals are never prepared to sacrifice their own welfare for the sake of another's. They are egoists because their interest in the welfare of others is too weak to counteract the concern they have for themselves. They are egoists because they say "me first."
7. Egoism, understood in terms of thesis (E1), makes the following claim about the preferences an agent will have in a given choice situation:

## *From a biological point of view*

For each other-directed preference O, there exists a self-directed preference S, such that the agent has preference O only because O is an instrument for satisfying preference S.

We now can say that an agent has O only because it is an instrument for obtaining S precisely when O and S are related as follows:

	+O	-O
+S	4	4
-S	1	1

For example, extreme egoists may prefer that their children do well rather than badly, but their reason for doing this is simply that the welfare of their children is correlated with whether they themselves feel good or bad.

8. Of course, perfect precision isn't necessary. A device can make both type-one and type-two errors and still evolve, if it is superior to the other devices with which it competes. This is discussed in more detail in Essay 3.
9. Singer (1981) develops an argument of this type. His idea is that "if there are advantages in being a partner in a reciprocal exchange, and if one is more likely to be selected as a partner if one has genuine concern for others, there is an evolutionary advantage in having genuine concern for others" (p. 44). Notice that Singer is proposing a correlation between psychological altruism and being chosen as a partner in exchanges. It also is worth noting that Singer's argument relies on there being a behavioral difference between egoism and altruism. My analysis will be compatible with their being behaviorally indistinguishable.
10. The concept "my children" isn't essential. The point is that the beliefs must have contents that are *about other people*.
11. For discussion pro and con see Maynard Smith (1978), Gould and Lewontin (1978), Orzack and Sober (1994), and Sober (1993b).
12. Batson (1991, chapter 4) suggests that we should prefer the egoistic hypothesis on grounds of parsimony if there were no behavioral evidence favoring altruism over egoism. However, Batson argues that there *is* behavioral evidence that supports the hypothesis of altruism.
13. I develop more general reasons for being skeptical of such parsimony arguments in Sober (1988a) and in Essay 7.

## REFERENCES

- Batson, D. (1991). *The Altruism Question: Towards a Social-Psychological Answer*. L. Erlbaum.
- Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. Reprinted, Princeton University Press, 1981.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
- Gould, S., and Lewontin, R. (1978). The Spandrels of San Marco and the

## *Did evolution make us psychological egoists?*

- Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proc. Royal. Soc. London* 205: 581–98.
- Maynard Smith, J. (1978). Optimization Theory in Evolution. *Annual Review of Ecology and Systematics* 9: 31–56.
- Mayr, E. (1961). Cause and Effect in Biology. *Nature* 134: 1501–6.
- Orzack, S., and Sober, E. (1994). Optimality Models and the Test of Adaptationism. *American Naturalist* 143:361–80.
- Singer, P. (1981). *The Expanding Circle: Ethics and Sociobiology*. Farrar, Straus, and Giroux.
- Sober, E. (1981). The Evolution of Rationality. *Synthese* 46: 95–120.
- Sober, E. (1984). *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. MIT Press.
- Sober, E. (1988a). *Reconstructing the Past: Parsimony, Evolution, and Inference*. MIT Press.
- Sober, E. (1988b). What Is Evolutionary Altruism? In B. Linsky and M. Matthen (eds.), *New Essays on Philosophy and Biology (Canadian Journal of Philosophy Suppl. Vol. 14)*, University of Calgary Press, pp. 75–99.
- Sober, E. (1989). What Is Psychological Egoism? *Behaviorism* 17: 89–102.
- Sober, E. (1992a). The Evolution of Altruism – Correlation, Cost, and Benefit. *Biology and Philosophy* 7: 177–88.
- Sober, E. (1992b). Hedonism and Butler’s Stone. *Ethics* 103: 97–103.
- Sober, E. (1993a). Evolutionary Altruism, Psychological Egoism, and Morality: Disentangling the Phenotypes. In M. Nitecki (ed.), *Evolutionary Ethics*, SUNY Press, pp. 199–216.
- Sober, E. (1993b). *Philosophy of Biology*. Westview Press.
- Williams, G. (1966). *Adaptation and Natural Selection*. Princeton University Press.
- Wilson, D. (1991). On the Relationship Between Evolutionary and Psychological Definitions of Altruism and Egoism. *Biology and Philosophy* 7: 61–68.