Articles

BEYOND THE SOCIOBIOLOGICAL DILEMMA: SOCIAL EMOTIONS AND THE EVOLUTION OF MORALITY

by Alejandro Rosas

Abstract. Is morality biologically altruistic? Does it imply a disadvantage in the struggle for existence? A positive answer puts morality at odds with natural selection, unless natural selection operates at the level of groups. In this case, a trait that is good for groups though bad (reproductively) for individuals can evolve. Sociobiologists reject group selection and have adopted one of two horns of a dilemma. Either morality is based on an egoistic calculus, compatible with natural selection; or morality continues tied to psychological and biological altruism but not as a product of natural selection. The dilemma denies a third possibility—that psychological altruism evolves as a biologically selfish trait. I discuss the classical treatments of the paradox by Charles Darwin ([1871] 1989) and Robert Trivers (1971), focusing on the role they attribute to social emotions. The upshot is that both Darwin and Trivers sketch a natural-selection process relying on innate emotional mechanisms that render morality adaptive for individuals as well as for groups. I give additional reasons for viewing it as a form of natural, instead of only cultural, selection.

Keywords: altruism; Charles Darwin; group selection; morality; selfish-gene theory; social emotions; sociobiology; Robert Trivers

---

The seemingly immoral outcome of the struggle for existence undoubtedly accounts for the difficulty that many people have in accepting the theory of evolution.

—David S. Wilson (1997, S1–S2)

Alejandro Rosas is Associate Professor of Philosophy at the Universidad Nacional de Colombia, Ciudad Universitaria, Bogotá, Colombia; e-mail arosasl@unal.edu.co.

[*Zygon*, vol. 42, no. 3 (September 2007)]
© 2007 the author.
Journal compilation © 2007 by the Joint Publication Board of Zygon.
ISSN 0591-2385

In this essay I argue that a version of morality involving psychological altruism can be explained as an adaptation at the individual level. The argument exposes a dilemma that has entrapped the debate on the evolution of morality, even though it is avoidable following suggestions in the writings of Robert Trivers and even Charles Darwin. In the first two sections I clarify the concepts of biological and psychological altruism. The dilemma is introduced in the second section.

BIOLOGICAL ALTRUISM

A behavior is biologically altruistic when it benefits other organisms by producing a net reproductive loss to the altruist and to the genes encoding for altruism. Whether biological altruism can evolve is still a debated issue in the biological theory of social behavior (Sober and Wilson 1998). If selection processes operate at the individual level or at the level of the genes producing the individual traits that are selected for (henceforth Individual Selection: IS), as claimed by sociobiology or selfish-gene theory, biologically altruistic traits cannot be adaptive and there can be no selection for them. Under IS, altruistic traits are those that lose in competition against selfish traits. But if selection operates between groups and its force overrides the force of selection operating between individuals within a group, biologically altruistic behavior evolves through the differential survival of groups (henceforth Group Selection: GS) (Sober 1993; Sober and Wilson 1998).[1]

Some behaviors explained through IS are still commonly called altruistic because of their short-term costs to the altruist. These behaviors are not, strictly speaking, instances of *biological* altruism. Trivers gave this view a memorable expression: "Models that attempt to explain altruistic behavior in terms of natural selection are models designed to take the altruism out of altruism" (Trivers 1971, 35). William Hamilton's demonstration that degree of relationship is a predictor of altruistic behavior provided one such model. It takes altruism out of altruism, because altruism to kin is a gene's way of benefiting its copies in related organisms (Hamilton 1964; Dawkins [1976] 1989). In a similar way, Trivers's model of reciprocal altruism takes altruism out of altruism by denying a net cost to the altruist: "under certain conditions natural selection favors these altruistic behaviors because in the long run they benefit the organism performing them" (1971, 35).

Kin altruism and reciprocal altruism may retain their brand names, but they do not involve *biological* altruism. They evolve through natural selection operating at the individual level. Without GS, there is no real biological altruism in nature. Biological altruism shares a common fate with GS: They stand or fall together.

HUMAN ALTRUISM

The attempt to explain human altruism through the theory of evolution has made sociobiology notorious among the public and among scientists interested in human affairs. We would expect sociobiologists to deny that humans are biologically altruistic; some in fact do, and endorse a conception of morality that denies psychological altruism[2] as well. However, two conspicuous figures of selfish-gene theory, Richard Dawkins and George Williams, believe that humans are capable of psychological altruism and that the same trait that makes them psychologically altruistic makes them biologically altruistic as well.

Dawkins claimed in his *The Selfish Gene* (both editions) that we humans can “defy the selfish genes of our birth” and “deliberately” cultivate “pure disinterested altruism” because “We, alone on earth, can rebel against the tyranny of selfish replicators” (Dawkins [1976] 1989, 200–201). In other words, we are, or can become, *psychological* altruists by rebelling against the forces of nature that make us *biologically* selfish. Why does it make sense to speak of rebellion here? Certainly, only because Dawkins thinks that biological selfishness leads naturally to psychological selfishness and to nothing else. Psychological altruism cannot be the natural outcome of selfish genes. Humans can rise to benevolence only by rebelling against them.

This is exactly how Williams read those words and expanded their meaning in his 1988 paper on Thomas Henry Huxley. He read them as the rebirth of Huxley’s thesis of the combat that we humans lead against natural selection in the name of morality: “Huxley viewed the cosmic process as an enemy that must be combated. Mine is a similar but more extreme position, based on the more contemporary view of natural selection as a process for maximizing selfishness” (Williams 1988, 399). Williams means that natural selection maximizes biological and psychological selfishness at the same time. He thinks, with Dawkins, that the first leads naturally to the second. This is why it makes sense for him to ask: “How could the maximization of selfishness produce an organism often capable of advocating, and occasionally practicing, charity towards strangers and even towards animals?” (p. 399).

Both conspicuous sociobiologists believe that humans display psychological altruism. Moreover, they both believe that acting on altruistic motives will result in a net cost in fitness. We rebel against selfish genes when we act as psychological altruists. For both authors, psychological altruism entails biological altruism. But because they reject GS and group-level adaptations they do not view psychological altruism as an adaptation at all. Its existence results from the combination of evolutionary accident and the effects of culture and conscious teaching against the contrary force of IS. It is ultimately sustained as an artifact of human will that somehow manages to thwart the power of IS.

One may reasonably wonder why Dawkins and Williams avoid common cause with other selfish-gene theorists for whom morality is an individual adaptation. Trivers's model of reciprocal altruism is an IS theory of moral behavior and moral psychology, a model that takes altruism out of altruism, because there is no net cost to the donor. Richard Alexander (1987) expanded this theory into a biological explanation of morality based on reputation or indirect reciprocity, also an adaptation due to IS. Why should any sociobiologist, let alone prominent ones, admit of biological altruism in humans? The answer must surely lie in their unwillingness to cast off their belief in psychological altruism. On account of their belief in psychological altruism in humans, and because they believe that this entails biological altruism, they are at odds with standard sociobiologists, who believe in neither.

Beneath this disagreement, however, a deeper common ground exists. Williams and Dawkins are espousing an opposition between natural selection and morality, argued first by Huxley in *Evolution and Ethics* ([1894] 1995) and conveniently labeled "Calvinist sociobiology" by Frans de Waal (1996, 13). This label applies also to sociobiologists who explain morality through IS and understand it as Thomas Hobbes did, namely, as instrumental to a self-benefit (Hobbes 1651, 65–66; especially his definition of free gift, 66). This is clearly the case in *The Biology of Moral Systems* (Alexander 1987, 88, 118f., 190f.).[3] It also is the usual understanding of Trivers's reciprocal altruism. Assuming that psychological altruism is a necessary component of morality, "Calvinist sociobiology" includes two alternatives in the form of a dilemma. If genuine morality exists, IS cannot explain it (Williams, Dawkins); but if IS explains it, morality can only be what Hobbes believed it to be: a form of self-interest and of instrumental rationality (Alexander).

BEYOND THE DILEMMA

The dilemma does not affect selfish-gene theorists alone. *Mutatis mutandis,* it affects most theorists interested in evolutionary ethics. When these do not endorse selfish-gene theory, they usually believe that morality—that is, psychological altruism toward non-kin—can evolve only if GS is a real evolutionary force.[4] However, I intend to show that the dilemma is not compelling, that psychological altruism and a version of morality involving it can evolve through IS and instantiate biological selfishness.

Parental care is one unproblematic case where psychological altruism and biological selfishness go together. In species whose newborns cannot survive on their own, parental care is necessary for reproductive success and is biologically selfish. Yet, in species with intentionality, parental care requires psychological altruism for reasons of reliability (Sober and Wilson 1998, chap. 10). Psychologically altruistic parents will be more consistent across different contexts in their caring behavior. Selfish parents will tend to neglect their children in cases where altruistic parents will not. Given their higher reliability, altruistic parents will enjoy higher fitness. Altruistic motivations will evolve if they are genetically available. But psychological altruism toward kin is unproblematic in a way that cannot apply to psychological altruists toward non-kin. How can the latter secure biological selfishness?

Surprisingly, clues to resolve this issue were given by both Darwin and Trivers. The literature on the topic typically overlooks this. It forces both authors into one of the two horns of the dilemma. Darwin is read as the first one to have noted the tension between IS and human altruism. Interpreters credit him with solving this tension by working out an explanation in terms of group selection (Sober 1993, 91–92; Sober and Wilson 1998, 4; Richards 1987, 212–17). As regards Trivers, his sketch of an IS theory of moral psychology (1971) is usually read as implying a conception of morality based on self-interest. This way of reading Trivers has become standard in different disciplines. Reciprocal altruism is seen as involving a calculus of altruistic acts to produce future benefits. Authors usually agree that a similar mechanism was already recognized by Darwin who correctly labeled it as a "low motive"[5] (Williams 1988, 390; Richards 1987, 60). The following passage from an experimental economist follows the same interpretation:

> The explanatory power of inclusive fitness theory and reciprocal altruism (Hamilton 1964; Trivers 1971; Williams 1966) convinced a generation of researchers that what appears to be altruism—personal sacrifice on behalf of others—is really just long-run self-interest. . . . We do believe, however, that the evolutionary success of our species and the moral sentiments that have led people to value freedom, equality, and representative government are predicated upon . . . motivations that go beyond inclusive fitness and reciprocal altruism. (Gintis et al. 2003, 153–54)

In the following two sections I argue that both Darwin and Trivers explained the evolution of psychological altruism through a process of IS. Their views lie beyond the horns of the dilemma by implicitly invoking a sense of selection that is no longer metaphorical but surprisingly literal. In the evolution of human morality, natural selection works through interacting agents who consciously, and guided by social emotions, select altruistic partners for cooperative enterprises. Darwin's views are somewhat obscured by the fact that he was not explicit about the motivational underpinnings of moral praise and blame. However, the thesis that he was not relying exclusively on group selection deserves to be highlighted. The passages discussed below have not received due attention (Ruse 1980; Richards 1987, 212–17; Gould 2002, 133–35).

In the final section I address some aspects of the model that concern empirical plausibility. Although they were hardly discussed by Darwin or Trivers, they deserve discussion today.

TRIVERS: THE ADVANTAGE OF BEING AN ALTRUIST

Trivers gave three concrete examples of reciprocal altruism in nature. One of them was human altruism. As is usual in evolutionary biology, reciprocity and altruism are treated as behavioral types defined through fitness effects, not motives. But at the end of his essay, Trivers did something new: he attempted to illuminate the details and the biological function of the psychological mechanisms that regulate "the expression of altruistic and cheating impulses to the selective advantage of individuals" (1971, 43). He reviewed experimental data relevant to moral and social motivations in humans. In this context, he expressed his position regarding psychological altruism.

Focused on moral psychology and character traits, Trivers discusses friendship, gratitude, fairness, cheater detection, moralistic aggression, trustworthiness, guilt and reparation, and rules of exchange. He views all of these phenomena as mechanisms designed to control the strategy he called *reciprocal altruism,* which enables participation in cooperative interactions and their benefits. Trivers clearly thinks that altruistic motivation is one of those mechanisms. He claims that selection favors an ability to detect the motivations underlying the altruistic acts of others: "Selection may favor distrusting those who perform altruistic acts without the emotional basis of generosity or guilt because the altruistic tendencies of such individuals would be less reliable in the future" (1971, 50–51).

This passage distinguishes between the nature of the acts and the nature of the underlying motivations. I understand the reference to generosity and guilt as entailing psychological altruism: a concern for the welfare of others for the others' sake. Without this motivation, Trivers claims, altruistic *acts* will not be consistent throughout time, "because the calculating spirit that leads this subtle cheater now to compensate may in the future lead him to cheat when circumstances seem more advantageous" (p. 51). Trivers claims that psychological altruists are more reliable over time than those who act altruistically out of a selfish calculus. Those who calculate an altruistic act with a view to an ulterior benefit are psychologically identical to subtle cheaters. They display as altruists for reasons of convenience. The connection between reliability and altruistic motivation is plausible as it stands and points to one of the weaknesses of Hobbes's conception of moral behavior as instrumental to selfishness.

But why should this reliability entail a higher relative fitness for the psychological altruist? Why are psychological altruists serving their biological selfishness after all?[6] The reason lies in the fact that people usually assess the reliability of prospective partners. People choose partners only after judging the motives behind their acts. Motives are more important than acts because they disclose whether altruistic manifestations are calculated for the occasion or reliable over time. It is in the biological interest of all of us to be able to make these judgments accurately in order to be protected from cheaters. Selection favors "the ability to detect and discriminate against subtle cheaters" (1971, 50). Moreover, it is in the biological interest of all of us to avoid being marked as subtle cheaters. But given accurate discrimination, only psychological altruists can consistently avoid this. Selfish calculators cannot and are thus excluded from reciprocal cooperation and from its benefits. If one's fitness is significantly determined by the amount of cooperation achieved, selfish calculators suffer a significant decrease in fitness. In fact, the process by which humans choose altruists as partners in cooperative enterprises is the very same process by which individual selection favors the existence of psychological altruism.

It is worth stressing that Trivers was not naive in his acceptance of psychological altruism; he looked for support in the psychological literature of his time:

> . . . there is ample evidence to support the notion that humans respond to altruistic acts according to their perception of the motives of the altruist. They tend to respond more altruistically when they perceive the other as acting "genuinely" altruistic, that is, voluntarily dispatching an altruistic act as an end in itself, without being directed towards gain. . . . Krebs (1970) has reviewed the literature on this point and . . . concludes that, "When the legitimacy of apparent altruism is questioned, reciprocity is less likely to prevail." Lerner and Lichtman (1968) have shown experimentally that those who act altruistically for ulterior benefit are rated as unattractive and are treated selfishly, whereas those who apparently are genuinely altruistic are rated as attractive and treated altruistically. (Trivers 1971, 51)

Trivers's review of the social-psychological literature reveals the human disposition to condition cooperation with others on the perception of their motives. It presupposes a capacity to discriminate between altruistic motivations and a calculating spirit that unreliably prompts altruistic acts. This capacity would be pointless unless the motivations distinguished were real. However, what if this distinction turned out to be illusory? I return to this point in the last section.

There is a darker side to Trivers's sketch of a theory, because he explicitly claimed that selection will make us not only psychological altruists but subtle cheaters as well: "natural selection will rapidly favor a complex psychological system in each individual regulating both his own altruistic and cheating tendencies and his responses to these tendencies in others" (1971, 48). This design is less than perfect, but natural selection is not a perfect designer and does not guarantee a human psychology free from internal conflict. Our psychological makeup implies internal conflict, because we are designed as psychological altruists and subtle cheaters at the same time.

Nothing in this model for the selection of psychological altruism suggests a mistaken view of natural selection. The model rests on plausible psychological assumptions and connects these assumptions correctly to biological selfishness. We will return to the issue of the empirical plausibility after looking at Darwin's similar views on the evolution of morality.

DARWIN: GROUP SELECTION IN SYNERGY WITH INDIVIDUAL SELECTION

The received view of Darwin's evolutionary explanation of morality points out that he was the first to note the tension between morality and natural selection. He solved this tension only by bringing group or community selection into his explanation. The received view stops there. Yet Darwin, although he did appeal to GS, did not believe it to be anywhere near the complete explanation for the existence of moral traits. He certainly noted that IS would apparently work against the evolution and spread of morality within groups, but this was not his considered view. He was sensitive to the fact that selection against morality within groups would make the existence of groups of altruists all too mysterious. After noting that a tribe of predominantly altruistic individuals would promote the dissemination of moral qualities by conquering and supplanting other tribes with few or no altruists, he explicitly posed the inevitable question: How could moral traits increase *within* groups or tribes in the first place, if it is true that IS works against them?

> But it may be asked, how *within* the limits of the *same tribe* did a large number of members *first* become endowed with these social and moral qualities, and how was the standard of excellence raised? It is extremely doubtful whether the offspring of the more sympathetic and benevolent parents . . . would be reared in greater numbers that the children of selfish . . . parents. . . . Therefore it hardly seems probable that the number of men gifted with such virtues, or that the standard of their excellence, could be increased through Natural Selection . . . ; *for we are not here speaking of one tribe being victorious over another.* Although the circumstances, leading to an increase in the number of those thus endowed *within* the same tribe, are too complex to be clearly followed out, we can trace some of the probable steps. (Darwin [1871] 1989, 130; emphases added)

Standard discussions (Ruse 1980; Richards 1987, 214; Gould 2002, 134–35) have not acknowledged the implications of this passage. Darwin is here demanding an explanation to the increase of moral traits *within groups*. The question is crucial, because GS would not be able to spread morality unless the number of altruists increased first within a group. The problem, as Darwin notes, is that individual selection seems to work against moral qualities. But he mentions this problem in order to solve it, as the last sentence of the quotation clearly indicates. This sentence introduces his explanation of the process that led to an increase of moral virtues within groups. GS cannot play any part in this process. So GS relies on a process of individual selection going on in parallel. In his short sketch of the "probable steps," only two steps or stages are clearly distinguished in terms of the mechanisms involved.

Darwin first mentions long-term prudence, the ability to see, by way of self-interested calculus, that it pays to donate benefits to others in long-term relationships because this encourages reciprocity. This is the "low motive" that interpreters have called attention to: reciprocity interpreted as selfishness. To the degree that the individual foresees an exchange of benefits in the long term, this ability promotes cooperation, but not on altruistic motivations. Self-interested calculus is plainly a form of psychological selfishness and of biological selfishness as well. Those who do not have it fail to share in the benefits of social cooperation and have lower fitness.

This prudential ability may explain an increase in altruistic acts, but it does not explain the evolution of morality or social virtues. Obviously, Darwin did not conceive of morality as Hobbes did. Morality cannot depend on selfishness alone, not even in long-term prudence. Selfish calculators are not wholly reliable; they cannot support a stable moral environment, because their calculating spirit will lead them to cheat whenever the circumstances seem advantageous.

The second stage is the crucial one. It justifies a parallel to Trivers's argument for psychological altruism, although the relevant psychological details are not equally explicit. Darwin claims that social emotions such as the love of praise and the dread of blame were powerful factors in the evolution of the moral virtues. Is it possible to connect love of praise and dread of blame to the evolution of psychological altruism? The answer is yes, if praise and blame determine who shares and who does not in the benefits of cooperation and if both primarily judge motivations and character rather than external conduct. Do we praise and blame manifest conduct or characters and motivations? Can love of praise and dread of blame mold our motivations as well as our manifest behavior?

Some clues indicate that Darwin linked love of praise and dread of blame to altruistic motivations. He says repeatedly in chapter 4 of *The Descent of Man* that praise and blame, and love of the former and dread of the latter, result from the workings of sympathy. If sympathy entails a genuinely altruistic concern for others, and interpreters agree that it does in Darwin (Richards 1987, 218–19; Darwin [1871] 1989, 106), it is plausible to conceive of praise and blame as social practices that monitor whether members of our social group are motivated by an ultimate concern for others. In one passage, Darwin treats sympathy, praise, and blame as the original components of the human social instincts, explicitly distinguishing this view from the philosophical conception where selfishness is the whole foundation of morality. Even at an early period in the history of man, he says, "the social instinct . . . together with sympathy (which leads to our regarding the approbation and disapprobation of others), [had] served as the primary impulse and guide. Thus the reproach is removed of laying the foundation of the noblest part of our nature in the base principle of selfishness" ([1871] 1989, 121).

In chapter 5 of the same work he makes perhaps the only direct statement as to the objects of praise and blame. The passage allows an interpretation of praise as aimed at motivations and character, if the social approval it mentions is granted to conduct that *intends* the general good as an end in itself, not only to conduct that in fact produces it: “It is obvious, that the members of the same tribe would approve of conduct which appeared to them to be for the general good, and would reprobate that which appeared evil. To do good unto others—to do unto others as ye would they should do unto you—is the foundation-stone of morality” (p. 131). If we credit Darwin with a conception of sympathy where altruistic motivation plays an essential role, we can as well grant him a conception of praise and blame where character, and not just manifest behavior, is the target of evaluation.

One may still wonder whether praise and blame do not appear too late in human evolution to be able to account for altruistic motivation as a heritable trait, as opposed to a culturally transmitted one. Praise and blame are highly dependant on language, as Darwin himself stresses (pp. 109f.). This puts praise and blame in sharp contrast to sympathy, which appeared much earlier, probably at the same time as the parental and filial affections, which lie “at the base of the social instincts” (p. 105).

Yet morality requires altruistic motivation, which can exist only in intentional agents who distinguish between benefiting another for the self’s sake and benefiting another for the other’s sake. It also extends beyond relatives. Praise and blame, with their dependence on language and social rules, connect to the appropriate motivations of intentional agents and explain their extension beyond the family. Darwin, therefore, needs praise and blame to account for morality; and, although he believes that they belong to human nature since primeval times, one can reasonably ask whether this is plausible or whether they are cultural practices instead, that is, culturally and not genetically transmitted. I return to this point in the next section.

We can now appreciate the parallel between Darwin and Trivers. Darwin is saying that a concern for the approval of our neighbors will put evolutionary pressure toward psychological altruism. Because of the social practices of praise and blame, humans put pressure, and also are put under pressure, to develop psychological altruism and moral virtues. As in the case of Trivers, the practices of praise and blame enhance the fitness of those who have them, because these practices select for altruistic motivations, and thus for reliability in others. On the other hand, in a social context shaped by these practices, psychological altruism benefits individuals who evolve it as a heritable disposition, because they will more often be chosen as partners in cooperative interactions. Psychological altruism turns out to be biologically selfish. The process of selection happens within the group and does not need GS in order to work, although GS can enhance it, and Darwin surely believed it did.

OPEN QUESTIONS

Darwin and Trivers described very similar processes of IS for the evolution of psychological altruism. In both authors, psychological altruism—a concern for the welfare of others for the others' sake—evolves as an individual adaptation and is therefore biologically selfish. The basic idea in both is similar to the one developed by economist Robert Frank (1988) in his theory of the moral sentiments as commitment devices. Human beings are emotionally committed to psychological altruism and prefer partners who cooperate out of altruistic motivations rather than out of a selfish calculus. These emotional preferences function as selection forces in social interactions. Given that humans are accurate judges of the underlying motivations (individual selection ensures this, to protect us from cheaters), success in reaping the benefits of social cooperation will be greater for psychological altruists than for selfish calculators. Altruists will thus have higher fitness.

This model is conceptually cogent, but questions of detail could compromise its overall plausibility. I address here three questions in ascending order of importance. The first is more a conceptual than an empirical issue and is, in my view, hardly open. The second and third address empirical issues that are legitimately open for further research.

The first question arises from a conceptual confusion. Some people believe that it is logically inconsistent to talk of altruistic motivations conditioned to reciprocity. They interpret reciprocity as meaning "I will help her so that I may get something from her afterward." But altruism conditioned to reciprocity means something altogether different, namely: "I will respect and take your interests into account as long as I receive the same treatment from you." The altruistic part of this attitude entails that I am sincerely disposed to view others' ends as having the same weight as my ends. This deserves the name of altruism. The reciprocity part of this attitude entails that I will condition altruism to the existence of a similar altruistic disposition in others. This is fairness. Conditionality just means that altruism, to be sincere, need not violate a basic norm of fairness. Thus, reciprocal altruism is consistent with psychological altruism.

The second point is empirical. Suppose Trivers and Darwin are factually wrong and psychological altruism does not exist. People may believe, though wrongly, in an actual difference between selfish calculators and psychological altruists. They may have, on this false belief, a real preference for psychological altruists, but if this preference aims at an illusory property, selection for psychological altruists is not taking place. The process of selection based on preferences as Darwin and Trivers envisaged it does not really happen.

This undoubtedly is correct, but it cannot invalidate the model as long as the existence of psychological altruism remains an open issue. The prospects for confirming its existence are promising. Expert opinion and experimental evidence, both in social psychology and in experimental economics, has been shifting lately in favor of psychological altruism. Humans show a disposition not only to behave fairly but also to interact preferentially with those who show this disposition.[7] Assuming that empirical evidence speaks for, rather than against, psychological altruism toward non-kin, selection at the individual level can explain the existence of this motivation. Individual selection operates through the emotional commitment, typical for humans, to prefer altruists in cooperative interactions. If everybody prefers to enter into cooperative enterprises with psychological altruists, it pays biologically to be a psychological altruist, because many benefits are available only through cooperation. Psychological altruism turns out to be compatible with biological selfishness.

One last point concerns the nature/nurture divide. How are the social practices of praise and blame and our susceptibility to them (for example, the propensity to feel shame) best understood in relation to this issue? We saw that Darwin linked these practices to language. This could point to their late appearance in human evolution. Could these practices and the accompanying emotions be learned and culturally transmitted instead of biologically built into our nature? This is surely an open empirical issue. Such practices do not need to be products of natural selection in order for them to help spread psychological altruism within the group.

Herbert Gintis and collaborators (2003) have argued that the distinctive human capacity for building social institutions can account for the evolution of altruistic behaviors. They follow the hypothesis advocated by Elliot Sober and David Sloan Wilson for human evolution (Sober and Wilson 1998; Wilson 1997). According to them, egalitarian social norms and punishment of free riders synergized with the pressure of intergroup competition during human evolution. By attenuating the advantage of selfish individuals within the group, egalitarian norms retard the action of IS against altruistic behaviors, allowing for the force of group competition to effectively select for altruism. In their reasoning, such institutions do not rest in genetically inherited behavioral tendencies. They are culturally transmitted artificial devices, sustained in their evolutionary success by the presence of intergroup competition and the danger of group extinction (Gintis et al. 2003, 165).

They do not apply this idea to the social practices of praise and blame. But, borrowing from the social constructionists, praise, blame, and emotions such as shame, guilt, and the sense of fairness could be viewed as social institutions in the above sense. Even if social scientists should confirm their universal presence across cultures, they could still be very successful inventions that every culture had either independently adopted or copied from others. There must be some innate capacity to learn such emotional dispositions; the question is whether the innate capacity is of a general sort or a specific one. Deciding one way or the other depends on the outcome of current lines of empirical research. Among these is research into the neurophysiologic patterns of the emotions. If unique patterns exist for different social emotions, and these patterns match across cultures, this speaks for social emotions as biologically selected traits. Shame and guilt, for example, are included in Paul Ekman's provisional list of fifteen basic emotions, where *basic* means, among other things, "hardwired" during biological evolution (Ekman 1999). Another relevant line of research involves looking for homologies of the social emotions in other primates, a line prominent in de Waal's writings (de Waal 1996; Flack and de Waal 2000; Brosnan and de Waal 2003).

As in the case of psychological altruism, it is empirically plausible, in the context of current research, to view these emotions as innate, in the sense of being insensitive to environmental factors in development (Griffiths 2002). One important line of evidence comes from the conduct disorder known as pyschopathy that is attributed to subjects who are difficult, even impossible, to socialize. Experimental data show that psychopaths suffer from emotional shallowness and absence of guilt feelings. They do not show autonomic responses to distress cues in others. This emotional inability remains constant over their lifetime and is not correlated to socioeconomic status or to negative parenting practices. It reflects a permanent inability and suggests that the dysfunction is the result of neurocognitive impairments. Neuroimaging studies show that psychopaths suffer from amygdala dysfunction. Dysfunction in the orbitofrontal cortex is also possible (Blair 2001; 2003). This selective impairment suggests that our brain evolved a *concern mechanism* that explains the ability to feel for others—that is, psychological altruism (Nichols 2001).

The outcome of this empirical question will not, however, affect our stance toward the role of individual and group selection in the evolution of morality. Even if empirical evidence should favor conceiving of praise and blame as cultural institutions, the reason for believing that GS is necessary for the evolution of altruism as fairness would disappear. Practices such as praise and blame do more than merely attenuate the advantage of selfish individuals within the group; they alter substantially the cost/benefit calculations that decide whether group selection is or is not necessary. They impose such costs on selfishness that they invert the relative fitness of selfishness and altruism within the group. Because of their reliability, psychological altruists, rather than selfish free riders, are the fitter type, both socially and biologically. As a result, the role of group competition has to be reevaluated. GS is no longer required for the evolution of altruism; rather, its influence adds to the individual selection for altruism achieved through the action of praise and blame as cultural forces.

CONCLUSION

My purpose has been to challenge the sociobiological dilemma. The dilemma denies the bare possibility that psychological altruism toward nonkin is an instance of biological selfishness. This is far from impossible. A model based on human preferences for psychologically altruistic partners, where these preferences are underwritten by emotions and operate as selection forces through social interaction, is arguably present in both Darwin's and Trivers's treatment of the subject. It is both conceptually cogent and empirically plausible.

NOTES

I thank the Research Division of the Universidad Nacional de Colombia, project DIB 8005002, and an anonymous referee for comments.

1. Elliot Sober and David Wilson (1998) claim that indisputable evidence for GS exists in some species of parasites and in species with biased sex ratios. They also argue for the relevance of GS in human evolution.

2. Psychological altruism is defined by the motivation underlying altruistic acts, namely, when one desires the benefit of some other person not as a means to one's own benefit but as an end in itself (Sober and Wilson 1998, 6, 201, 228; Batson and Shaw 1991, 108; Rosas 2002, 97f.).

3. "Thus we have evolved to be egoists, but the 'particle of the dove in our makeup' that puzzled Hume is our evolved tendency to favor relatives, invest in reciprocators, and portray ourselves favorably to potential reciprocators" (Alexander 1987, 194).

4. A recent example: ". . . cooperation is based in part on the distinctive capacities of humans to construct institutional environments that limit within-group competition and reduce phenotypic variation within groups, thus heightening the relative importance of between-group competition, and hence allowing individually costly but in-group-beneficial behaviors to coevolve with these supporting environments through a process of interdemic group selection" (Gintis et al. 2003, 165).

5. This is Darwin's passage from *The Descent of Man*: "as the reasoning powers and foresight of the members became improved, each man would soon learn that if he aided his fellow men, he would commonly receive aid in return. From this low motive he would acquire the habit of aiding his fellows" ([1871] 1989, 130f.).

6. The morality described by Trivers, as well as the one describe below by Darwin, is advantageous both for the individual and for the group. The similarity to cooperation through mutualism is, however, superficial. Biologists use the term *mutualism* when an organism's selfish behavior inevitably transfers benefits to others without incentives to cheat. In the case of morality there is an incentive to cheat. Special devices must evolve to protect cooperators, in particular the ability to detect reliability through attribution of motivations and character.

7. See Batson and Shaw 1991; Sober and Wilson 1998; Gintis et al. 2003; Fehr and Fischbacher 2003; Fehr and Gächter 2002. Frank (1988) reports experiments that confirm a human preference to choose altruists as partners. See also Sheldon et al. 2000.

REFERENCES

Alexander, Richard. 1987. *The Biology of Moral Systems.* New York: Aldine de Gruyter.

Batson, Daniel, and Laura Shaw. 1991. "Evidence for altruism: Towards a pluralism of prosocial motives." *Psychological Inquiry* 2 (2):107–22.

Blair, R. J. R. 2001. "Neuro-cognitive models of aggression, the antisocial personality disorders and psychopathy." *Journal of Neurology, Neurosurgery and Psychiatry* 71:727–31.

———. 2003. "Neurobiological basis of psychopathy." *British Journal of Psychiatry* 182:5–7.

Brosnan, Sarah F., and Frans de Waal. 2003. "Monkeys reject unequal pay." *Nature* 425 (18 September): 297–99.

Darwin, Charles. [1871] 1989. *The Descent of Man*. *The Works of Charles Darwin*, ed. Paul H. Barrett and R. B. Freeman, Vol. 21. New York: New York Univ. Press.

Dawkins, Richard. [1976] 1989. *The Selfish Gene.* Oxford: Oxford Univ. Press.

De Waal, Frans. 1996. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals.* Cambridge: Harvard Univ. Press.

Ekman, Paul. 1999. “Basic Emotions.” In *Handbook of Cognition and Emotion*, ed. T. Dalgleish and M. Power, 45–60. Sussex: John Wiley.

Fehr, Ernst, and Urs Fischbacher. 2003. “The nature of human altruism.” *Nature* 425: 785–91.

Fehr, Ernst, and Simon Gächter. 2002. “Altruistic punishment in humans.” *Nature* 415: 137–40.

Flack, Jessica, and Frans de Waal. 2000. “Any animal whatever. Darwinian building block of morality in monkeys and apes.” *Journal of Consciousness Studies* 7:1–29.

Frank, Robert. 1988. *Passions within Reason: The Strategic Role of the Emotions.* New York: W. W. Norton.

Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr. 2003. “Explaining altruistic behavior in humans.” *Evolution and Human Behavior* 24:153–72.

Griffiths, Paul E. 2002. “What is innateness?” *The Monist* 85:70–85.

Gould, Stephen Jay. 2002. *The Structure of Evolutionary Theory*. Cambridge: Harvard Univ. Press.

Hamilton, William D. 1964. “The genetical evolution of social behavior.” *Journal of Theoretical Biology* 7:1–52.

Hobbes, Thomas. 1651. *Leviathan.* Ed. C. B. Macpherson. London: Penguin.

Huxley, Thomas Henry. [1894] 1995. “Evolution and Ethics.” In *Issues in Evolutionary Ethics*, ed. Paul Thompson, 111–50. Albany: State Univ. of New York Press.

Nichols, Shaun. 2001. “Mindreading and the cognitive architecture underlying altruistic motivation.” *Mind & Language* 16:425–55.

Richards, Robert. 1987. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior.* Chicago: Univ. of Chicago Press.

Rosas, Alejandro. 2002. “Psychological and evolutionary evidence for altruism.” *Biology & Philosophy* 17:93–97.

Ruse, Michael. 1980. “Charles Darwin and group selection.” *Annals of Science* 37:615–30.

Sheldon, Kennon M., Melanie Skaggs Sheldon, and Richard Osbaldiston. 2000. “Prosocial values and group-assortation within an N-person prisoner’s dilemma game.” *Human Nature* 11 (4): 387–404.

Sober, Elliot. 1993. *Philosophy of Biology*. Boulder, Colo.: Westview.

Sober, Elliot, and David Sloan Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard Univ. Press.

Trivers, Robert. 1971. “The evolution of reciprocal altruism.” *The Quarterly Review of Biology* 46:35–57.

Williams, George C. 1988. “Huxley’s Evolution and Ethics in Sociobiological Perspective.” *Zygon: Journal of Religion and Science* 23:383–407.

Wilson, David Sloan. 1997. “Introduction: Multilevel selection theory comes of age.” *American Naturalist* 150:S1–S4.