

ALTRUISM AND THE GOLDEN RULE

by Jonathan Goodman

Abstract. This essay addresses recent claims about the compatibility of the sociobiological theory of reciprocal altruism with standard Western formulations of the Golden Rule. Derek Parfit claims that the theory of reciprocal altruism teaches us to be “reciprocal altruists,” who benefit only those people from whom we can reasonably expect benefits in the future. The Golden Rule, on the other hand, teaches us to benefit anyone regardless of their intention or ability to return the favor, or as Parfit puts it, the Golden Rule teaches us to be “suckers.” I argue that this distinction is founded on a misconception of the nature of the theory of reciprocal altruism, which is sociobiological as opposed to moral, and that this distinction accordingly confuses *is* with *ought*. Sociobiological theories may explain underlying psychological motivations in individuals (and perhaps even in populations), but these theories do not prescribe any sort of moral behavior. Furthermore, the theory of reciprocal altruism does not imply mental states of which agents are aware. The unconscious motivations assumed by this theory are in fact compatible with certain formulations of the Golden Rule; I will accordingly argue for the view that certain words with moral content related to the Golden Rule—such as “altruism” and “selfishness”—exist only insofar as they are social tools, which can further the self-interests of an individual in any group.

Keywords: evolutionary biology; Golden Rule; group selection; morality; reciprocal altruism; social Darwinism; Ludwig Wittgenstein

THE POSSIBILITY OF EVOLUTION AND MORALITY

One of the most common English formulations of the Golden Rule is “treat others as you would like to be treated.” Some philosophers, most notably Derek Parfit (2011), argue that this and similar formulations prescribe the prosocial behavior of “unconditional altruism.” Herbert Gintis et al. (2006) argue that persons who practice unconditional altruism will be overcome (evolutionarily speaking) by those who practice “strong reciprocity,” which consists of benefiting only those whom we believe observe the cultural norms of cooperation. Such behavior helps prevent our being taken advantage of by others, and increases the odds that our social circles

Jonathan Goodman is a postgraduate student in philosophy at Birkbeck College, University of London. He may be contacted at Department of Philosophy, Birkbeck College, Malet Street, London WC1E 7HX, UK; e-mail: uctyjrg@live.ucl.ac.uk.

will consist only of similarly prosocial individuals. This theory may follow from Robert Trivers's (1971) theory of reciprocal altruism, which suggests that benevolent behavior tends to be evolutionarily favored over egoism.

Parfit argues that reciprocal altruism and related sociobiological theories¹ are incompatible with the Golden Rule because such theories promote conditional as opposed to unconditional altruism, whereas the Golden Rule teaches us to be what sociobiologists call "suckers." If I follow the Golden Rule I will help someone regardless of her intention or ability to help me in return, but if I am a "reciprocal altruist," as Parfit puts it, I will help only those who are likely to help me in return. There is a self-interested aspect to these sociobiological theories that is not present in the most common English formulations of the Golden Rule.

Charles Darwin ([1871]2011) maintained that a regrettable part of human nature is that a person should act altruistically only toward another when he expects something in return. Two things are interesting about this: first, that if such behavior is intrinsic to being human, it is strange that anyone, including Darwin, would find it immoral; and second, that unlike George Williams (1966), Darwin appears to assume that the motivation for such selfishness is conscious. A selfish person is far more successful in her actions if no one is aware that she is egoistic, and people tend to treat others they regard as altruistic better than those they believe are selfish.

Parfit's objection to the compatibility of the Golden Rule and the theory of reciprocal altruism rests on certain assumptions that are subject to challenge. I suggest that reciprocal altruism explains certain human tendencies through its fundamental role in human psychological development. People tend to act selfishly or altruistically not because it is right or wrong, but because they are conditioned (genetically and/or culturally speaking) in certain ways. Unconditional altruism is as peculiar as extreme and obvious selfishness. The famous case of George Price giving up all of his earthly belongings for the good of others is as ludicrous as meticulously obeying the dicta of Ayn Rand. Sociobiology is explanatory but not prescriptive.

Nonetheless, Darwin's claim about human nature would probably be accepted by most people. Very few nonphilosophers (and even few philosophers) would defend the view that selfishness is a quality that everyone should regularly exhibit. In fact the word "selfishness" has such a deep stigma attached to it that most people would be reluctant to describe themselves as remotely selfish in public settings. No religious historical figure has ever proclaimed that "everyone ought to behave selfishly toward his fellow man." Religion, among many other things, helps to promote prosocial behavior between its constituents. But in spite of venerating what many people call "true" altruism, many theorists insist that we are actually selfish; we only fool ourselves into believing otherwise.

But if that is true, it seems strange that natural selection, as some evolutionary psychologists argue, would have instilled in us the further belief

that selfishness is wrong. On the other hand, Jesse Prinz (2007), as well as Sharon Street (2006), argue that although evolution has given us certain tendencies, we often disregard them in favor of moral principles produced by culture. Although I do not doubt the importance of enculturation, we must necessarily accept that evolution has given us the *ability* to question our natural tendencies.² But culture (and perhaps indirectly evolution) gives us a set of rules to follow and feelings to go with them; we just happen to call these things “manners” and “morals,” respectively.

THE POSSIBILITY OF UNCONSCIOUS SELF-INTEREST³

Let us accept that Darwin’s claim about human nature is generally true: those of us who tend to be the most successful—evolutionarily speaking—are those who tend to help only those who are likely to return the favor. George Williams sidestepped the moral problems of this claim by suggesting that there is no need for either party to be aware of the reciprocal properties intrinsic to altruistic actions. Most of us would readily admit that we are more likely to help a stranger if we believe he is honest than someone we think is lying, though Parfit (1984) contends that in a large society it makes no difference whether we help one or the other.

Robert Frank (1988) elucidated these unconscious components of altruistic-yet-self-interested behavior by arguing that personal belief in one’s own good intentions is a necessary quality for social and evolutionary success in altruistic societies. If a person believes she is “truly” altruistic and behaves accordingly, it raises the likelihood that others around her will believe that her actions are genuine (i.e., genuinely altruistic). This, in turn, increases the odds that others will help her in the future. Matt Ridley (1998) uses Frank’s theory to explain the motivation for voting when one knows it will make no statistically significant difference, or for leaving a tip at a restaurant where one has no intention of returning. These are both cases where personal beliefs promote an action that has little chance of being remotely beneficial to the agent. This person does not leave her waiter a generous tip with the expectation that he will give the money back to her at some point, or that he will tell others how wonderful she is. Instead she might believe that he is deserving of it, or that leaving a generous tip is the right thing to do; in either case, selfish motivations need not come into her conscious psychological picture.

The significance of this phenomenon can be highlighted by comparing two scenarios, where in the first we know only the following:

- (1) Yaroslav makes a large donation to a charity. He is known among his friends for his generosity, never fails to pay back his debts, and always tips generously at restaurants. He makes sure both to vote and keep his political affiliations relatively private.

And in the second scenario we know only that:

- (2) Vasili makes a large donation to a charity. He is known among his friends for his generosity and does not fail to pay back his debts unless his lender is unable to reach him. He only tips at restaurants he knows he will visit again. He does not vote but is exceedingly upfront with his political views.

Both scenarios are obviously extremes but most of us will agree that these “types” of people exist. What is important is that, if we were presented with either of these cases in some unbiased news publication where only this information were presented, every one of us would make certain judgments about both Yaroslav and Vasili, of which the most common would be that Yaroslav made his donation in a genuinely altruistic way, whereas Vasili *wanted* to appear altruistic, but instead made the donation for his own indirect benefit (i.e., public praise). This is true even though the actions they are receiving attention for—the charitable donation—are identical in both cases. What differs is only the smaller behaviors that give us insights into each of their characters.

The relevance of this contrast to the Golden Rule is the following: believing in and practicing certain formulations of the Golden Rule are indirectly good for us, in that our own beliefs tend to affect whether others perceive us as genuine. The benefits for the agent diminish gradually as society grows larger—as Parfit points out—but that does not mean that there is nothing to be gained for someone who is perceived as generous. Societies are large but communities, clubs, neighborhoods, or whatever, can still be quite small. Practicing, if not preaching, the Golden Rule still has its benefits, even if those benefits are not readily apparent to or expected by the agent.

If, several years after making his donation, each philanthropist loses his fortune, we could expect the public to be much more likely to come to Yaroslav’s aid than to Vasili’s. Yet this is probably not something Yaroslav expected when he made his initial donation: he was presumably not expecting these long-term benefits. Vasili, on the other hand, will probably not be in public favor, even though his altruistic action was identical to Yaroslav’s.

We have here a plausible case where reputation is more important to the agent than his altruistic behavior. Most people would rather be unknown than infamous for disingenuousness. It follows that those who not only practice, but *believe* that the Golden Rule is right—in the way that Frank describes—are more likely to be favored by others than people who are known to be selfish. Beliefs directly influence both public and private behaviors, but often it is how we behave when we think others are not watching that determines one’s reputation.

Of course none of this is true universally, or else we could reasonably expect everyone to be altruistic all the time, but as Robert Axelrod ([1984]2006) showed, certain decisions in prisoner's dilemmas are generally, but not always, favored over others. Much of moral decision-making is nothing if not a large-scale prisoner's dilemma; every day each of us tries to guess whether we should be altruistic based on the merits of others, and to ascertain whether we are being cheated in various situations.

This is true regardless of the formulation of the Golden Rule a person adheres to. It is easier to follow the Judaic form, which takes the negative form "do not do unto others . . ." rather than the Christian form, which takes the positive form "do unto others . . ." Depending on a person's background, he might believe that either of these or any other formulation of the Golden Rule is right, but he will still make subjective calculations for every circumstance where the rule might be followed. It may be easier not to hurt someone than to help him, but that does not mean that a person who believes the Judaic formulation is right will never harm someone. An employee told to fire a subordinate by his manager might believe in the Judaic formulation, but will still determine that he must fire the employee to keep his own job.

We can therefore reasonably say, if this kind of calculated decision-making is ubiquitous—and it obviously is—that the theory of reciprocal altruism does not "prescribe" behavior so much as it explains *how* and *why* we make certain decisions with regard to other people. These decisions will not always be "right," both in the moral sense and with respect to what is best for ourselves, but nonetheless we all possess the ability and psychological motivation to not want to be cheated, and to tend to help those who we believe most deserve it.

Parfit's objection to the possibility that reciprocal altruism is compatible with the Golden Rule is correct if we assume that conditional reciprocity is something that we *ought* to practice. But he mistakes an *ought* for an *is*. We are careful about maintaining cooperative norms because those who are not are exploited by those who tend to cheat. We have had to develop the ability to make precautionary psychological assessments about the people we deal with. This does not mean that we ought to *only* help those who are likely to help us in return; such an expectation, as the contrast of Yaroslav and Vasili above illustrates, can be detrimental to our well-being.

One important purpose of the Golden Rule, then, is to further disguise from ourselves the self-interested aspects of practicing reciprocal altruism. If I believe, as Parfit puts it, that I am a sucker, as the Golden Rule prescribes, others will be more likely to believe I am a Yaroslav-type than a Vasili-type. The less I know about the self-interested side of my own altruistic action, the better it is for me.

If this is true, the problem Immanuel Kant ([1785]2012) raises that the Golden Rule does not cover our duties to ourselves is to an extent

mitigated. The advantages for the agent of practicing reciprocal altruism are not readily apparent in any formulation of the Golden Rule. This is obvious insofar as no sane person would give money to a stranger who is a heroin addict over a homeless war veteran. We judge that the veteran deserves our charity more, that it is fairer that he gets the money. In neither case can we reasonably expect to be paid back, but any third party that happens to be passing will make judgments about us—the agent—based on whom we help and how we do so. Subconsciously we are all aware that good will from the public is as important as material wealth, though thousands of years of evolution and enculturation have hidden this self-interested side of altruism from us.⁴

THE POSSIBILITY OF THE PRACTICE OF THE GOLDEN RULE BEING BOTH ALTRUISTIC AND SELF-INTERESTED

It is possible to invert Kant's objection and claim that the Golden Rule *only* covers our duties to ourselves, or more broadly claim that all altruistic action is ultimately selfish. Such lines of thought are undoubtedly the motivations for popular philosophical movements that laud egoism as some moral ideal, as Ayn Rand did. Helena Cronin (1991) may have been partly motivated by the self-interested quality of altruistic action to try to solve the so-called "Problem of Altruism," which questions the possibility of "true" altruism in a world governed by natural selection. I contend that these views rest on mistaken definitions for "altruism" and "selfishness," which ought to be seen as entirely social.

Richard Alexander (1987) claims that neither biologists nor philosophers have been clear on how, exactly, to define altruism. But Thomas Nagel (1970) defines altruism to mean that the agent intends to act in the interests of others. Richard Dawkins (1979) conversely defines altruism only in terms of the effect an action has on others; the intention, biologically speaking, is irrelevant. Cronin applies this argument to certain birds who raise the offspring of other birds that imitate their own. Even though these birds believe they are raising their own chicks, Cronin argues that they are behaving altruistically by raising the young of others. Cronin calls this "true, albeit involuntary" altruism (1991).

Christine Clavien and Michel Chapuisat (2012) identified and qualified four different types of altruism that are regularly used in academic discourse (psychological, reproductive, behavioral, preference). These authors argue that different academic disciplines rely on one definition for altruism in debates with other disciplines that refer to something entirely different. This problem is especially clear if we imagine a biologist discussing the effect of some action with a philosopher, whose definition refers to the agent's intention.⁵

For the sake of simplicity I group psychological and preference into a *private* sense of altruism, which by definition no one but the agent can know, whereas behavioral and reproductive are *public*; they are observable to others. David Lahti (2003) makes a similar distinction between *ostensible* and *intentional* altruism, where an action is only *ostensibly* altruistic if it leaves open whether the agent *intended* to behave altruistically.⁶

The ultimate problem with each of these definitions is an epistemic one: we cannot know the minds (Nagel 1986) (and consequently the intentions) of others. Nagel's definition for altruism—which relies on intention—is impractical because we cannot know whether the agent is intending to act in the interests of others, or whether he only appears that way.⁷ Dawkins's and Cronin's definitions—which have only to do with effect—are impractical because people have a tendency to question the intentions of others, sometimes regardless of the results of the actions in question.

Here let us add a twist to the cases of Yaroslav and Vasili (remember that Yaroslav is known publicly for behaving altruistically in private engagements, whereas Vasili is known for behaving selfishly under similar circumstances). Yaroslav, whose action is lauded by the public, is actually making his charitable donation for tax purposes (i.e., for selfish reasons), whereas Vasili, who is publicly loathed, makes his donation because he believes it is the right thing to do. Public opinion of them remains the same even if these additional premises are true, because their intentions are private and thus not observable. The public has then made a mistake with regards to judgment of both persons. The possibility of such mistakes shows that we make nothing more than reasonably justifiable (but uncertain) assessments about the intentions of others behind their actions; moral judgments about other people, in other words, are to an extent “educated” guessing games.

One might be surprised at the possibility that in such extreme cases, Yaroslav's (private) intentions are selfish whereas Vasili's are benevolent. But such incredulosity is exactly what I wish to call attention to: we tend not to believe certain people are genuine based solely on the observable evidence we have. We are programmed—whether by evolution or by culture, or perhaps both—to make these judgments out of necessity; it is best for us to be associated with those who have proven themselves to be genuinely altruistic. That is precisely why the “twist” added to the cases of Vasili and Yaroslav is implausible. It is hard to believe that two people with such intentions would behave contrary to their dispositions, especially when those dispositions are altruistic ones.

Jack Wilson (2002) argues that for this reason biological altruism ought to be discussed only in ethology and not in philosophy. It is possible to imagine that some person might “accidentally” be altruistic in this way: if so-and-so makes a charitable donation at some event, assuming that

the donation is mandatory when it is optional, others may think she is behaving altruistically, even though the action in question was ultimately a mistake on the part of the agent. This is akin to some of the phenomena described by Bernard Williams ([1973]1999) and Nagel (1986) in their respective articles on moral luck. Wilson argues that in such cases the action is biologically altruistic, but not altruistic insofar as philosophers and laymen think of it, because our agent had no benevolent intention.

Similarly, an action can be perceived as selfish even if the agent has altruistic intentions. If the same agent makes a large charitable donation because she believes it is the right thing to do, others may claim she only made it for personal gain, let us say, for public praise. The effect of the donations are the same but the circumstances under which they are made, as well as the descriptions they are given, differ widely.

This phenomenon can be broadened to everyday circumstances under which people apply the Golden Rule. Some people follow it because they believe it is a moral truth that ought to be followed, whereas some people only follow it so as not to be called selfish by others. Intention is philosophically relevant but still unknowable to others; one only tends to—but does not definitely—increase the likelihood that others will believe their intentions are altruistic by believing that the Golden Rule is right.

We have, then, the possibility of an agent having altruistic intentions (i.e., believing that the Golden Rule is right), coupled with self-interested benefit (the public appearance of an action and the subsequent judgments made about our character). This possibility escapes both Kant's objection that the Golden Rule does not cover one's duties to oneself, as well as the inverted objection that the Golden Rule *only* covers one's duty to oneself. What matters is not so much the direct effect of the action as the judgments made about the action by others.

GIFT-GIVING AND THE SOCIAL USES OF MORAL IDEAS

Marcel Mauss ([1925]1967) shows that every person who is a member of some culture or society follows some set of customs or rules that give a structure to local "gift-giving." In some cultures these customs are so sophisticated that a gift is believed to retain a spiritual quality that is independent of the material object given.⁸ These various rules, customs, and beliefs about gifts create situations where individuals attempt to get the most from gift-giving by employing different tactics, as Axelrod would put it. People often make self-interested gains through gift-giving.

Mauss suggests that gift-giving, generally speaking, is a social tool that people often consciously or unconsciously employ for self-interested ends. An agent may give someone else a gift solely because he wants the person to do something for him. Similarly he might make the gift with the knowledge

that the favor cannot be returned, and wishes to make that publicly known. Many a family has been ruined throughout the world solely because they received a lavish gift such that they could not possibly return.

The larger point is that people often have self-interested reasons for gift-giving, and as a result we have developed hypersensitive—perhaps to the point of paranoid—mental apparatuses for making judgments about why we have been given something. Lev Tolstoy's Lukashka in *The Cossacks* immediately assumes that the prodigal Olenin wants something from him when the latter confers upon him a horse for no readily apparent reason. Olenin had altruistic intentions, as Nagel would put it, and the immediate effect was altruistic, as Dawkins would put it, and yet neither are sufficient for the action being called altruistic so much as stupid and tactless. We are missing some quality aside from intention and effect that is necessary for an action to be called altruistic.

For example, no one would think any better of a person who brings a gift to a friend's birthday party; they would, however, think *worse* of a person who failed to bring a gift for his friend on such an occasion. That person has failed to observe a custom of his culture and is judged accordingly.

But if another agent donates to a charity for underprivileged children's birthday gifts, we would tend to think well of her, especially if we should discover that she attempted to make her own gift anonymous. Such evidence points not to the agent's observing custom but to her benevolent character.⁹

A contrast between the person who brings a friend a birthday gift and the one who donates to the birthday charity shows—even if the gift for both recipients is identical—that what is important is neither the intention nor the effect of the action, but *the manner in which that action is judged by others*. The action is deemed altruistic or selfish by those who witness or hear about the action in question; this is akin to Ludwig Wittgenstein's ([1953]1997) claim that whether a student understands some formula is a judgment made by the teacher, and does not necessarily imply anything about the student's psychological state. Much as it is the teacher rather than the student who determines when it is that we ought to say, “the student understands this formula,” it is those that judge an agent's actions that determine whether he is altruistic; most adult members of a society are considered qualified judges simply because of enculturation. It is not for the agent to linguistically qualify himself as altruistic, and we tend to be wary of those who do.

Michael Ruse ([1989]1998) argues that the term “altruism”—and in fact morality in its entirety—developed as a set of tools that directly and indirectly promote a person's social (and thus evolutionary) success.¹⁰ But although Ruse argues that morality is altogether an illusion created by natural selection, I am claiming only that words with moral content,

despite how and why they have come into use, can be manipulated in ordinary language to the speaker's advantage. Just as a person would make judgments as to the motivations of Vasili and Yaroslav, who each make a charitable donation, so might Vasili claim that Yaroslav has made his donation for selfish purposes, or vice versa to make themselves look better by comparison. The possibility of manipulating moral discourse does not preclude the existence of morality altogether; it only highlights the idea that any kind of tool, whatever its origin, can be used to the advantage of a clever individual.

From this we can draw the larger conclusion that the use of the words "altruism" and "selfishness"—and perhaps of all words with moral content—is more about subjective differentiation than objective qualification. Whether the action is altruistic in the objective sense is irrelevant; there is no action that would uniformly be judged altruistic by all people under all possible circumstances. There will always be emotions such as greed, envy, jealousy, and the like, that will cause certain people to doubt even the most well-intentioned people. Either Vasili or Yaroslav might qualify as altruistic in the objective sense, and yet be considered selfish by someone. This is because it is natural to try to determine the intentions of the people we deal with or hear about; there is always the possibility that we are being fooled, and regardless how much we trust any person, the fact that we cannot *know* her intentions means there is always a degree of uncertainty.

Making oneself stand out as an altruist has far more to do with effectively using language and activity in a way that makes one appear altruistic; one's moral appearance takes precedence over reality. The game of much of moral discourse, in the public sphere, is about learning to appear altruistic effectively. Every altruistic act is only ostensibly altruistic—that is, not certainly altruistic because it is impossible to know the intentions of others. And it is also impossible to qualify some action as altruistic solely for the effect that action has.

I suggest that altruism be viewed not as some lofty, ideal quality that an action or a person has, but rather as a possible maneuver in the larger game of social competition. It is a praxis, to be used and described according to what is in the interests (or perceived interests) of the individual. Both a benevolent action as well as a vicious action can be described as altruistic; what matters is not the label "altruism" itself, so much as how, when, and by whom the label is used. This may well be the motivation for Wittgenstein saying: "It would not matter what you had done, you might even have killed somebody: what would matter would be how you talked about it, or whether you talked about it at all" (McGuinness 2005, 33).

But even if we accept Ruse's claim that morality is an illusion that hides our selfish nature from ourselves, it is unfair to call those people selfish who believe in and act from what they call moral truth. The fact that we

tend to think that such people are admirable is evidence that they are doing something right.

A PROBLEM FOR PROPONENTS OF “GROUP ALTRUISM”

Some theorists in economics, biology, and philosophy claim that altruism is a characteristic exhibited by members of groups that helps the group to function more effectively.¹¹ Katarzyna de Lazari-Radek and Peter Singer (2012) most recently argued that this “group” altruism is necessary for the possibility of universal benevolence, in the sense that Henry Sidgwick ([1874]1893) uses the term .

Gintis et al. (2006) argue that successful groups in Western societies are made of “altruistic punishers”—a quality that all “strong reciprocators” have—who behave negatively toward those who violate the norms of cooperation. Groups of altruistic punishers, these theorists argue, will fare better than groups of people who are either unconditionally altruistic or uniformly selfish. I argue in this section that “group” altruism is incompatible with the view of altruism I put forth in the preceding section of this essay.¹²

William Hamilton claims that

... with most traits that can be called social in a general sense there is some question. For example, as language becomes more sophisticated there is also more opportunity to pervert its use for selfish ends: fluency is an aid to persuasive lying as well as to conveying complex truths that are socially useful. (Hamilton 1998, 332)

More generally, George Williams claims that traits cannot develop that benefit a group without benefiting the individuals that comprise it. Language, the vehicle of moral discourse and moral judgments, can be perverted from what we think of as its more general purpose—that is, to communicate—to a tool that allows clever people to advance their own interests, without making their intention for that advancement outwardly noticeable. This is the spoken version of Mauss’s culturally universal, self-interested form of gift-giving.

Returning to the comparison between Vasili and Yaroslav, consider the same scenarios again, where in the first we know only that

- (3) Yaroslav makes a large donation to a charity. He is known among his friends for his generosity, never fails to pay back his debts, and always tips generously at restaurants. He makes sure both to vote and keep his political affiliations relatively private. Yaroslav is making the donation for self-interested reasons.

And in the second scenario we know only that:

- (4) Vasili makes a large donation to a charity. He is known among his friends for his generosity and does not fail to pay back his debts unless his lender is unable to reach him. He only tips at restaurants he knows he will visit again and where no one knows who he is. He does not vote but is exceedingly upfront with his political views. Vasili also makes the gift for self-interested reasons.

Here, we can reasonably expect that the public will look upon Yaroslav more favorably than upon Vasili, because the “facts” point to Vasili’s selfishness, whereas Yaroslav appears benevolent. The optimistic view put forth by proponents of “group” altruism suggests that cases such as this one range from highly improbable to impossible in advanced Western societies. It would not be possible for Yaroslav, in a group of “altruistic punishers,” to achieve his selfish goal without being found out. But despite what these theorists hope is true, some bad people do get away with bad things without ever being discovered.

Just as Yaroslav uses his charitable donation for self-interested reasons, so do many people use moral language-use to their advantage every day. The more experienced, clever, and rhetorically gifted a person is, the more she is able to steer moral conversations in the direction of whatever she is trying to promote. Moral discourse is a language game, and like any other it has a set of rules that are grounded in the culture and language in which the discussion takes place. The action in question matters far less than how it is described by the agent and the observers; the person thought of as the most altruistic might actually be the most self-interested.

And of course that is the nature of reciprocal altruism: just as a system of norms of cooperation can be infiltrated and manipulated to an egoist’s advantage, so can the system of norms be changed to account for the egoist’s tactic, if it is discovered. Another egoist comes and the process repeats itself.

Much as Parfit mistakes an *ought* for an *is* when assuming that evolutionary principles are prescriptive as opposed to explanatory, so do the proponents of “group” altruism mistake an *is* for an *ought*. The Golden Rule is not meant to steer us toward unconditional altruism, or even disguised self-interest, but to help us form the belief that fairness *ought* to be our most foundational principle, not our immediate desires, nor the immediate desires of others.

There will always be those who attempt to use moral discourse to their own selfish advantage, either by parading as an altruist through action or differentiating themselves with a clever use of words. Assuming that each member of our “group” is a “strong reciprocator” is to put oneself in danger of being taken advantage of by someone gifted in the art of social

manipulation. To assume that no such person exists is a mistake. We ought to treat well those whom we believe deserve it, but there is always a chance, despite all possible evidence, that we are mistaken. It is therefore a good idea to always be careful.

CONCLUSION

I have argued that: (1) evolutionary principles can explain human behavioral predispositions but do not directly prescribe moral belief or action; (2) widespread acceptance of the Golden Rule is evidence that the practice of an altruistic principle is not incompatible with evolutionary influences, and further that it is possible for the practice of a moral principle to be good for one without one realizing it; (3) some words with moral content are only social tools for differentiation, and (4) any tool for social differentiation can be used to one's personal advantage, regardless of any group affiliations that a person might have. We ought to practice fairness, but that does not mean that everyone does.

It is important to note, finally, that this essay is not at all concerned with prescribing moral values so much as with calling attention to patterns of human behavior that have been mistaken for evidence of universal selfishness or altruism.¹³ Predispositions do not imply anything about the morality of a person or a group of people, but rather explain the parameters around which moral values emerged after millions of years of enculturation.

ACKNOWLEDGMENTS

I would like to thank David Lahti, Paul Snowdon, Michael Otsuka, and Derek Parfit for their helpful comments on previous drafts of this article.

NOTES

1. For example, Gintis, Bowles, Boyd, and Fehr in *Moral Sentiments and Material Interests* (2006).

2. Kitcher (1993 and 1998) discusses the "evolutionary origins" of altruism and morality in an effort to elucidate the philosophical importance of "questioning our nature," which is to say, going against our instincts in an attempt to institute a change in accepted moral code.

3. I use the term "self-interest" to suggest the possibility that an agent may not be aware of the advantages to her of her actions. Hence an action may be both self-interested and selfish, but self-interest does not imply selfishness (or egoism, in Bernard Williams's [1973]1999 sense).

4. This is not to suggest that people who tend to be altruistic will be evolutionarily favored over people who tend to be selfish. Often a person who might otherwise behave selfishly behaves altruistically instead, because of certain circumstances. We ought not view people as "altruists," "egoists," "reciprocal altruists," and so forth; instead, we ought to look at the *tendencies* of a person's character to gauge whether we ought to form a relationship—of whatever kind—with them.

5. It is probably for this reason that Richard Alexander suggested that I do away with the use of the term "altruism" altogether, and talk instead of "beneficence" (personal communication).

6. Note that this has no bearing on whether the agent intended to *appear* to behave altruistically.

7. This is not, however, true for the agent's knowledge of his own intentions. One can apply Williams's ([1973]1999) "litmus test" for altruism by applying the following thought experiment: If I am a doctor who is trying to save a person's life, would I rather that the person's life be saved, but I go about believing that I have failed, or that I do not save the person's life, but go about believing that I saved her life? If I prefer the former, I know that my intentions are altruistic, but if I prefer the latter, I must accept that my intentions are selfish. I thank Cristian Constantinescu for bringing up this point.

8. Mauss quotes a Maori informant on this subject: "I shall tell you about *hau*. . . . Suppose you have some particular object, *taonga*, and you give it to me; you give it to me without a price. . . . Now I give this thing to a third person who after a time decides to give me something in repayment for it (*utu*), and he makes me a present of something (*taonga*). Now this *taonga* I received from him is the spirit (*hau*) of the *taonga* I received from you and which I passed on to him. The *taonga* which I receive on account of the *taonga* that came from you, I must return to you. . . . If I were to keep this second *taonga* for myself I might become ill or even die. Such is *hau*. . . the *hau* of the *taonga*. . ." (Mauss [1925]1967, 8).

9. It is interesting to note that in the hierarchy of the Judaic forms of charity, *Tzedakah*, anonymity takes precedence over almost any other quality the action might have.

10. De Waal (2006) attributes a similar theory to Thomas Huxley, who calls it "vener theory." De Waal argues that "altruism in nature," or rather behavior in animals that we call altruistic, serves as evidence against Huxley's view.

11. Here, I mainly refer to the works of Gintis et al. (2006), Sober and Wilson (1999), and De Lazari-Radek and Singer (2012).

12. Alejandro Rosas (2007) similarly argues against the idea that biological altruism is incompatible with self-interest. He claims that psychological altruism is evolutionarily favored; if people believe that you are altruistic, they are more likely to cooperate with you. But this ignores the possibility that some psychological egoist is just an excellent actor and is able to pass himself off as an altruist. Although Rosas's argument relies on the evolutionary success of the psychological altruist, the present issue for "group" theorists is based on the social way the terms "altruism" and "egoism" are used; psychological states merely predict social tendencies, but do not guarantee how one is publicly perceived.

13. I use the words "selfishness" and "altruism" is in terms of social use and not in terms of moral quality.

REFERENCES

- Alexander, Richard. 1987. *The Biology of Moral Systems*. New York: Aldine de Gruyter.
- Axelrod, Robert. [1984]2006. *The Evolution of Cooperation*, rev. ed. New York: Perseus Books.
- Clavien, Christine, and Michel Chapuisat. 2012. "Altruism across Disciplines: One Word, Multiple Meanings." *Biology and Philosophy* 28. doi:10.1007/s10539-012-9317-3
- Cronin, Helena. 1991. *The Ant and the Peacock: Altruism and Sexual Selection from Darwin to Today*. New York: Cambridge University Press.
- Darwin, Charles. [1871]2011. *The Descent of Man*. CreativeSpace Independent Publishing Platform.
- Dawkins, Richard. 1979. "Twelve Misunderstandings of Kin Selection." *Z. Tierpsychol.* 51:184–200.
- De Lazari-Radek, Katarzyna, and Peter Singer. 2012. "The Objectivity of Ethics and the Unity of Practical Reason." *Ethics* 123:9–31.
- De Waal, Frans. 2006. *Primates and Philosophers: How Morality Evolved*. Princeton, NJ: Princeton University Press.
- Frank, Robert. 1988. *Passions within Reason*. New York: W. W. Norton.
- Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr. 2006. *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge, MA: MIT Press.
- Hamilton, William. 1998. *Narrow Roads of Gene Land*, vol. 1 New York: Oxford University Press.
- Kant, Immanuel. [1785]2012. *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor and Jens Timmermann. Cambridge: Cambridge University Press.

- Kitcher, Philip. 1993. "The Evolution of Human Altruism." *The Journal of Philosophy* 90:497–516.
- . 1998. "Psychological Altruism, Evolutionary Origins, and Moral Rules." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 89:283–316.
- Lahti, David. 2003. "Parting with Illusions in Evolutionary Ethics." *Biology and Philosophy* 18: 639–51.
- Mauss, Marcel. [1925]1967. *The Gift*. New York: W. W. Norman.
- McGuinness, Brian. 2005. *Young Wittgenstein: Wittgenstein's Life, 1889–1921*. New York: Oxford University Press.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. London: Oxford University Press.
- . 1986. *The View from Nowhere*. Oxford: Oxford University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. London: Oxford University Press.
- . 2011. *On What Matters*. London: Oxford University Press.
- Prinz, Jesse. 2007. *The Emotional Construction of Morals*. London: Oxford University Press.
- Rosas, Alejandro. 2007. "Beyond the Sociobiological Dilemma: Social Emotions and the Evolution of Morality." *Zygon: Journal of Religion and Science* 42: 685–700.
- Ridley, Matt. 1998. *The Origins of Virtue*. New York: Penguin Books.
- Ruse, Michael. [1989]1998. *Taking Darwin Seriously*. Amherst, NY: Prometheus Books.
- Sidgwick, Henry. [1874]1893. *The Methods of Ethics*, 5th ed. London: Macmillan.
- Sober, Elliot, and David Sloan Wilson. 1999. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127:109–66.
- Trivers, Robert. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46:35–57.
- Williams, Bernard. [1973]1999. *Problems of the Self*. Cambridge: Cambridge University Press.
- Williams, George. 1966. *Adaptation and Natural Selection*, Princeton, NJ: Princeton University Press.
- Wilson, Jack. 2002. "The Accidental Altruist: Biological Analogues for Intention." *Biology and Philosophy* 17: 71–91.
- Wittgenstein, Ludwig. [1953]1997. *Philosophical Investigations*, trans. Gertrude Anscombe. Malden, MA: Wiley-Blackwell.