

Naturalism

James Rachels

[This essay originally appeared in *The Blackwell Guide to Ethical Theory*, ed. Hugh LaFollette (Oxford: Blackwell, 2000), pp. 74-91.]

Twentieth century philosophy began with the rejection of naturalism. Many modern philosophers had assumed that their subject was continuous with the sciences, and that facts about human nature and other such information were relevant to the great questions of ethics, logic, and knowledge. Against this, Frege argued that “psychologism” in logic was a mistake. Logic, he said, is an autonomous subject with its own standards of truth and falsity, and those standards have nothing to do with how the mind works or with any other natural facts. Then, in the first important book of twentieth century ethics, *Principia Ethica* (1903), G. E. Moore also identified naturalism as the fundamental philosophical mistake. Moore argued that equating goodness with any of the natural properties of things is “inconsistent with the possibility of any Ethics whatsoever” (Moore, 1903, p. 92).

Frege, Moore, and other like-minded thinkers inaugurated a period in which logic and language were the dominant philosophical subjects and confusing conceptual with factual issues was the greatest philosophical sin. During this period, philosophy was thought to be independent of the sciences. This may seem a strange notion, especially where ethics is concerned. One might expect moral philosophers to work in the context of information provided by psychology, which describes the nature of human thinking and motivation; sociology and anthropology, which describe the forms of human social life; history, which traces the development of moral beliefs and practices; and evolutionary biology, which tells us something about the nature and origins of human beings. But all these subjects were counted as irrelevant to the philosophical understanding of morality.

Of course, naturalism never disappeared completely—John Dewey was the century’s most influential naturalist—and in the latter part of the century it has made a comeback even among analytical philosophers. Naturalistic theories of mind, knowledge, and even logic are once again being defended. In ethics, naturalism remains under suspicion. Almost a hundred years after Moore, books of moral philosophy still routinely explain why it cannot be true.

I

Ethical naturalism is the idea that ethics can be understood in the terms of natural science. One way of making this more specific is to say that moral properties (such as goodness and rightness) are identical with “natural” properties, that is, properties that figure into scientific descriptions or explanations of things. Ethical naturalists also hold that justified moral beliefs are beliefs produced by a particular kind of causal process. Thus C. D. Broad observed that “If naturalism be true, ethics is not an autonomous science; it is a department or an application of one or more of the natural or historical sciences” (Broad, 1946, p. 103).

The most plausible form of ethical naturalism begins by identifying goodness with satisfying our interests, while “interests” are explained in turn as the objects of preferences. Protecting our eyesight, for example, is in our interests because we have desires that would be frustrated if we could not see; and that is why unimpaired eyesight is a good thing. Again, protecting children is a good thing because we care about children and we do not want to see them hurt. As Hobbes put it, “Whatever is the object of any man’s appetite or desire, that is it which he for his part calleth good” (Hobbes, 1651, p. 28). Reasoning about what to do, therefore, is at bottom reasoning about how to satisfy our interests. In what follows, I will elaborate this view and consider its merits.

Moore believed that no such view can be correct, for two reasons. First, he said, if we focus our attention on what we mean by “good” and what we mean by “satisfies our interests” we will see that they are not the same. We need only to think clearly about the

two notions to realize they are different. Second, Moore devised an argument against naturalism that came to be known as the “open question” argument. The question “Are the things that satisfy our interests good?” is an open question, and to say that those things are good is a significant affirmation. But, the argument goes, if goodness and interest-satisfaction were the same thing, this would be like asking “Do the things that satisfy our interests satisfy our interests?” An analogous argument can be given with respect to any other natural property with which goodness is identified. This seems to show that goodness cannot be identical with anything other than itself, and so, Moore concluded, ethical naturalism cannot be right.

Are these arguments effective? It depends on what, exactly, we take naturalism to be. Naturalism can be construed as a thesis about the meaning of words—that, for example, the word “good” means “satisfies our interests.” If this is how we understand naturalism, Moore’s arguments are plausible. (I will not discuss whether in fact they are sound.) But ethical naturalism can also be understood, more interestingly, as an idea about what goodness is—that it is, for example, the same thing as the property of satisfying our interests. That is how I understand it, and Moore’s arguments do not touch this idea at all. If his arguments were sound, they would also show that the Morning Star cannot be identical with the Evening Star. If we focus our attention on what we mean by those terms, we will see that they are not the same—the first is a star seen in the morning, while the second is a star seen in the evening. And the question “Is the Morning Star the Evening Star?” was an open question the answer to which was unknown for many centuries. But in fact the two are identical. (The same could be said about water and H₂O and about Lee Harvey Oswald and the man who shot John Kennedy.) So Moore’s arguments cast no doubt whatever on ethical naturalism, understood as a thesis about the nature of things.

There is, however, another argument that many philosophers believe is devastating to ethical naturalism. David Hume is credited with first observing that we cannot derive “ought” from “is.” In perhaps the most famous passage in modern moral philosophy, Hume wrote:

In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. This change is imperceptible; but is, however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. (Hume, 1739, p. 468)

Max Black dubbed this idea “Hume’s Guillotine”: factual judgments and evaluative judgments are fundamentally different, and no amount of purely factual information can logically entail any evaluation. It is commonly assumed that, if this is true, the naturalistic project is doomed.

But is it true? On the face of it, it appears that sometimes evaluative conclusions can be drawn from factual premises. P entails Q if and only if there is no possible world in which P is true and Q is false. But consider:

P: The only difference between doing A and not doing A is that, if we do A, a child will suffer intense prolonged pain. Otherwise, everything will be the same.

Q: Therefore, it is better not to do A.

It certainly seems that, in any world in which P is true, Q will also be true. Therefore, P entails Q. Why, then, do philosophers think there is an “unbridgeable chasm” between facts and values?

Hume himself believed that the gap between is and ought is a consequence of the fact that ought-judgments have a connection with conduct that factual judgments (and other deliverances of reason) do not have:

Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv'd from reason; and that because reason alone, as we have already prov'd, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason. (Hume, 1739, p. 457)

When a person has a moral belief—a belief that a certain way of acting is right or wrong, or that an action should or should not be done—this necessarily involves the person's being motivated to act accordingly. Suppose, for example, I say that gambling is wrong. But then you discover that I play in a high-stakes poker game every Friday night. You might find this puzzling. What is to be made of it?

There are several possibilities. (1) Maybe I lied; I do not really believe that gambling is wrong. Perhaps I was teasing you or speaking sarcastically. (2) Another possibility is that I was only reporting what other people believe—in saying “Gambling is wrong” I only meant something like “Most people think it is wrong.” I did not mean to be expressing my own view of the matter. (3) Or it may be that I am morally weak. Although I think it is wrong, and I resolve not to gamble, when Friday comes around I am overcome by temptation. Afterwards I kick myself for having succumbed. (The kicking can be avoided, however—it may be that I escape self-reproach by refusing to think very carefully about what I am doing. I want to gamble, so I do it but without giving it any thought. This is a common pattern: people often believe, for example, that they should give money to feed the starving rather than spending it on luxuries for themselves. So it is convenient for them to put the starving out of mind; then they can

have their luxuries without self-reproach. Later we will see that this sort of cognitive disengagement makes moral perception impossible.)

Other explanations might be tried. But the point is that my conduct requires some such explanation; otherwise it is unintelligible. I cannot gamble happily, without hesitation, and reflect on my conduct with no sense of self-recrimination, if it is true that I believe gambling is wrong. It is not necessary, of course, that I always do what I think is right, but it is necessary that I have at least some motivation to do it. This motivation need not be so powerful that it cannot be overcome by other desires. But it must supply some inclination, however slight, so that there is something there for the other desires to overcome.

Beliefs about what should be done, then, are necessarily motivating, to at least some extent. But aren't purely factual beliefs also motivating in the same way? Suppose I believe the room is on fire and I rush out. We might say that I left because of that belief. This explanation, however, is incomplete. It must be added that I didn't want to be burned. If I did not have that desire, the fire would have been a matter of indifference to me. Of course we may think it is too obvious for words that I don't want to be burned. But its obviousness should not cause us to overlook its importance in the explanatory scheme. Knowledge of facts together with appropriate attitudes prompts action; knowledge of facts alone does not.

This is why Hume thought that ought-judgments are "entirely different" from judgments of fact. Ought-judgments are motivating; factual judgments by themselves are not. But this is itself a puzzling feature of ought-judgments that requires explanation. What is it about ought-judgments that gives them this peculiar power "to produce or prevent actions?" How is their motivational content to be explained? Hume believed, sensibly enough, that we can account for this only by understanding moral beliefs to be associated with "sentiments," or feelings. Sentiments alone have the right kind of connection with conduct. If they are not expressions of sentiment, the motivational force of moral beliefs is inexplicable.

This, however, has a surprising implication for the idea that we cannot derive “ought” from “is.” Rather than explaining why such derivations are impossible, it helps to explain how they are possible. In 1964 Max Black offered this example:

Fischer wants to mate Botwinnik.

The one and only way to mate Botwinnik is for Fischer to move the Queen.

Therefore, Fischer ought move the Queen.

Black argued that this is in fact a valid chain of reasoning: if the premises are true, the conclusion must be true also. But the premises concern only matters of fact. They include no “ought” judgments. The conclusion, however, is about what ought to be done. Thus it seems that we can derive “ought” from “is.”

One might wonder whether the example commits some sort of subtle logical error. Perhaps the conclusion should be rewritten to say only “Therefore, Fischer wants to move the Queen,” since the major premise concerns only what Fischer wants and not what he ought to do. But this would render the argument obviously invalid. (Fischer may not want to move the Queen because he may not realize that this move would lead to mate.) Or one might wonder whether a premise about what ought to be done needs to be added. To satisfy the maxim “No ought in the conclusion without an ought in the premises,” the argument might be rewritten:

Fischer ought to mate Botwinnik.

The one and only way to mate Botwinnik is for Fischer to move the Queen.

Therefore, Fischer ought to move the Queen.

But the notion that Fischer ought to mate Botwinnick is decidedly peculiar. And even if this different argument is valid, how does that show that the original example was invalid? Why couldn’t the original one be valid as well?

We can, in fact, explain why Black's example works, by attending to the relationship between ought-judgments, reasons, and preferences. Any judgment about what should be done requires reasons in its support. If I say you should get out of the room, you may ask why. If there is no reason, then it isn't true that you should leave—my suggestion is merely strange. Suppose, however, I tell you the room is on fire. That provides a reason; and if you believe me, you will no doubt leave at once. But whether this is a reason for you will depend on your attitudes. If you want to avoid being burned, then the fact that the room is on fire is a reason for you to leave. In the unlikely event that you don't care whether you are burned, this fact may have no importance for you. It will not provide a reason for you to leave.

Other examples come easily to mind. Suppose you and I are sitting in a movie theater; we have come to see The General. I realize, however, that we are in the wrong theater. So I say we should move next door. Why? Because The General is being shown there, not here. Again, this is a reason for you to move only if you want to see that movie. If you don't care which movie you see, it is not a reason for you to move.

These examples illustrate one common type of practical reasoning. In each case, there is a judgment about what should be done (You should leave the room, We should change theaters) and a reason is supplied why this should be done (The room is on fire, The General is being shown next door). The fact that you have a certain desire (to be safe, to see The General) explains why the reason cited is a reason for you to do the indicated action. There is no mystery, then, why this pattern of reasoning is valid:

You do not want to be burned.

The room is on fire, and the only way you can avoid being burned is to leave.

Therefore, you should leave.

We want to see The General.

The only way we can see The General is to go next door.

Therefore, we should go next door.

One qualification needs to be added. It may be objected that these conclusions do not follow because there may be reasons not yet noticed why you should not leave the room, or why we should not go next door. Perhaps, for example, we should not go next door because a maniac is there shooting people. So it would not follow that we should go next door simply because we want to see The General. But this observation, while it is true enough, requires no fundamental change in our analysis. The new information will only form part of another argument of the same form: We don't want to be shot; there is a maniac next door shooting people; therefore we should not go next door. Whether we should go next door "all things considered" will just depend on whether we prefer missing to movie to being shot.

Black's example trades on the same features of practical reasoning. It says that Fischer should move the Queen, a judgment that is true only if there are good reasons in its support. Then just such a reason is provided (because moving the Queen is the only way to mate Botwinnik). And finally, the relevance of this reason is secured by asserting that Fischer has the required attitude (he wants to mate Botwinnik). If Fischer wants to mate Botwinnik, there is a good reason for him to move the Queen. So it follows that he should move the Queen.

Hume was wrong, then, to say that we can never derive "ought" from "is." But he was wrong for a reason that his own analysis exposes. If our premises include information about a person's relevant desires, we may validly draw conclusions about what he or she should do. This result is not out of keeping with the spirit of Hume's view. Indeed, it is probably better to express Hume's view as the idea that we cannot derive ought-judgments from facts about how the world is independently of our desires and other attitudes regarding it. That is the point of Hume's Guillotine.

We now have the rudiments of a naturalistic theory of reasons. According to this theory, attitudes are not themselves reasons, but they explain why facts are reasons. The reason you should leave the room is simply that it is on fire; but your desire to avoid being burned explains why this is a reason for you to act rather than its being a matter of indifference.

Sometimes “following reason” is just a matter of adopting the appropriate means to achieve what one already wants to achieve. The examples about the fire and The General are like this. Reflecting on such examples, we might be tempted to conclude that reason only directs action toward goals that are already fixed by pre-existing attitudes. We may call this The Simple Picture. According to the Simple Picture, each of us begins with a set of attitudes that are simply “given,” independently of thought and reflection. We want (or approve, or care about) some things and we don’t want (or approve, or care about) others. Then, reason comes in to tell us what we must do in order for those attitudes to be satisfied. Hume seemed to accept The Simple Picture when he wrote that “Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them” (Hume, 1739, p. 415).

But The Simple Picture is mistaken, because what people care about is itself sensitive to pressure from the deliberative process. The process of “thinking through” the facts surrounding an issue can affect one’s attitudes regarding it. Thinking things through can have the effect of strengthening the feelings one already has, but it can also cause feelings to weaken, to be modified, or to disappear; and it can cause new feelings to form. Thus our cognitive capacities can play a significant part in forming, shaping, and sustaining our attitudes. They need not merely “serve and obey” whatever attitudes we already happen to have.

This is a common phenomenon. Prior to deliberation, a person might have all sorts of attitudes, including, say, a love of smoking cigarettes, a hatred of Jews, and a resentment of demands that he contribute to charities. But then—if he is the sort of person who is willing to think seriously about such matters—he might consider that cigarette smoking can shorten his life; that contrary to the anti-Semitic stereotypes, Jews

are people much like everyone else; and that there are, after all, children starving. As a result, his feelings about such things might change.

Thinking through the facts is not a simple matter. Philosophers from Aristotle to David Falk have emphasized that after “the facts” are established, a separate cognitive process is required for one to understand fully the import of what one knows. It is necessary not merely to know the facts, but to rehearse them carefully in one’s mind in an impartial, nonevasive way. Aristotle even suggested that there are two distinct kinds of knowledge here: first, the sort of knowledge had by one who is able to recite facts “as a drunken man may mutter the verses of Empedocles,” but without understanding their meaning; and second, the sort of knowledge that one has when one has carefully thought about what one knows (Aristotle, 1147b). We all know, for example, in an abstract sort of way, that many children in the world are dying from starvation and easily preventable diseases; yet for most of us this makes no difference to our conduct. We will spend money on trivial things for ourselves, rather than using it to help them. How are we to explain this? The Aristotelian explanation is that we “know” the children are dying only in the sense in which the drunkard knows Empedocles’s verses—we simply recite the fact. Suppose, though, that we thought carefully about what life is like for those children. Our attitudes, our conduct, and the moral conclusions we reach might be substantially altered.

Thus moral judgments express attitudes, but not just any attitudes: they express attitudes that are evoked and sustained by the deliberative process. Thus, to say that an action is morally required is to say that one cannot help but feel a “sentiment of approbation” (to use Hume’s term) toward it when the full range of our cognitive capacities are functioning as well as they can.

This apparently leads to relativism. Because people's preferences could differ, even when they are being attentive and thoughtful, what counts as a reason for one person need not be a reason for another. However, naturalists emphasize that people do share many of the same values, because people are basically alike in their interests, needs, and psychological makeup. (This is an empirical matter; it could have been otherwise.) To complete this thought, naturalists have appealed to the idea of a common human nature, shared by all people—or at least, the vast majority of people—that gives rise to our basic values. Thus Hume wrote that “When you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it” (Hume, 1739; p. 469; italics added).

What is “the constitution of our nature” like, and what motivational structure does it sustain? Since Darwin, one of the most fruitful sources of information about human nature has been evolutionary theory. The most pervasive and general features of human psychology—our attitudes, dispositions, and cognitive capacities—can be viewed as the products of natural selection. This helps us to understand how those psychological features work and why we have them. Simply put, we have the psychological characteristics that enabled our ancestors to win the competition to survive and reproduce.

What is moral behavior like, viewed from this perspective? Moral behavior is, at the most general level, altruistic behavior, motivated by the desire to promote not only our own welfare but the welfare of others. Can we understand this as the product of natural selection? Darwin realized early on that this is not easy.

The problem is that altruism involves acting for the good of others even at some cost to oneself. Therefore the tendency to behave altruistically seems to work against reproductive success. The altruist increases the chances of others' surviving, by helping them, while at the same time decreasing the chances of his own survival, by giving something up. Therefore we would expect natural selection to eliminate any tendency toward altruism. Darwin saw this and wrote:

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 an average perish in larger numbers than other men. Therefore it seems scarcely possible (bearing in mind that we are not here speaking of one tribe being victorious over another) that the number of men gifted with such virtues, or that the standard of their excellence, could be increased through natural selection, that is, by the survival of the fittest. (Darwin, 1871, p. 163)

Nevertheless, the “social instincts,” as Darwin called them—the impulse to help others, even at some cost to oneself—obviously do exist, not only in humans but in other animals as well. It is most conspicuously manifested in the behavior of family members toward one another. What accounts for the selfless devotion of parents, when their sacrifices decrease their own chances of survival? Darwin was mystified. “With respect to the origin of the parental and filial affections,” he wrote, “which apparently lie at the basis of the social affections, it is hopeless to speculate” (Darwin, 1871, pp. 80-1).

The mystery remained unsolved until 1964 when W. D. Hamilton announced the theory of kin selection. Hamilton’s idea was based on the observation that many individuals are genetically similar to one another—typically, one shares half the genes of one’s siblings, one-eighth the genes of one’s cousins, and so on. Therefore, acting in such a way as to increase the chances of a genetically similar individual’s surviving is a way of increasing the chances of one’s own genes being passed on to later generations. This being so, we would expect natural selection to favor a tendency to altruism toward one’s near kin. This fits well with the phenomenon of altruism as we commonly observe

it: individuals do behave far more solicitously toward their relatives than toward strangers.

Viewed in this light, there is nothing mysterious about the self-sacrificial altruism shown by parents and siblings. It is no more than we should expect, given how natural selection operates. Of course, the point is not that individuals calculate how to ensure the propagation of their genes—no one does that. The point is that these are types of genetically-influenced behavior that will be preserved by same mechanism that preserves any other beneficial characteristic.

But not all altruism is kin altruism. Animals may be observed to sacrifice their own interests to help others who are not closely related to them, and this is more difficult to explain. To deal with this problem, evolutionary psychologists have turned to the idea of “reciprocal altruism.” The idea is that an individual performs a service for another because doing so increases the likelihood that a similar service will be performed for him—a monkey picks the external parasites off the back of another monkey, and then the favor is returned. It is easy enough to see that such reciprocal aid, when practiced by all (or even most) members of a group, will work to the advantage of all; but it is not so easy to see how, on the principles of natural selection, this sort of cooperative behavior could become established in the first place.

Perhaps the problem could be solved if we could regard the Darwinian “struggle for survival” as a competition between groups. Then we could argue that social cooperation gives the members of a group a competitive advantage—we could say that the members of groups that cooperate are more likely to survive and reproduce than members of groups that do not cooperate. But ever since Darwin, the struggle for survival has been conceived as a struggle between individuals, not groups. (This is reflected in the quotation from Darwin above, in which he says we must bear in mind that “we are not here speaking of one tribe being victorious over another.”) Some observers see no justification for such an exclusive emphasis on the individual as the unit of selection. R. C. Lewontin views it as merely an “article of ideological orthodoxy,

virtually unchallenged at present by any student of evolution, . . . that the individual organism is the object seen directly by natural selection”:

What is being explicitly denied is that characteristics favorable to the population as a whole will evolve by natural selection, except as a secondary consequence of the greater fitness of individuals over others within the population. So, for example, we are not allowed to claim that linguistic communication between humans was favored by natural selection by arguing that a group of protohumans who could talk to each other would be at an advantage in warfare or hunting over other groups who were without language. (Lewontin, 1998, p. 60)

Why not? Sober and Wilson (1998) have argued that group selection has an important place in the explanation of cooperative behavior.

Meanwhile, another possible solution to the problem has come from an unexpected source. The Prisoner’s Dilemma, a problem studied by game theorists, has had surprising implications for a variety of issues in ethical theory, and it may be useful in explaining the origins of reciprocal altruism.

In the Prisoner’s Dilemma, two players must choose whether to cooperate or defect, with neither knowing what the other will do. For each player the payoff for defecting is higher than for cooperating, no matter what the other does. Therefore, it seems that the winning move must be to defect. But if both defect, the payoff is less than if both cooperate. The payoffs are stipulated as:

Both cooperate: both get 3

Both defect: both get 1

One cooperates and one defects: the defector gets 5, the cooperator gets 0

So each player is tempted to defect because he gets the highest payoff if he defects while her “opponent” cooperates; while at the same time, both fear cooperating because if the

opponent defects they end up with zero. Nonetheless, both will be better off cooperating than defecting.

The Prisoner's Dilemma models a deep problem about social cooperation. Like the monkeys who need to have their backs groomed, we will all be better off cooperating than if each of us goes it alone. Yet, if any one of us could become a "free rider," gaining the benefits of other people's cooperation without cooperating ourselves, we would be even better off.

In such circumstances, what is the best strategy? The social theorist Robert Axelrod investigated this question by setting up a Prisoner's Dilemma tournament in which players had to decide repeatedly whether to cooperate. In each round, a player and his opponent would play the game 200 times. Each would decide, privately, whether to cooperate; and then the choices would be revealed and the payoffs noted. Then the process would be repeated.

Obviously, many strategies are possible. You could always cooperate or always defect; you could alternate cooperating and defecting; you could always do the opposite of what your opponent did on the previous move; and on and on. The possibilities are infinite. Axelrod arranged for his tournament to be played on computers and invited professional game theorists to submit programs that would play against one another. The winning program, called Tit-for-Tat, was submitted by Anatol Rapoport of the University of Toronto. After it was known that Tit-for-Tat had won the first tournament, a second was held and 62 programs were submitted in the attempt to find a better strategy; but again Tit-for-Tat was victorious.

Tit-for-Tat was surprisingly simple. It contained only two instructions: On the first move, cooperate. On each subsequent move, do whatever the other player did on the previous move. It is remarkable how these instructions correspond to common moral feelings. It is a "nice" strategy, in that it begins by cooperating, and it will continue to cooperate as long as the opponent is cooperating. It will never try to "take advantage" by being the first to defect. But it will not allow itself to be exploited more than once: when the opponent stops cooperating, Tit-for-Tat stops immediately. On the other hand, Tit-

for-Tat is also a “forgiving” strategy. When the other player resumes cooperating, Tit-for-Tat resumes cooperating immediately.

Ethical theorists have found this result to be important for various reasons. Contract theorists, who see moral requirements as arising from agreements of mutual self-interest, have cited the Prisoner’s Dilemma as exemplifying the kind of social situation which requires us to “bargain our way into morality”—we agree to cooperate by telling the truth, not harming others, and so on, so long as others obey those rules as well (Gauthier, 1986). Utilitarians, meanwhile, have taken the victory of Tit-for-Tat as indicating the kind of approach we should adopt in our dealings with other people if we want everyone to be as well-off as possible. (Singer [1995, 142-52] draws a number of conclusions of this kind.) For naturalists, however, it has a different significance.

Axelrod’s work might help us to understand how reciprocal altruism could have evolved. Suppose, as Axelrod’s research indicates, that in social settings Tit-for-Tat players end up better off than individuals who adopt other strategies. Then, in Darwinian terms, Tit-for-Tat players are more likely than others to survive and reproduce; and this means that any genes that dispose one to play Tit-for-Tat will be more likely to be represented in future generations. Thus we would have evolved as Tit-for-Tat players.

It would be a mistake, however, to think that because Tit-for-Tat won Axelrod’s tournaments, it has been shown to be the best strategy in any environment. Axelrod’s tournaments featured programs following a large number of strategies. But suppose there were only two strategies, Tit-for-Tat and a player (or players) who invariably defected. Then, Tit-for-Tat would lose every time. This poses a problem for the idea that social cooperation evolved because the tendency to play Tit-for-Tat was individually advantageous. It would be advantageous only if the environment already included other individuals disposed to cooperate. At this point some supplementary idea seems required—perhaps that individuals with a variety of behavioral dispositions appear as Darwinian “variations,” with Tit-for-Tat prevailing in the familiar way that advantageous variations prevail.

At any rate, none of this means that we consciously play the game. Rather, we come equipped with a set of emotions that dispose us to deal with others as though we were playing the game. At least three kinds of emotions are involved. Friendliness and a general beneficence dispose us to be helpful to others so long as they return our good will. But we take offense when others are hostile; we resent bad treatment from them and we stop treating them well. And finally, we are disposed to forgive people their past misdeeds, so long as they express remorse and take the first step toward reform. Together, these emotions cause us to make the major moves in Tit-for-Tat. It is this combination of emotional dispositions that natural selection preserves, in the same way that it preserves any characteristics that contribute to reproductive fitness.

Evolutionary naturalism has been an important option in recent philosophy, not only in ethics but in epistemology and philosophy of mind. Naturalists have sometimes made exaggerated claims about the importance of evolution for ethics. E. O. Wilson, for example, has argued that moral philosophy in the traditional sense is no longer needed, because biology explains ethics “at all depths” (Wilson, 1975, p. 3). The naturalistic theory we are considering here, however, is content with a more modest claim. An evolutionary understanding of the origins of altruism—whatever the details of that account might turn out to be—fills in some of the details of our conception of human nature. It helps us to understand what human beings are like, and why; and that in turn enables us to discern the values that human beings will normally have in common.

IV

Ethical naturalism says that goodness, rightness, and other such properties are “natural” properties. According to the version of ethical naturalism we are considering, which natural properties are they?

Moral properties turn out to be similar to what John Locke called “secondary qualities.” The distinction between primary and secondary qualities is tricky, but for our

purposes it may suffice to say, paraphrasing Locke, that secondary qualities are powers that objects have to produce effects in the consciousness of observers. Color is the classic (if disputed) example of a secondary quality. Primary qualities, such as a box's shape and mass, are what they are independently of observers. The box's shape and mass would be the same even if there were no conscious beings in the universe. But what of its color? Color is not a thing spread upon the box like a coat of paint. Rather, the box's surface reflects light-waves in a certain way. Then, this light strikes the eyes of observers, and as a result the observers have visual experiences of a certain character. If the light-waves falling on the box were different, or if the visual apparatus of the observer were different, then the box would have a different color, and this would no more be the box's "real" color than any other. The box's color, then, just consists in its power, under certain conditions, to cause a certain kind of observer to have a certain kind of visual experience.

Much the same may be said about other powers of things to induce other sorts of experiences in us. What does it mean for something to be sour? A lemon is sour because, when we put it to our tongues, we experience a certain kind of taste. What is sour for humans might not be sour for animals with different kinds of sense-organs; and if we were made differently, lemons might not be sour for us. Moreover, what is sour for one human being may not be sour for another—although we do have a notion of what is normal for our species in this regard. But despite all this, to say that lemons are sour is not a "subjective" remark. It is a perfectly objective fact that lemons have the power to produce the sensation in us. Whether they are sour is not merely a matter of opinion.

On the version of ethical naturalism we are considering, moral properties are properties of this kind—they are powers to cause us to have certain sorts of attitudes or emotions. (See McDowell, 1985) Being evil consists in having what it takes to provoke a thoughtful person to such responses as hatred, opposition, and contempt. When we think of a murderer and his victim, the (ordinary) facts of the matter are such that they evoke feelings of horror in us; the evil is simply the power to call forth this reaction.

Similarly, being good would consist in being so constituted as to evoke our support and approval.

Dewey compares “desirable” with “edible,” and the analogy is illuminating (Dewey, 1929, p.166). Whether something is edible is a matter of fact; if something is not edible, we cannot simply decide to make it so by adopting a positive attitude toward it. Yet whether something is edible for us depends as much on the kind of creature we are as on the kind of thing it is. If we were different, what is edible for us might be different too. However, we would not say that a certain food was inedible simply because, for special reasons, it could not be eaten by a few people. We have a conception of what is normal for humans, given the kind of creature a human being is, and what is edible for humans is what may be eaten by the representative human being, whether or not it can be eaten by everyone.

It may be thought that the analogy with “edible” breaks down at a crucial point. “Edible” means, roughly, “capable of being eaten,” whereas “desirable” (in the sense relevant to ethics) means, again roughly, “worthy of being desired.” So it may be thought that Dewey’s theory trades on a confusion between two senses of “desirable”—“capable of being desired” and “worthy of being desired.” The charge is that he thinks he has defined the latter, ethically relevant, sense of “desirable,” but really he has only given a definition of the former sense of the term, which is not relevant to ethics. But this charge is not well-founded. The essence of this sort of view is that to be worthy of being desired is to be capable of being desired, under the circumstances of thought and reflection. Perhaps this is not correct, but it is not a confusion of the theory: it is the theory itself.

This sort of view is, in an obvious way, a compromise between objective and subjective views of ethics. It is objective in that identifies good and evil with something that is really “there” in the world outside us, but at the same time, what is there is the power to produce feelings inside us.

Arguments against philosophical theories often turn out to be nothing more than thinly disguised descriptions of the theory, together with the insistence that it cannot be true. Naturalists have suffered from this treatment more than most. Despite his eminence as a logician, when John Stuart Mill said that “The sole evidence it is possible to produce that anything is desirable is that people do actually desire it” (Mill, 1861, p. 34), he was accused of committing a simple fallacy: it does not follow from the fact that something is desired that it is desirable. But Mill was only expressing the basic naturalistic idea that desirability is an empirical matter. We learn that a lemon is sour by tasting it; the proof of the sourness is the taste in our mouths. Similarly, we learn that something is desirable by contemplating it and as a result experiencing desire.

Complications arise, however, because such feelings can have other sources—a feeling might be the product of prejudice or cultural conditioning rather than proceeding from what Hume called “a proper discernment of its object.” The problem is how to distinguish between sentiments that indicate the presence of moral properties and those that do not. Moral properties are powers to provoke sentiments in us. But might we not have the sentiment in the absence of such properties?

The problem will be solved if there is a way to conduct the process of deliberation that eliminates those other sources—if we can set aside prejudice and other influences and take an objective look at the matter under consideration, in such a way that we will be justified in concluding that it is the thing itself that evokes our response. As Hume knew, doing this might require a considerable expenditure of cognitive energy:

But in order to pave the way for such a sentiment and give a proper discernment of its object, it is often necessary, we find, that much reasoning should precede, that nice distinctions be made, just conclusions drawn, distant comparisons formed, complicated relations examined, and general facts fixed and ascertained. (Hume, 1752, p. 173; also see Dewey, 1939, pp. 31-2, and Falk, 1986)

A sentiment indicates the presence of moral qualities only to the extent that it results from this kind of thinking. Otherwise, we cannot infer that the object itself has what it takes to cause the feeling.

The secondary quality view also fits well with the facts of ethical agreement and disagreement. There is, of course, a vast amount of agreement in ethics. We agree about good and evil for the same reason we agree about what is sour. Lemons affect most of us in the same way because we have similar sense organs; and murder affects us in the same way because we are similar in what Hume called “the constitution of our natures.” The similarity is not hard to understand: we have evolved, by natural selection, not only as moderately altruistic beings, but as creatures with common desires and needs. We all want and enjoy friends. We take pleasure in our children. We respond to music. We are curious. We all have reason to want a peaceful, secure society, because only in the context of such a society can our desires be satisfied. Such facts support large-scale agreements about respect for life and property, truthfulness, promise-keeping, friendship, and much more.

All this is compatible, however, with a degree of disagreement. As we have already observed, what is sour for one human being may not be sour for another—although we do have a notion of what is normal for our species. As a practical matter, however, we need never assume that moral disagreements are expressions of intractable differences between people. More mundanely, and more frequently, disagreement will be the result of ignorance, prejudice, self-deceit, and the like. Our working hypothesis may be that we are enough alike that we could be brought to agree about most things, if only the sources of error could be eliminated.

VI

The most important objection to ethical naturalism is that it leaves out the normative aspect of ethics. Since the whole point of ethics is to guide action, there could hardly be a more serious complaint. The objection can be expressed in various ways.

One way, which we have already considered, is to say that we cannot derive “ought” from “is.” Another is to say that ethical assertions are prescriptive, whereas their naturalistic translations are merely descriptive. Or it may just be said: look at the whole naturalistic account and you will find nothing that tells you what to do.

Is this objection sound? On the one hand, the version of ethical naturalism that we have considered does incorporate normative elements. On this view, ethics is action-guiding because of its connection with our desires and interests. Ethical reasoning engages our preferences, and the ethical “properties” of things are powers they have to influence our preferences. So we have some explanation of the normative aspect of ethics. Somehow, though, this does not seem to dispel the worry.

There is another way to put the complaint, due to Thomas Nagel (1997), that explains why the worry persists. We may compare our naturalistic account of ethics with naturalistic accounts of other subjects, such as mathematics. Why did Frege object to psychologistic theories of mathematics? Mathematics, he insisted, is an autonomous subject with its own internal standards of truth. Consider Euclid’s famous demonstration that there are an infinite number of primes:

Take any list of prime numbers you like, and multiply them together. Then add one. Call the resulting number n . n is not evenly divisible by any of the numbers on your list, because there will always be a remainder of one. Therefore, either n is itself a prime number, or it is divisible by some prime not on your list. Either way, there is at least one prime not on your list. Since this will be true no matter what list you begin with, it follows that there are an infinite number of primes.

If you think through this reasoning, and understand it, then you know that there are an infinite number of primes. But notice that the reasoning makes no reference to you, your cognitive capacities, your brain, your education, the social setting that produces and sustains mathematicians, the mathematical beliefs of people in your culture, the English language, or your understanding of the English language. And no information about any

of those things needs to be added to the proof: it is sufficient as it is. Moreover, no information about those other matters can cast doubt on the conclusion that the primes are infinite. That could be accomplished only by showing that there is something wrong with the proof itself. As Nagel puts it, in so far as the primes are concerned, the proof is “the last word.”

Now consider this example of ethical reasoning:

If we perform the operation on the child without first administering the anesthetic, she will be in agony. The anesthetic will have no ill effects on her; it will simply render her temporarily unconscious and insensitive to pain. Therefore, the anesthetic should be administered.

If you think through this reasoning, and understand it, then you know that the anesthetic should be administered. But like Euclid’s proof, this reasoning makes no reference to your cognitive capacities, your emotions, the beliefs of people in your society, and so on. And no information about these other matters can cast doubt on the conclusion that the anesthetic should be used. That could be accomplished only by showing that there is something wrong with the reasoning itself. This is the sense in which ethics is autonomous: the other matters do not figure into the reasoning about what should be done. In so far as the anesthetic is concerned, the reasoning is the last word.

If the reasoning is sufficient unto itself, what does our naturalistic account add? The naturalistic account explains what is happening as you think through the reasoning: you are considering facts that have the power to influence your attitudes, considering, in Hume’s words, “the constitution of your nature.” Your nature, formed by natural selection, includes feelings of protectiveness toward children and a disposition to care about other people, at least until they give you reason to resent them. Therefore, the result of thinking through the facts is a “sentiment of approbation” toward administering the anesthetic, which is expressed by the conclusion that it should be done. But this is what the process looks like when viewed from the outside. From the outside, we see

your human nature, your feelings, the operation of your cognitive capacities, and the interaction of all these with the facts about children and anesthetics. From the inside, however, you do not consider such things: you simply think through the argument.

This is the sense in which naturalism leaves out the normative. Naturalism provides a view from the outside, and from that perspective, it provides all sorts of interesting information. But it misses something that can be experienced only from the inside, namely the normative force of the reasoning. Examined from the outside, the normative aspect disappears, because when we begin to talk about the reasoning—quoting it, commenting on it—we cease to say anything that compels action.

So a naturalistic account of ethics seems deficient. But should we say it is wrong-headed and reject it altogether, or should we say only that it is incomplete? If we reject it altogether, we forego a great deal of understanding and insight into the nature of ethics. So long as we confine ourselves to thinking through ethical arguments “from the inside,” we can say nothing about the relation between moral thinking and the emotions, the nature of the properties to which moral terms apparently refer, or any of the other matters that have traditionally engaged moral philosophers. Indeed, we can say nothing about ethics at all—we cannot even say, along with the critics of naturalism, that it is autonomous. Perhaps our conclusion should be that, while moral thinking itself is done from the inside, moral philosophy is done from the outside.

References

Adams, E. M.: Ethical Naturalism and the Modern World-View (Chapel Hill: University of North Carolina Press, 1960).

Aristotle: Nicomachean Ethics, trans. W. D. Ross, in The Basic Works of Aristotle (New York: Random House, 1931).

Axelrod, Robert: The Evolution of Cooperation (New York: Basic Books 1984).

Black, Max: "The Gap Between 'Is' and 'Should'," The Philosophical Review, 73 (1964), 165-181.

Broad, C. D.: "Some of the Main Problems of Ethics," Philosophy, 31 (1946) 99-117.

Darwin, Charles: The Descent of Man, and Selection in Relation to Sex (London: John Murray, 1871).

John Dewey: The Quest for Certainty (New York: Capricorn Books, 1960; originally published in 1929).

John Dewey: Theory of Valuation (Chicago: University of Chicago Press, 1939).

Falk, W. D.: Ought, Reasons, and Morality (Ithaca: Cornell University Press, 1986).

Gauthier, David: Morals By Agreement (Oxford: Oxford University Press, 1986).

Hamilton, W. D.: "The Genetical Evolution of Social Behavior." Journal of Theoretical Biology, 12 (1964), 1-16, 17-32.

Hobbes, Thomas: Leviathan, ed. Edwin Curley (Indianapolis: Hackett, 1994; originally published in 1651).

Hume, David: An Enquiry Concerning the Principles of Morals. Reprint ed. L. A. Selby-Bigge (Oxford: Oxford University Press, 1902; originally published in 1752).

Hume, David: A Treatise of Human Nature, ed. L. A. Selby-Bigge (Oxford: Clarendon Press, 1888; originally published in 1739-40).

Lewontin, R. C.: "Survival of the Nicest?" New York Review of Books, 22 October 1998, 59-63.

McDowell, John: "Values and Secondary Qualities," Morality and Objectivity, ed. Ted Honderich (London: Routledge and Kegan Paul, 1985), 110-29.

Mill, John Stuart: Utilitarianism, ed. George Sher (Indianapolis: Hackett, 1979; originally published in 1861).

Moore, G. E.: Principia Ethica (Cambridge: Cambridge University Press, 1903).

Nagel, Thomas: The Last Word (New York: Oxford University Press, 1997).

Singer, Peter: How Are We To Live? (Amherst, NY: Prometheus Books, 1995).

Sober, Elliott, and David Sloan Wilson: Unto Others: The Evolution and Psychology of Unselfish Behavior (Cambridge, MA: Harvard University Press, 1998).

Wilson, E. O.: Sociobiology: The New Synthesis. (Cambridge, MA: Harvard University Press, 1975).