

A Biological Alternative to Moral Explanations

Joseph Millum

National Institutes of Health

Abstract

Some moral realists claim that moral facts are a species of natural fact, amenable to scientific investigation. They argue that these moral facts are needed in the best explanations of certain phenomena and that this is evidence that they are real. In this paper I present part of a biological account of the function of morality. The account allows the identification of a plausible natural kind that could play the explanatory role that a moral kind would play in naturalist realist theories. It is therefore a candidate for being the moral kind. I argue, however, that it will underdetermine the morally good, that is, identifying the kind is not sufficient to identify what is good. Hence this is not a natural *moral* kind. Its explanatory usefulness, however, means that we do not have to postulate any further (moral) facts to provide moral explanations. Hence there is no reason to believe that there are any natural moral kinds.

1. Introduction

In *The Nature of Morality*, Gilbert Harman argued that moral facts, were they to exist, would be explanatorily irrelevant.¹ In any situation where someone makes a moral judgment, the explanation of why they made the judgment need not mention any moral facts. Consequently, the moral judgments people make would be the same whether or not there were any such facts. This, according to Harman, implies that nonreductive forms of moral realism are false.

Joseph Millum is a Fellow in the Department of Bioethics at the National Institutes of Health. His research interests include foundational issues in ethics, the origin and content of parental rights and responsibilities, and the ethics of international research.

Those moral realists who accept Harman's characterization of their position, that is, those who accept that evidence for or against moral facts could be given by their role in explanations, have developed a two-part response.² First, they deny that the bare fact that moral facts *could* be omitted from explanations is sufficient to justify disbelief in them on the grounds of parsimony. If this were the case, then most of the facts cited by nonphysics sciences should be similarly rejected. The invocation of parsimony could only be effective if it were shown that moral facts do no explanatory work, that is, they are not needed in the best explanations of any type of phenomenon. Second, they present examples of phenomena that they claim are best explained by citing moral facts.

My aim is to show that this response is not sufficient to defend moral realism against the charge of explanatory impotence. I argue that even if we accept that these realists have identified phenomena that could be explained by moral facts, these phenomena are better explained by facts that are not moral facts. I present part of an evolutionary biological account of the origins of moral beliefs, sentiments, and rules. Using this account it is possible to identify a plausible natural kind that could play the explanatory role that a moral kind would play in naturalist realist theories. It is therefore a candidate for being the moral kind. I argue, however, that it will underdetermine the morally good, that is, identifying the kind is not sufficient to identify what is good. Hence this is not a natural moral kind. Its explanatory usefulness, however, means that we do not have to postulate any further (moral) facts to provide moral explanations. Hence there is no reason to believe that there are any natural moral kinds.

I focus on the naturalist realism of Richard Boyd, David O. Brink, and Peter Railton. Each holds that moral facts are a distinct natural kind (though they differ somewhat on how to characterize this kind). I take them as representative of this type of contemporary moral realism, which I call *substantive naturalism*; I therefore intend to raise a general problem with this position.³

2. Locating Substantive Naturalism

Moral realists hold, minimally, that moral judgments are truth-valuable (i.e., they attempt to make statements about matters of moral fact) and that some moral judgments are true (i.e., there are moral facts to be talked about).⁴ They thereby contrast with *noncognitivists*, who deny that moral judgments are truth valuable, instead interpreting them as expressions of some type of affective attitude, and with *error theorists*, who accept cognitivism but deny that moral judgments are ever true, because they deny that there are any moral facts.⁵

Substantive naturalism should be distinguished from various other realist positions. As a *naturalist* position it contrasts with nonnaturalist realism, according to which moral facts are of an ontologically different type from those mentioned by the sciences. These facts are *sui generis*; as Russ Schafer-Landau puts it, “they introduce an element of normativity that cannot be captured in the records of the natural sciences.”⁶ As a *substantive* position, substantive naturalism contrasts with definitional naturalism, which agrees that moral facts are natural facts but holds that an examination of the meaning of certain normative concepts could reveal which natural facts constitute moral facts.⁷ Finally, substantive naturalism is objectivist—it holds that the status of moral facts is independent of anyone’s beliefs about those facts. Consequently it should be distinguished from subjectivism and response-dependent realism.⁸ My arguments in this paper are not intended to bite against any of these alternative views.

I turn now to a more detailed exposition of substantive naturalism and, in particular, how the substantive naturalists think moral facts may figure in explanations.

3. Moral Facts and Explanations

Moral facts, if there are such, supervene on other natural facts.⁹ This means, minimally, that for any moral fact, M, there are nonmoral facts, N, such that any change in M entails a change in N.¹⁰ Consequently, if moral facts are natural facts, then evidence for their existence must come from their explanatory independence from the supervenient base, that is, our best explanations of some phenomena must make essential reference to moral facts. If moral facts are not needed in our best explanations, then we have no reason to think that there are any such facts.¹¹ This condition applies to moral terms just as it applies to other supervenient natural kind terms, such as chemical elements, genes, beliefs, and so forth; if we can replace the terms with lower-level terms without loss of explanatory power, then we should disbelieve in their existence, on the grounds of parsimony.¹²

The substantive naturalists I am considering claim that positing moral facts can help explain certain considered moral judgments and some social phenomena. In neither case do they assume that the moral facts are the sole cause of the judgments or social phenomena, since other psychological, sociological, economic, and historical causes may need to be cited in a full explanation; however, the moral facts play an ineliminable role. In the following two sections I show what part moral facts would play in such explanations.

3.1 Moral Judgments

Moral judgments are natural candidates for phenomena that may be explained using moral facts. Indeed, if moral judgments could *never* be explained this way, but moral facts were natural facts, then it seems that we could not have moral knowledge. This follows from a minimal condition on an account of knowledge: if I know that P, then my belief that P must be nonaccidentally related to the fact that P. For example, if I see a man hitting a child and form the belief “that’s wrong,” this will only count as knowledge if it was the wrongness of the action that led to my judgment, not, say, a mistaken belief that the man was tousling the child’s hair. Similarly, there needs to be some causal link between moral judgments and moral facts in order to secure reference to moral facts (assuming a causal theory of reference).¹³

More generally, in order for it to be plausible that we could acquire a correct moral theory through scientific inquiry, a significant subset of our initial moral beliefs must be at least approximately true. Otherwise, reflection on those beliefs (using, say, the method of reflective equilibrium) might not lead to increased verisimilitude but might lead us further away from the facts.¹⁴ And if there is an interplay between observation and theory that facilitates improvements in our moral theory, then this reflects a relationship between our moral judgments and the facts. This would not imply that all or even most of our present moral judgments are true, but it would mean that some of them can be (partially) explained by the moral facts.¹⁵

3.2 Social Phenomena

Social phenomena, such as social unrest in the face of institutionalized injustice, are also amenable to moral explanation.¹⁶ If we consider, for example, the social unrest in South Africa under apartheid, it is natural to explain it with reference to the injustice of the social institutions; that is, when asked why there was so much social unrest, we point out that the system was set up so that there were unjustified inequalities between different groups. Lower-level explanations in terms of economic and social conditions are possible, but they fail to explain relevant counterfactuals: the unrest would exist under different economic and social conditions, provided that they were still unjust. Brink explains:

it seems better to cite racial oppression as a cause of political instability and social protest in South Africa than the particular social, economic, and political restrictions, precisely because there would still have been racial oppression and instability and protest

under somewhat different social, economic, and legal restrictions, and the only thing this large set of alternate possible social, economic, and legal bases of oppression have in common is that they realize racial oppression (it is very unlikely that there is a natural—nonmoral—social category that corresponds to this set).¹⁷

In Section 5, I argue that there is indeed a natural nonmoral category that appropriately (for explanatory purposes) corresponds to this set and that can help explain moral judgments. This category can be identified using an evolutionary account of moral phenomena.

4. A Biological Account of Morality

The previous section showed how substantive naturalists think moral facts should figure in explanations. In this section, I sketch a biological model of the origins and function of morality, which suggests rival explanations of the same phenomena.¹⁸ The account is selective, so I can explain the details that are relevant to my argument against moral realism in a suitably small space. Consequently, I omit some important aspects of how a system of cooperation, like morality, could arise among humans. As I argue in Section 8, these omissions should not affect the conclusions I draw.

I should emphasize at the outset that the term “moral” is here being used in a descriptive sense. Thus to explain the origins of morality is to explain the origins of human capacities for moral beliefs and sentiments, as well as the general form of those beliefs and sentiments; but the claims I make about morality are not normative claims. Similarly, I shall describe certain actions, or strategies for acting, as “rational”; but by this I mean here only that they will maximize an organism’s fitness.¹⁹

4.1 Modeling the Evolution of Cooperation

The evolutionary biological account postulates that morality arose in order to facilitate mutually advantageous cooperation between members of a community in situations where each individual’s short-term interests would be better served by not cooperating. Internalized moral rules and sentiments play a role in interactions which have the form of iterated prisoner’s dilemmas.

A two-person prisoner's dilemma has the following generic payoff matrix.²⁰

Prisoner's Dilemma: Generic Payoff Matrix

	Player B		
	(A,B)	Cooperates	Defects
Player A	Cooperates	(r_1, r_2)	(s_1, p_2)
	Defects	(p_1, s_2)	(t_1, t_2)

Where $p_i > r_i > t_i > s_i$ for all i ;
and $r_i > (p_i + s_i)/2$ for all i .²¹

Example Payoff Matrix 1

	Player B		
	(A,B)	Cooperates	Defects
Player A	Cooperates	(3, 3)	(0, 5)
	Defects	(5, 0)	(1, 1)

Each participant in a prisoner's dilemma will maximize his or her payoff by defecting, whatever the other participant does. However, if both defect, then the payoff each receives is less than if they had both cooperated: rational action leads to a suboptimal outcome. In situations where the same individuals interact over and over again, this suboptimal outcome can be avoided by adopting some sort of cooperative *strategy*.²² The tit-for-tat strategy, for example, involves cooperating on the first interaction and thereafter doing whatever one's partner did on the previous interaction. A strategy of this type allows two cooperative individuals to reap the benefits of cooperation but does not leave them open to excessive exploitation by individuals who defect. I assume these ideas are familiar.²³

It is relatively straightforward to show how reciprocal cooperation might evolve in iterated two-person prisoner's dilemmas. In populations where pairs of individuals interact frequently, tit-for-tat is an evolutionarily stable strategy (ESS)²⁴ and will be more successful than defecting strategies

when there are enough cooperating individuals in the population. There are various ways in which enough cooperators could evolve to make tit-for-tat a viable strategy. One is via kin selection: because they share genes it can be advantageous to cooperate with kin when it would not be to do so with nonkin.²⁵

Prisoner's dilemmas need not occur just between pairs of individuals; they can occur in groups of any size. However, it is less straightforward to explain the evolution of cooperation between individuals in larger groups. To see why, suppose that each individual is asked to bear a certain cost, C , in cooperating (we assume the cost of defecting is 0), and this then contributes a benefit, B , to the total benefit to the group (where $B > C$). The total benefit, B_N , is then shared equally among all N members of the group. For each individual it is clearly rational to defect, but if all defect this will result in a suboptimal payoff to each compared to universal cooperation, that is, this is a prisoner's dilemma. However, as the number of cooperators increases, the gain to each from cooperating drops: for any individual the benefit they receive from cooperating, whatever others do, is $B/N - C$. This rapid drop in the benefits of cooperation means that even with modestly sized groups, cooperative strategies are unlikely to increase when rare and are very vulnerable to invasion by defecting strategies when common.²⁶

Robert Boyd and Peter Richerson propose the following solution to this problem.²⁷ They argue that conformism (the matching of individual behavior to the behavior of the majority) is independently advantageous. This is because many of the learned behaviors of the adult members of a population are likely to be adapted to local conditions.²⁸ Conformism is a force that slows or prevents the change in behavior of a subpopulation. If there are subpopulations that are cooperative (perhaps groups within the population that have evolved in the ways described above), then conformism can make them resistant to being undermined by selfish strategies. This is because conformism will mask the effect of genotypes that cause behavior that is not in conformity with the behavior of the majority. If the phenotypic expression, in this case of the defecting strategy, is prevented, then it cannot be selected for. This makes it easier for *group selection* to take place.²⁹ The effects of group selection are generally reckoned to be slight: group selection is too slow to seriously counter the effects of gene selection at the level of the individual.³⁰ In this case, however, gene selection for defecting behavior does not occur. Hence, the competition between groups can be evolutionarily significant. It is clear, moreover, that a group of cooperators is likely to do much better than a group of defectors, considering the extent to which humans are dependent on each other and

how well they are able to prosper from cooperation.³¹ Thus we may expect cooperating groups to drive noncooperating groups to extinction, either literally or by assimilating their members.³²

4.2 The Importance of Cooperation

Opportunities for reciprocal cooperation are common in many areas of human social activity, in many different social systems. Hunting and gathering may present such opportunities, as will food distribution systems. Farming requires cooperation, at least within families and potentially in larger groups, and common land may require cooperative management.³³ Personal property would not be possible if people could not rely on others not taking the property. Mutual defense (against predators and against rivals) may require extensive cooperation with opportunities for gain through defection. People's day-to-day activities in any community require that those we live with do not harm us each time they have the opportunity.³⁴ And trade, insofar as it relies on each giving what they say they will give, would not be possible if everyone looked only to their immediate gains. This list could be substantially extended. It should suffice to indicate the importance of cooperation for humans and the frequent overlap between situations where reciprocal cooperation is possible and situations that are governed by moral rules.

The considerations given so far allow us to see when and how reciprocal cooperative behavior can be fitness enhancing. Further, in situations where humans tend to engage in such cooperative behaviors, moral considerations are frequently believed to apply. This, in turn, allows us to assign an evolutionary function to the evolved mechanisms in humans that facilitate the internalization of moral rules and generate moral sentiments. These mechanisms, which I term the human *moral apparatus*, have the function of encouraging cooperation and so help to maximize the total benefits to interacting individuals.³⁵ They thus act as a referee for interactions where the long-term interests of the members of a community differ from their short-term interests. Moral sentiments motivate cooperation in an individual and their expected presence reassures others that the individual will cooperate. Where cooperative expectations are not met, censorious sentiments may arise that help prevent the breakdown of future cooperation; these may be self-directed (like guilt or shame) or other-directed (like moral outrage).³⁶ Thus moral rules and sentiments are mechanisms for using complex cooperative strategies in iterated prisoner's dilemma situations.³⁷

4.3 Mutually Advantageous Combinations

There is an aspect of the prisoner's dilemma payoff matrix whose significance is often overlooked:³⁸ the variation that is possible in the payoffs to each interactant. For each individual the factors relevant to the rational action to perform are the relationships between the possible payoffs *to that individual*—the payoffs to other individuals are irrelevant. As long as the payoff to A from defecting is greater than the payoff from cooperating (no matter what B does), it will be rational to defect in a one-shot prisoner's dilemma; and as long as the payoff to A from mutual cooperation is better than the average payoff of alternately successfully defecting and unsuccessfully cooperating, it will be rational to engage in cooperative strategies in iterated prisoner's dilemmas. This means that the payoffs to the different individuals who find themselves in a prisoner's dilemma can be very different from one another. For example:

Example Payoff Matrix 2

	Player B		
	(A,B)	Cooperates	Defects
Player A	Cooperates	(3, 15)	(0, 20)
	Defects	(5, 5)	(1, 10)

Here, despite the fact that B receives a much larger payoff than A for cooperating, it is still rational for A to follow a cooperative strategy, if the alternative is a destructive series of mutual defections. The same considerations apply *mutatis mutandis* to situations with more than two interactants. Hence in situations where cooperative behavior is an option, the payoffs can be distributed in various ways while preserving the payoff relationships that lead to cooperation. For example, where resources, such as food, are acquired through cooperative endeavor, the distribution of the resources need not be egalitarian. Resources may be distributed on the basis of need, on the basis of effort expended, on the basis of the status of the cooperators, and so forth; these substantial inequalities in distribution will not preclude cooperation to secure the resources.

Moral rules and sentiments, I suggested, have the function of leading to cooperative behavior in iterated prisoner's

dilemma situations.³⁹ We can view a moral rule as a prescription for cooperative behavior in situations falling under certain descriptive criteria. For example, the rule may prescribe the sharing of food in situations where one individual has more than he or she (loosely speaking) needs while others lack food. The considerations of the previous paragraph suggest, however, that a moral rule needs to be more precise than this. Where there are various possible distributions of resources, the rule should give guidance about that distribution, that is, moral rules will prescribe cooperation, give a description of the situations in which cooperation is prescribed, and prescribe the appropriate payoffs to the individuals involved. The fact that a particular type of situation is one in which mutually beneficial cooperation is possible does not decide which of many possible moral rules should govern the situation.⁴⁰

Some terminology will be useful here. I call a situation (or situation-type) in which mutually advantageous cooperation is possible a *cooperative situation*.⁴¹ Each of the possible allocations of payoffs that makes it rational for the individuals to engage in cooperative strategies (in the sense of rational used above) is a *mutually advantageous combination*, or MAC. MACs prescribed by the moral rules of a society are *moral MACs*,⁴² and the instantiations of these MACs are maintained to some extent by the internalizing of the moral rules by members of a community.⁴³

What decides which of these possible rules a community will come to endorse? It is probable that various factors make a difference. First, a community is likely to already have a moral system. Consequently, how well a rule proposed for a novel situation coheres with already existing rules may be a factor: a plausible analogy with a moral principle others have already internalized, for example, may make a compelling case for a rule. Second, likewise, it will make a difference how well a rule coheres with moral sentiments, such as sympathy.⁴⁴ Third, the plausibility of a rule may be affected by the grounding it is supposed to have: its purported endorsement by tradition or religious institutions, for example.⁴⁵ Finally, power will be important: all else being equal, we should expect the MAC instantiated to reflect the abilities of the interactants to inflict costs and confer benefits on one another. Thus, to some extent moral rules should reflect the power relationships between potentially cooperating individuals in a community.⁴⁶ In sum, various sociological (including economic), historical, and psychological factors will affect the moral rules a community comes to follow. This will prove relevant for the comparison between my explanations and those proposed by the substantive naturalists.

4.4 Nonparadigmatic Rules and the Function of Morality

It is perfectly possible to accept the considerations I have suggested about the advantages of cooperation and yet doubt that the moral apparatus is an adaptation with the function of facilitating cooperation in prisoner's dilemma situations. First, it might be doubted that the moral apparatus as described is an adaptation at all. Perhaps it is a spandrel—a by-product of selection for something else.⁴⁷ Further, even if it is an adaptation, moral considerations are often believed to apply to situations that do not have the form of prisoner's dilemmas. For example, self-regarding duties (e.g., develop one's talents) appear only to benefit the agent, and self-sacrificing duties (e.g., give alms to the poor) appear only to benefit others. Why should we think that it is the function of the moral apparatus to govern cooperative situations rather than these?⁴⁸

A complete defense of the claim that the moral apparatus is an adaptation would require another paper. However, its plausibility can be readily established. It is shown by the psychological and anthropological evidence that morality is a human universal,⁴⁹ that it has a physiological basis,⁵⁰ and that it follows a common developmental trajectory in children worldwide.⁵¹ Further, the biological account of morality suggests a function for the moral apparatus that explains why we can expect selection for such a thing.

We can ascertain the likely function of the moral apparatus through a process of elimination. First, since it is an adaptation, it cannot have as its function something that is normally deleterious to the fitness of the organism that possesses it. Thus it did not evolve to make people follow rules that are genetically self-sacrificing (although, of course, it might be adaptive to sincerely encourage others to such self-sacrifice). Second, in general we should not expect complex adaptations where simpler adaptations could perform the same function. Persuading people to behave in ways that are obviously in their self-interest does not generally require the complex machinery of morality: there are easier ways to spread and internalize rules of prudence. Finally, it is clear that the moral apparatus does have features that fit it for enabling mutually advantageous cooperation, that there are genetic benefits to be gained from such cooperation, and that there do not appear to be other biological mechanisms that are able to generate these benefits. Thus of the roles that morality appears to play, the only one for which it could be an adaptation, and would be needed as an adaptation, is the one proposed by the biological account.

These considerations imply that the internalizing of moral rules that prescribe MACs is likely to be the function of the moral apparatus. Thus we can expect humans to preferentially internalize these types of rule, rather than the alternatives I just considered. Naturally, this does not prevent the appropriation of the moral apparatus for the proselytizing of rules that are not MAC-prescribing, nor does it downplay the importance of the influencing factors mentioned in the last section; but it does indicate how we should expect humans to behave in the absence of such distorting factors.⁵²

This biological account of morality has been brief. Moreover, aspects of the account are liable to be challenged.⁵³ This does not matter: the purpose of this paper is not to defend a particular evolutionary account of the origins of morality, but to demonstrate a meta-ethically significant consequence of such accounts. As long as the correct account still involves adaptations whose function is to detect and take advantage of opportunities for mutually advantageous cooperation, the account here will be accurate enough to support my argument against substantive naturalist realism. In Section 8, I state more precisely the features that must be accurate for my argument to be sound. To see why, we must turn to that argument.

5. Explanation on the Biological Account

We can now see how the biological account can help to explain the phenomena that the substantive naturalist realists claim moral facts are needed to explain, such as moral judgments and social unrest. In each case, I argue, facts about MACs provide explanations analogous to those provided by positing moral facts.

5.1 Moral Judgments

In the normal case, that is, the case where morality (still descriptively speaking) is functioning as it was selected to function, a necessary condition for an accurate judgment that some action is morally good is that the action would promote cooperation for mutual benefit against the short-term interests of the interactants, that is, the action would lead to consequences fitting the payoff structure of a MAC. Further, since the capacity for moral beliefs and sentiments is an adaptation, we should expect humans to have an ability to recognize such situations. Such recognition will not, of course, be sufficient for moral judgment: the judge must also have a clear grasp of the facts of the situation, have internalized relevant moral rules, and possess (potentially motivating) moral sentiments. Nevertheless, a key factor that separates situations

where moral judgments are appropriate from those where they are not is whether the situation makes a MAC possible, and this factor helps to explain the moral judgment, as it does the presence of the moral rule in the community.

As I noted in Section 3, the naturalist realists accept that there are social and psychological conditions that are necessary for considered moral judgments to be made (and that will cause the judgments to be inaccurate where these conditions deviate from the ideal). They also think that there is a further, deeper explanatory factor corresponding to a moral fact of the matter, which may be whether the proposed action in this situation is just (Brink), in society's objective interests (Railton), or promoting human goods (Boyd). I assume the parallel with the biological account is apparent.

5.2 Social Phenomena

The biological explanation of social unrest makes use of the same facts about MACs. Consider a situation where cooperation for mutual benefit would be possible. Suppose that the rules of the society stipulate cooperation in this situation but that the payoffs that are stipulated are powerfully weighted in favor of one group rather than another so that the rules fail to instantiate a MAC. It becomes in the interests of one group to cooperate, but not in the interests of the other, that is, the benefit is not mutual. In such a situation, insofar as the members of the discriminated against group are aware that they are not benefiting, we can expect resistance to the imposition of the rule, and we can expect that the group that benefits will need to use sanctions to impose the rule. Even in those cases where there is no conscious awareness that a rule is not mutually advantageous, its deviation from a MAC can have effects: insofar as human beings are well adapted to their environment we should expect them to detect and respond negatively to situations that are not in their interests, and this need not be conscious.⁵⁴ Thus social unrest is to be expected where the rules of a society fail to be in the interests of all its members; the greater the discrepancy and the more who fail to benefit, the greater the social unrest expected.

Note, too, that this account allows an explanation of why the fact that something is, for example, unjust will explain social unrest while the particular economic and social form of the injustice varies.⁵⁵ The unrest is explained by the fact that the social institutions are not in the interests of all the members of the society, even though there are a variety of ways in which social institutions might fail to be in everyone's interests.

Of course, this is not the only way in which the moral rules in a society can affect its members. When social institutions

violate moral rules that have already been determined (i.e., members of the community have already internalized these rules), social unrest is liable to result as well. Inegalitarian institutions of property, for example, may generate unrest because they seem to violate established moral rules about fairness. In such cases, however, the most obvious explanations would seem to make reference to the beliefs of the members of the society, not directly to moral MACs (or, indeed, moral facts).

The biological account seems to do the explanatory work that the naturalist realists want moral facts to do. If an account like this is correct, then there is no need to postulate any further moral facts to explain considered moral judgments and the relevant social phenomena. This implies that either moral facts simply are facts about MACs or moral facts are unnecessary for these explanations. I now argue that the MACs underdetermine whether something is morally good, and so they cannot be moral facts.

6. Biological Facts as Moral Facts

The facts that the biological account uses have a number of similarities to those suggested by naturalist realists. In particular, they are essentially related to the interests of human beings. On the biological account those interests are (at least initially) genetic interests; but genetic interests are the distal cause of many of the interests that we normally take humans to have: interests in food, shelter, sex, the good of their families, security, reliable friends, freedom to act, and so forth. These interests correspond to many of the goods that substantive naturalists mention as possibly constitutive of moral goods.⁵⁶ In both cases the “moral” facts are held to result from human interests and the relations of individuals to them; “good” individuals will promote the pursuit of interests by other members of the community, both directly (e.g., through public works) and through noninterference (e.g., not stealing). I take the similarity between the proposed facts posited in the realists’ explanations and in my explanations as evidence that the facts about MACs that I have considered are the right sort of facts to be identical with the facts that substantive naturalists claim are moral facts. Thus I am proposing an alternative to the realist explanations, not merely replacing their explanations with explanations in terms of lower-level (subvenient) entities, as Harman appeared to do.

Unfortunately, though they do the explanatory work that moral facts were supposed to do, the biological facts are not moral facts. If we know that a particular situation is one in which mutually advantageous cooperation is possible, then we know that a moral rule may appropriately prescribe that

people should cooperate (and avoid defecting) in this situation, that is, we know that a moral MAC will be appropriate. There are, as argued above, a number of possible MACs that could be instantiated. Any of these might be judged to be moral in a society, and the fact that mutually advantageous cooperation is possible does not tell us which MAC to instantiate, that is, which MAC is right (or good). For example, there are various ways in which material resources can be distributed in a society whereby the distribution is in each person's interests. These range from egalitarian systems (such as those of hunter-gatherer societies) through to very stratified and inegalitarian systems (such as feudal societies). Each of these may instantiate moral MACs, and this fact will be (partly) explanatory of why the system is stable and believed to be just. The distributions, however, clearly rest upon different principles. Thus the relevant explanatory property is a property of the "moral" principles of a society, but it is not determinate of those principles. It therefore appears that these biological facts are not identical with moral facts.

7. Two Quick Responses

If a substantive naturalist accepted the accuracy of the biological account I have sketched, then there are two (related) ways in which she might reply to my objection. Both involve biting the bullet by identifying moral facts with facts about MACs. First, she might suggest that the indeterminacy in question is just a consequence of some vagueness in the moral facts.⁵⁷ With goodness, as with species membership, or a property like "health," there are likely to be borderline cases, where not only is it unclear whether something has the property, but it is plausible that there is no fact of the matter about whether it has the property. Thus the moral realist could suggest that the underdetermination of moral principles by the available facts simply shows that, for example, there is no clear answer to the question of where duty ends and beneficence begins. A second response would be to treat the facts about MACs as moral facts and the particular rules instantiated in a society as an expression or interpretation of those facts. Just as we can view the different funeral practices of the Greeks and the Callatians as different ways of showing respect for the dead,⁵⁸ so particular cooperative practices could just be local ways of realizing the same values.⁵⁹

I suggest that neither of these responses takes sufficient account of the extent to which MACs underdetermine moral rules. The fact that a moral rule would instantiate a MAC allows for rules that prescribe gross inequality or wide-ranging equality, retribution or forgiveness, liberal or puri-

tanical sexual practices, and so forth.⁶⁰ Moral conflicts and questions generally concern where between these extremes the correct answer lies, not whether the answer is between them. If moral facts do not give guidance about such issues then there will be few cases in which the moral facts are able to ground particular moral claims. This would make the search for moral truth seem pointless.

8. Requirements on Substantive Naturalist Theories

My argument against substantive naturalist realism reveals a general requirement that such theories must meet. It is not enough that these realists show that some additional explanatory factor must be cited in the best possible explanations of some phenomena. In addition, they must show that they have identified an explanatory factor that is of a type such that answers to questions about that factor are of a degree of specificity that allows them to count as answers to moral questions. If they are not of the right degree of specificity, this is evidence that they are not moral facts. Thus, putative natural moral facts must satisfy two criteria: they must be necessary for the best possible explanations of some phenomenon, and they must provide answers to moral questions. I have argued that substantive naturalism cannot satisfy both criteria.

These considerations also indicate which aspects of the biological model must be accurate in order for an argument of the type I have developed to be successful, that is, what must be true for the biological facts to underdetermine moral judgments, and so for it to be clear that the biological facts are not identical with moral facts. First, there must be variation in the MACs that could govern different situation-types, that is, different mutually advantageous distributions of the spoils of cooperation must be possible in at least some situations. Second, humans must have some ability to weigh the actual and possible payoffs of cooperation. This ability must be joined with a tendency to prefer MACs over non-MACs when making moral judgments and a tendency to prefer certain distributions, typically those that favor the self, relatives, or in-group members.⁶¹ The former of these two tendencies permits MACs the role in explaining moral judgments and social phenomena that I argued for in Section 5; the latter explains why different MAC-instantiating rules may be adopted by different groups to cover otherwise similar cooperative situations. Consequently, a criticism of my argument on the grounds of some empirical inadequacy in the biological account would have to show not only that some detail of my particular account is inaccurate, but that the correct account does not meet the conditions just listed. These

conditions are very weak: if the human moral apparatus is the product of evolution, then it is likely that they will be met.⁶²

9. Conclusion: The Explanatory Irrelevance of Moral Facts

I have argued that the substantive naturalists could be correct that there is a particular type of fact that is needed to do explanatory work where moral considerations are particularly important. However, this fact does not determine what is good or bad. Hence, it is not a species of moral fact. This argument undermines substantive naturalism, since it shows that it is not necessary to make reference to moral facts in order to give explanations of moral (or, indeed, nonmoral) phenomena. Such facts are not needed for our best explanations of phenomena and so are not facts that we have reason to think exist.

This conclusion is, of course, conditional on the truth of the biological account; but I have argued that it relies only on certain structural aspects of this account. If it can be shown that there exists an explanatory factor that does the work of the substantive realists' moral facts, and that it is not determinative of moral judgments, then this form of moral realism will be undermined.⁶³

Notes

¹ G. Harman, *The Nature of Morality* (New York: Oxford University Press, 1977), 3–10.

² See, for responses to Harman in particular, N. Sturgeon, "Moral Explanations," in *Morality, Reason and Truth*, ed. D. Copp and D. Zimmerman (Totowa, NJ: Rowman and Allanheld, 1985), and N. Sturgeon, "Harman on Moral Explanations of Natural Facts," *The Southern Journal of Philosophy* 24 Supplement (1986): 69–78. Harman replies in G. Harman, "Moral Explanations of Natural Facts—Can Moral Claims Be Tested Against Moral Reality?" *The Southern Journal of Philosophy* 24 Supplement (1986): 57–68. For careful examination of the issue of explanatory impotence, see G. Sayre-McCord, "Moral Theory and Explanatory Impotence," in *Essays on Moral Realism*, ed. G. Sayre-McCord (Ithaca: Cornell University Press, 1988). The brief history I provide will not do justice to the sophisticated responses developed by moral realists; it can only situate my contribution to the debate.

³ I take their views from R. Boyd, "How to be a Moral Realist," in *Moral Discourse and Practice: Some Philosophical Approaches*, ed. S. Darwall, A. Gibbard, and P. Railton (New York and Oxford: Oxford University Press, 1997); D. O. Brink, "Moral Realism and the Sceptical Arguments from Disagreement and Queerness," *Australasian Journal of Philosophy* 62, no. 2 (1984): 111–25; D. O. Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge

University Press, 1989); P. Railton, "Moral Realism," in *Moral Discourse and Practice: Some Philosophical Approaches*, ed. S. Darwall, A. Gibbard, and P. Railton (New York and Oxford: Oxford University Press, 1997). Nicholas Sturgeon is also a substantive naturalist realist.

⁴ Some definitions of realism also include the criterion that the truth values of moral statements must be independent of anyone's beliefs about their truth values. In order to include response-dependent theories, I do not include this *objectivist* criterion. I therefore follow Geoffrey Sayre-McCord's definition of realism (G. Sayre-McCord, "Introduction: The Many Moral Realisms," in *Essays on Moral Realism*, ed. G. Sayre-McCord [Ithaca: Cornell University Press, 1988], 5).

⁵ Richard Joyce defines an "error theory" as: "the position that holds that a discourse typically is used in an assertoric manner, but those assertions by and large fail to state truths." R. Joyce, *The Myth of Morality* (Cambridge and New York: Cambridge University Press, 2001), 9. He rejects the view that an error theory holds a discourse to consist of truth-valuable claims that are all false, amongst other reasons, because it may be the case that some claims are neither true nor false (see the discussion in Joyce, *Myth of Morality*, 6–9). The difference between these definitions does not make a significant difference to this taxonomy.

⁶ R. Schafer-Landau, *Moral Realism: A Defence* (Oxford: Clarendon, 2003), 4.

⁷ See M. Smith, *The Moral Problem* (Oxford: Blackwell, 1995) for a defense of one version of definitional naturalism.

⁸ See J. McDowell, "Values and Secondary Qualities," in *Morality and Objectivity: A Tribute to J. L. Mackie*, ed. T. Honderich (London and Boston: Routledge & Kegan Paul, 1985); J. McDowell, "Projection and Truth in Ethics," Lindley Lecture. Lawrence, KS: Department of Philosophy, University of Kansas, 1988; D. Wiggins, "A Sensible Subjectivism?" in *Needs, Values, Truth* (Oxford: Blackwell, 1987).

⁹ This is generally taken to be an *a priori* truth (see, e.g., Smith, *Moral Problem*, 21–22; 40–41). It is illustrated by the requirement that if someone makes different moral judgments about two situations then they should be able to justify the difference with reference to nonmoral differences between the two situations.

¹⁰ This is a very weak notion of supervenience. Stronger notions might invoke causal independence of the supervening facts from the supervened upon facts. The use of stronger notions would not make a difference to my conclusions, however: if a case cannot be made for the existence of even weakly supervening moral facts, then it will be even more implausible that there are any more robust supervenient facts.

¹¹ Harman, *Nature of Morality*; Sayre-McCord, "Explanatory Impotence."

¹² This is a shared assumption in the literature. The naturalist realists I am discussing take on the challenge of showing the explanatory relevance of moral facts and accept that a lack of explanatory relevance would have skeptical implications.

¹³ Boyd, "How to be a Moral Realist," 115–16.

¹⁴ Boyd, "How to be a Moral Realist," 124–26.

¹⁵ Brink implies that moral facts might not explain moral

judgments, since he asserts that moral facts do not have to explain any nonmoral facts (Brink, *Moral Realism*, 183). However, it seems grossly implausible—on a naturalist account—that there could be justified beliefs in moral claims without the normal explanation of the justification citing the truth (or approximate truth) of the belief (there might, of course, be Gettier-style examples where the truth was not part of the justification but these could not be the normal way that such beliefs were justified).

¹⁶ Brink, *Moral Realism*, 187–88. See also Railton on social unrest “when the interests of a group are discounted” (“Moral Realism,” 151).

¹⁷ Brink, *Moral Realism*, 195.

¹⁸ I provide these explanations in Section 5.

¹⁹ Nothing in my conclusions turns on this somewhat deviant use of rational.

²⁰ Payoffs are in terms of fitness gains or losses. For simplicity, my examples all have positive payoffs. This is not essential: only the relationship between the payoffs must be preserved. It could make a difference in cases where an individual had the option of whether to enter the prisoner’s dilemma situation.

²¹ This second condition ensures that the maximum total payoff to the participants is gained by mutual cooperation, not by alternately cooperating while the other defects and defecting while the other cooperates.

²² Where the individuals will not interact again, defection remains the only rational strategy.

²³ See R. Axelrod and W. D. Hamilton, “The Evolution of Cooperation,” *Science* 211 (1981): 1390–96 on strategies in prisoner’s dilemma games. For cooperative strategies to be successful there must exist some mechanism that can allow the detection of (past) defectors and cooperators. Minimally, this requires some consistency in behavior, the ability to recognize interactants, and the ability to remember past behavior, or functional equivalents to these conditions.

²⁴ An evolutionarily stable strategy is not vulnerable to being undermined by other strategies when it is dominant in the population. See J. Maynard-Smith, *Evolution and the Theory of Games* (Cambridge: Cambridge University Press, 1982), 10. He gives a proof that tit-for-tat is an ESS provided that there is a large enough probability of a future interaction in Appendix K (Maynard-Smith, *Theory of Games*, 202–03).

²⁵ Given that humans may be unable to ascertain exact relatedness, strategies that lead to cooperation with those likely to be kin (e.g., a strategy of cooperating with those who live in close proximity) may evolve. This would then make it possible for unrelated individuals to cooperate and thus give tit-for-tat and similar strategies the foothold needed to prosper in the population. See A. Rosenberg, “The Biological Justification of Ethics: A Best-Case Scenario,” in *Issues in Evolutionary Ethics*, ed. P. Thompson (Albany, NY: State University of New York Press, 1995), 366–71, for discussion of how interaction with relatives might allow the evolution of quite widespread cooperation.

²⁶ R. Boyd and P. J. Richerson, “The Evolution of Reciprocity in Sizable Groups,” *Journal of Theoretical Biology* 132 (1988): 337–56.

²⁷ R. Boyd and P. J. Richerson, *Culture and the Evolutionary*

Process (Chicago: University of Chicago Press, 1985), 232–35. See also J. Henrich, et. al., “Group Report: The Cultural and Genetic Evolution of Human Cooperation” in *Genetic and Cultural Evolution of Cooperation*, ed. P. Hammerstein (Cambridge, MA; London: MIT Press, 2003), 459–64.

²⁸It also has the advantage of solving various coordination problems, since it allows individuals to predict the behavior of others.

²⁹Group selection occurs when a population contains a number of subpopulations (groups who interbreed more than they breed with other members of the population) who have differing reproductive success. In order for strategies that benefit the group to be selected for, there must be some mechanism that mitigates the effect of selfish strategies that undermine the group (cooperative groups may be more successful than selfish groups, but selfish individuals do better than cooperative ones). One such mechanism is to have the groups periodically break up and reform. Another is the conformism discussed by Boyd and Richerson.

³⁰See G. C. Williams, *Adaptation and Natural Selection* (Princeton: Princeton University Press, 1966).

³¹See Section 4.2.

³²See J. Soltis, R. Boyd, and P. J. Richerson, “Can group-functional behaviours evolve by cultural group selection? An empirical test,” *Current Anthropology* 36 (1995): 437–94, on group extinctions in New Guinea as a result of warfare. Alternative solutions to the problem of group-wide cooperation have been proposed. These include considerations of the effects of punishment (see E. Fehr, and S. Gächter, “Altruistic Punishment in Humans,” *Nature* 415 (6868) [2002]: 136–40), and reputational effects (see M. Milinski, D. Semmann, and H.-J. Krambeck, “Reputation helps solve the ‘tragedy of the commons’,” *Nature* 415 (6870) [2002]: 424–26). My goal here is not to decide what the best account is, but to show that the problem of group-wide cooperation can be solved. As long as the correct account still involves adaptations whose function is to detect and take advantage of opportunities for mutually advantageous cooperation, my account in this section should be accurate enough to support the philosophical arguments of the rest of the paper.

³³Cf. the “tragedy of the commons” (G. Hardin, “The Tragedy of the Commons,” *Science* 162 (1968): 1243–48).

³⁴Cf. Hobbes’s war of all against all (T. Hobbes, *Leviathan* [London: Penguin Books, 1985], 184–85).

³⁵Although the existence of a moral rule prescribing cooperation will not, of course, guarantee that individuals will cooperate every time. The strategies that individuals pursue are likely to be adapted to local opportunities for free-riding (e.g., defecting without detection). Thus the rules that are followed may deviate from those that are prescribed. This complication does not affect my conclusions.

³⁶Cf. R. Frank, *Passions Within Reason* (New York: W. W. Norton & Co., 1988).

³⁷Brian Skyrms has recently argued that theorists considering the evolution of cooperation have placed too much emphasis on the prisoner’s dilemma (B. Skyrms, *The Stag Hunt and the Evolution of Social Structure* [Cambridge: Cambridge University Press, 2004], xii). He suggests, instead that we should be concerned with what he calls (following Rousseau) the “stag hunt,” a game in which it is rational to

cooperate if the other cooperates and to defect if the other defects (Skyrms, *Stag Hunt*, 3). Nevertheless, the criteria for the success of my argument set out in Section 8 would still be met if Skyrms were correct, though much of this biological account of morality would require modification.

³⁸ Though see A. Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, MA: Harvard University Press, 1990), 65–67, and Maynard-Smith, *Theory of Games*, 147–66.

³⁹ Note that it is not *particular* moral rules that have this function, but moral rules in general, that is, the adaptive trait is the ability to internalize local cooperative social strategies.

⁴⁰ Consider the different views on how resources should be distributed in contemporary liberal democracies. These range from redistribution for almost complete equality, through rules like Rawls's maximin principle and moderate taxation, to libertarian views that each should be allowed what he/she acquires in free market exchange, or conservative views that each should be allowed what his/her family allots them. Each of these rules can be expected to supply *some* benefits to everyone (over the absence of a rule) but some supply substantially more benefits to certain individuals.

⁴¹ Paradigmatically this is an iterated prisoner's dilemma between individuals capable of recalling previous actions by their fellow interactants and varying their responses accordingly.

⁴² The "moral" attached to "moral MAC" should, again, be read throughout as a purely descriptive term.

⁴³ Compare the "evolutionary bargaining situations" described by Allan Gibbard (*Wise Choices*, 66). What he calls a "mutually fitness-enhancing combination" corresponds to a MAC, though not all evolutionary bargaining situations have the structure of iterated prisoner's dilemmas.

⁴⁴ Consider the difficulty that Huckleberry Finn has in following the rules of his community once his sympathy for Jim has been engaged (M. Twain, *The Adventures of Huckleberry Finn* (London: The Cresset Press, 1950), 87–91).

⁴⁵ See, e.g., R. B. Brandt, *Hopi Ethics: A Theoretical Analysis* (Chicago: University of Chicago Press, 1954), 95, on the importance of appeals to tradition in Hopi justifications of ethical claims.

⁴⁶ See, e.g., N. W. Thornhill, "An evolutionary analysis of rules regulating human inbreeding and marriage," *Behavioral and Brain Sciences* 14 (1991): 247–93, on how power relationships help explain the rules governing marriage and sexual intercourse in different cultures. I take it that one way for individuals to have power over one another is to be able to inflict costs or confer benefits in a particular situation; another is their ability to do so in general. A proper account of power is beyond the scope of this paper.

⁴⁷ S. J. Gould and R. Lewontin, "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme," *Proceedings of the Royal Society of London*, B 205 (1979): 581–98.

⁴⁸ Thanks to Danielle Bromwich and Tom Hurka for pressing me on this point.

⁴⁹ G. P. Murdock, "The Common Denominator of Cultures," in *The Science of Man in the World Crisis*, ed. R. Linton. (New York: Columbia University Press, 1957), 124.

⁵⁰ M. L. Hoffman, "Is Altruism Part of Human Nature?" *Journal of Personality and Social Psychology* 40, no. 1 (1981): 121–37.

⁵¹ L. Kohlberg, *Essays on Moral Development. Volume II: The Psychology of Moral Development* (San Francisco: Harper & Row, 1984). John Snarey provides evidence that certain of the structures of justice reasoning that Kohlberg identified are universal and that these universal structures still leave substantial flexibility in the moral rules that humans internalize (J. R. Snarey, "Cross-Cultural Universality of Social-Moral Development: A Critical Review of Kohlbergian Research," *Psychological Bulletin* 97, no. 2 [1985]: 202–32).

⁵² Note, also, that I am not claiming that it is always fitness-enhancing to behave cooperatively (or to follow moral rules). While in some situations cooperation maximizes benefit, in others it is better to be selfish. This may apply even in situations governed by moral rules, if an individual is able to defect without detection.

⁵³ See note 32.

⁵⁴ Cf. Brink on "cases where the causal efficacy and explanatory power of moral facts precede their recognition" (Brink, *Moral Realism*, 189).

⁵⁵ Brink, *Moral Realism*, 195, quoted in Section 3.2.

⁵⁶ Railton defines the objective interests of an individual in terms of the desires that individual would give her nonideal self from an idealized situation (Railton, "Moral Realism," 142–43). Boyd enumerates goods that satisfy human needs including physical and psychological needs (such as friendship and intellectual expression), and the mechanisms that enable and unify those goods such as respect, democracy, education and the like. He writes:

Moral goodness is defined by this cluster of goods and the homeostatic mechanisms which unify them. Actions, policies, character traits, etc., are morally good to the extent to which they tend to foster the realization of these goods or to develop and sustain the homeostatic mechanisms upon which their unity depends. (Boyd, "How to be a Moral Realist," 122)

⁵⁷ Boyd, for example, suggests that we should expect occasional failures of bivalence (Boyd, "How to be a Moral Realist," 129), and Brink argues that, "even a moral realist can maintain that some genuine moral disputes have no uniquely correct answers. Moral ties are possible, and considerations, each of which is objectively valuable, may be incommensurable" (Brink, *Moral Realism*, 202). Nonetheless, he concedes that a realist should maintain that most moral questions have a single answer.

⁵⁸ Cf. Herodotus, *The Histories*, trans. A. de Sélincourt (Baltimore, MD: Penguin Books, 1966), 190–91.

⁵⁹ Cf. John Kekes's pluralism. Kekes distinguishes "primary values," which are common to all humans, and "secondary values," which are interpretations of primary values in a context, or new values that can be realized only within a particular cultural context (J. Kekes, *The Morality of Pluralism* [Princeton, NJ: Princeton University Press, 1993], 41–44).

⁶⁰ Cf. note 40.

⁶¹ It is highly probable that there are some biological constraints on the MACs that are chosen. For example, Alan Fiske argues that combinations of just four relational models are sufficient to explain

the structure of the rules humans use to govern their social relationships (A. P. Fiske, "The Four Elementary Forms of Sociality: Framework for a Unified Theory of Social Relations," *Psychological Review* 99 (1992): 689–723). Nonetheless, there are good reasons to think that humans are still appropriately flexible in the moral rules they are able to internalize (see, for example, J. Henrich et al., "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies," *AEA Papers and Proceedings* [May 2001]: 73–78, who provide a cross-cultural survey of the Ultimatum game, which shows some preference for fair distributions as well as variation that correlates with culture-relative notions of fairness).

⁶² The various evolutionary explanations of moral phenomena that have been suggested in the last couple of decades all agree on the explanandum—widespread mutually advantageous cooperation. They disagree on the roles played in the explanation by kin selection, direct and indirect reciprocal altruism, and cultural factors. For example, in Richerson and Boyd's most recent work dealing with the relationship between cultural and genetic evolution, they write: "We think that human social instincts very similarly [to language instincts] constrain and bias the kind of societies that we construct, with important details left to be filled in by the local cultural input." P. Richerson and R. Boyd, *Not By Genes Alone: How Culture Transformed Human Evolution* (Chicago and London: University of Chicago Press, 2005), 216.

⁶³ The opinions expressed in this paper are my own. They do not reflect any position or policy of the National Institutes of Health, U. S. Public Health Service, or Department of Health and Human Services. I would like to thank Danielle Bromwich, Philip Clark, Michael Garnett, Tom Hurka, Wayne Sumner, Paul Thompson, and audience members at a Central Division meeting of the American Philosophical Association for helpful comments on earlier drafts of this paper.