

EURAMERICA Vol. 40, No. 3 (September 2010), 529-571  
© Institute of European and American Studies, Academia Sinica  
<http://euramerica.org>

## The Prospects for Evolutionary Ethics Today

Neil Levy

Florey Neuroscience Institutes, University of Melbourne  
Level 2, Alan Gilbert Building, 161 Barry Street, Carlton South, Victoria, 3053  
Australia  
E-mail: [neil.levy@philosophy.ox.ac.uk](mailto:neil.levy@philosophy.ox.ac.uk)

### Abstract

One reason for the widespread resistance to evolutionary accounts of the *origins* of humanity is the fear that they undermine morality: if morality is based on nothing more than evolved dispositions, it would be shown to be illusory, many people suspect. This view is shared by some philosophers who take their work on the evolutionary origins of morality to undermine moral realism. If they are right, we are faced with an unpalatable choice: to reject morality on scientific grounds, or to reject our best-confirmed scientific explanation of our origins in order to save morality. Fortunately, as I show, we have no reason to accept the deflationary claims of some evolutionary ethicists: morality, as we ordinarily understand it, is fully compatible with evolution.

**Key Words:** evolution, ethics, meta-ethics, Spencer, Huxley

---

Invited article, Received March 9, 2010  
Proofreaders: Hsueh-mei Chen, Yi-Hsin Lai, Ying-bei Wang

Ever since the sociobiological revolution of the 1970s, we have witnessed an ever-growing interest in evolutionary explanations of morality. More and more philosophers accept that evolution has something to tell about, at very least, the *origins* of morality. But many are loathe to grant it a more substantial role. In part, I suspect, this hesitancy is the product of an obscure feeling that a more thoroughly evolutionary ethics would have deflationary, or even eliminativist implications. If morality can be significantly explained in evolutionary terms, it would be explained *away*, revealed as illusory. Moreover, this feeling is not confined to those who reject robustly evolutionary explanations of morality. Many of those who champion evolutionary ethics also take it to have deflationary implications. The best explanation for how and why we come to make moral judgments shows them to be unjustified, they claim.

We need urgently to articulate the fears of the opponents of a thorough-going evolutionary ethics, and to examine the deflationary arguments of its supporters. For the evolutionary explanation of morality is not merely by far the best confirmed hypothesis we have concerning the origins of our moral sense, it also seems to promise to throw light on its content as well. If the evolutionary explanation has normative, motivational and meta-ethical implications, we must discover what they are. The naturalizing tendency in philosophy has taken hold across most of its sub-disciplines, and for good reason. Its hold is growing in ethics as well. Our meta-ethics had better come to grips with this naturalizing current, if it is not to become an irrelevancy.

In this paper, I will review much of the recent work in evolutionary ethics, with these meta-ethical questions foremost in mind. Before we turn to this recent work, however, some background is necessary, which will illustrate the characteristic debates and methods in the field. Section One is largely historical, focusing on the debates over the moral implications of natural selection which, already in Darwin's day, divided evolutionists against each other. These debates are of more than narrowly

historical interest, since, as we shall see, they delineate positions which continue to mark the major polarities of the contemporary debate. Section Two briefly sketches the more sophisticated models of the evolution of morality developed by biologists and philosophers over the past three decades. It is the explanatory power of these models that constitutes the best evidence that morality is the product of natural selection, and it is therefore their implications which matter in this debate. Section Three turns to the main question of this article, the examination of the meta-ethical implications of the claim that morality is a product of natural selection. In particular, I shall be concerned with the extent to which it threatens to undermine, or to narrow, the concept of morality shared by most philosophers and by common-sense. Does evolution sound the death knell of (what we have hitherto regarded as) morality? As we shall see, there are good reasons to think that the robust evolutionary story told by many evolutionary ethicists does indeed imply that nothing answers to this concept of morality. We shall have to choose, between this evolutionary hypothesis and morality, as we hitherto understood it. It is far from clear, however, that a respect for science and an acknowledge of the truth of Darwinian evolution requires us to accept the full evolutionary hypothesis urged by the Darwinian fundamentalists. It may be that we can naturalize ethics without emptying it of most of its substance.

## I. The Rise and Fall and Rise of Evolutionary Ethics

The naturalistic tendency in moral philosopher has a very long history. Aristotle, in one way, and Hume, in another, both sought to give their ethical theories foundations in theories of human nature (indeed, Hume continues to serve as a model and inspiration for more recent naturalizers). Here, however, we are concerned more specifically with the *evolutionary* naturalizers:

philosophers (and sometimes biologists) who attempt to explain, or explain away, ethics by reference to Darwinian and neo-Darwinian natural selection.<sup>1</sup>

So understood, evolutionary ethics begins with Darwin himself. In both his major works, *The Origin of Species* and *The Descent of Man* (especially in the latter), Darwin makes scattered remarks of direct relevance to the claim that morality has an evolutionary origin. Most importantly, in *Descent* Darwin notices the problem of altruism, which has plagued evolutionary ethicists ever since, and offers the outline of a solution to it. The entire history of evolutionary ethics could be written in terms of this problem, and the changing fortunes of Darwin's solution. However, perhaps because Darwin had an aversion to unnecessary controversy, he did not play a conspicuous role in the debate over evolutionary ethics. Instead, this debate features most prominently two of his most able champions. Between them, Herbert Spencer and Thomas Henry Huxley set the tone for much of the debate which was to dominate evolutionary approaches to ethics for the next century. Though their works are seldom read today, evolutionary ethicists continue to work in their shadow, whether they realize it or not.

Evolutionary ethicists have been primarily concerned with two questions:

1. Can evolution provide us with knowledge of the ethical principles which ought to guide our behavior? This is the *normative* evolutionary question.

---

<sup>1</sup> A note on terminology. *Neo-Darwinism* is the product of the synthesis of Darwinian evolution with Mendelian genetics (which provided Darwinism with the mechanism of heredity it hitherto lacked). Though there are some signs of cracks in the neo-Darwinian façade today, the consensus version of evolution is still overwhelmingly neo-Darwinian. In this essay, I use terms like "evolution," "Darwinism," and "natural selection" to refer to this consensus view. According to it, evolution occurs through the process of random variation, differential reproduction and inheritance. There are (postulated) evolutionary mechanisms besides these (such as Lamarckian inheritance of acquired characteristics) but they are ignored here.

2. Might understanding the (putative) evolutionary origins of morality help us comprehend the structure of ethics, and to settle debates over the meaning and reference of ethical terms? This is the *meta-ethical* evolutionary question.

As we shall see, the relative importance accorded these questions in the ongoing debate has altered over time. Early evolutionary ethicists were more concerned with the first; today, it is the second that receives most attention. But both were already clearly articulated by Huxley and Spencer, and the answers they gave laid down the pattern for their successors.

Huxley and Spencer were both enthusiastic evolutionists, but they diverged radically in the extent to which each believed that ethics could be given a biological foundation. Huxley argued that ethics was an *autonomous* institution, independent of, indeed opposed to, mere biological nature. It is probably true, Huxley contends, that our moral sentiments are the product of natural selection. But, by the same token, so are our *immoral* sentiments. From a purely evolutionary point of view, there is no reason to prefer one set of sentiments to the other. They are each merely a set of motives which drives behavior that might be expected to have survival value. Thus, we cannot derive ethical principles from natural selection itself:

The thief and the murderer follow nature just as much as the philanthropist. Cosmic evolution may teach us how the good and the evil tendencies of man may have come about; but, in itself, it is incompetent to furnish any better reason why what we call good is preferable to what we call evil than we had before. Some day, I doubt not, we shall arrive at an understanding of the evolution of the aesthetic faculty; but all the understanding in the world will neither increase nor diminish the force of the intuition that this is beautiful and that is ugly. (Huxley, 1989: 138)

Thus, Huxley answers the meta-ethical question largely in the negative: at least by itself, evolution cannot tell us much about the meaning or reference of moral terms. “Good” does not mean

“adaptive,” or anything of that kind, for too many different kinds of actions and traits of character fit this description. Meta-ethics is relatively independent of biology.

What of the normative question? Here Huxley is unequivocal. If we were to follow nature’s counsels, then we would glorify death, suffering and selfishness. Natural selection systematically rewards violence and greed. If it can serve as a model for us, it is only negatively, by showing us what *not* to do:

Let us understand, once for all, that the ethical progress of society depends, not on imitating the cosmic process, still less in running away from it, but in combating it. (Huxley, 1989: 141)

Far from modelling morality on the process of natural selection, Huxley argued that we ought to oppose evolutionary processes in the name of morality.<sup>2</sup>

Spencer’s influential views could not be more different. It was Spencer, and not Darwin, who coined the phrase that in many minds still encapsulates evolution, “the survival of the fittest.” For him, evolution is an essentially *progressive* force, which systematically selects the best representatives of each species. If it is allowed to work without interference, therefore, natural selection improves the quality of species. We can therefore identify the process and the product of evolution with the good.

Spencer gives positive, and interrelated, answers to both the normative and the meta-ethical questions. For him, “good” just *means* “highly evolved”; the study of evolution shows that

---

<sup>2</sup> I take it, however, that Huxley does not mean that “the cosmic process” gives us an infallible—negative—guide to morality. If that were the case, then we would have an evolutionary analysis of goodness at hand: “good” refers to all and only those actions and processes which are disfavored by evolution. Huxley argues that both our moral and our immoral dispositions are the product of evolution. Accordingly, I take his point to be that evolution cannot offer us any way of distinguishing between them.

the conduct to which we apply the name good, is the relatively more evolved conduct; and that bad is the name we apply to conduct which is relatively less evolved. (Spencer, 1883: 25)

If “good” means “highly evolved,” then it is our moral obligation, in pursuing the good, to assist the process of evolution. Thus, Spencer’s meta-ethical views have direct implications for his answer to the normative question. We ought to allow natural selection to do its work. Spencer therefore opposes a short-term, sentimental, devotion to the welfare of the poor to a hard-headed, but ultimately also more ethical, concern to see that those who win out in the struggle for existence are those who are, genuinely, fit. He advises against organized charity, for instance, since it will tend to divorce fitness from reproductive success, while supporting private charities which aim to provide the deserving poor with opportunities they might otherwise lack. We should intervene in social processes only to undo the effects of past interventions, or to compensate for sheer bad luck, Spencer argued. But we ought to resist the temptation to ameliorate the suffering of the undeserving, of those genetically destined to fall by the wayside.

Spencer’s views, and similar positions, were massively influential in the early decades of the twentieth century. The Social Darwinism he advocated, and the closely related eugenics movement, had thousands of adherents, from all parts of the political spectrum. Not just conservatives, but socialists and feminists, as well as the majority of mainstream scientists, supported these programs, in whole or in part. They were politically successful, especially in Northern Europe, and in the United States. Eugenic policies, aiming at discouraging the immigration or birth of supposedly inferior individuals, became law in twenty-nine American states (Kevles, 1995). But it was the far more sinister eugenic policies of Nazi Germany which finally discredited the movement. After the war, eugenics went into a rapid decline. Its erstwhile intellectual leaders quickly distanced

themselves from it or were marginalized, the legislation which had put its policies into practice was repealed. It was not, it seems, because eugenics had been shown to be false, because the theories which it had elaborated had been refuted, that it suffered this fate. It was because of its association with crimes of an unprecedented enormity that it lost favor.

Because of the link between views like Spencer's and eugenics, the vogue for finding normative implications in evolution largely passed. The belief that some individuals were naturally fitter than others came to be associated with racism and with a morally retrograde and scientifically unsound elitism. On more purely intellectual grounds, Moore's "open question argument"—specifically formulated with Spencer in mind—seemed to defeat all attempts to identify goodness with being highly evolved (Moore, 1903). It would, Moore argued, always be an open question whether any identification of moral terms with natural properties was correct. The direct route, from meta-ethics to normative principles, was apparently blocked.

If Spencer won the debate between him and Huxley during their lifetimes, in the sense that his views were far more influential at the time, then perhaps we can say that Huxley emerged the eventual victor. After the war, most moral philosophers sided with him in agreeing that ethics was an autonomous domain, even if our capacity to engage in it was itself a product of evolution. Even when a more specifically evolutionary ethics began to revive in the 1970s and 1980s, spurred on by the success of sociobiology, the philosophers who engaged in it steered, for the most part, away from normative and motivational questions. Instead, they concerned themselves largely with questions that Huxley deemed appropriate, in particular with the question whether the fundamental bases of morality—whatever they turn out to be—might have a direct evolutionary origin.<sup>3</sup>

---

<sup>3</sup> Of course, given that natural selection is true, all human abilities and characteristics are the product, in some sense, of evolution. By asking whether

Gradually, however, their ambition grew. Indeed, the naturalizing impulse cannot appropriately be confined to genetic (in the historical, not Mendelian, sense) questions. If morality is the direct product of evolution, then we can explain its content in adaptive terms. Answering the genetic question in the affirmative will therefore have far-reaching implications for our account of moral motivation, for our meta-ethics, and for normative questions as well. The debate between Spencer and Huxley is not over: if morality can be directly tied to evolution (as Spencer claimed), it is a very different kind of phenomenon than the ethics which might emerge from the processes of cultural and intellectual elaboration on a biological base, the kind of morality that Huxley defended. In particular, as we shall see, a neo-Spencerian morality seems far more likely to be subjectivist, non-realist and restricted than a Huxleyan morality. It is therefore an urgent task to assess the plausibility of neo-Spencerian views.

Before we undertake this task, however, let us briefly sketch the kind of accounts which both sides in the continuing debate now rely upon as the central pillar of their claim that morality is—in some sense—a product of natural selection.

## II. The Problem of Altruism and the Evolution of Morality

Altruism is a puzzle for evolutionary ethicists, for fairly obvious reasons. Evolution systematically favors *phenotypic* traits—the bodily morphology and behavior of organisms—which are adaptive. A phenotypic trait is adaptive just in case it increases the fitness of the organism; fitness, in turn, we define in term of the reproductive success of the organism. Fit phenotypic traits are thus reproduction-relevant: they are characteristics that enable the

---

morality is a *direct* product of evolution, I ask whether its evolutionary history can throw light on its nature and function, or whether we are better off seeking to understand it in, say, exclusively cultural terms.

organism successfully to compete for reproduction-relevant resources, with conspecifics and with organisms from other species. Just which resources will be reproduction-relevant will vary from species to species, but they will always include the means of survival (food and shelter) and of attracting high-quality mates (assuming sexual reproduction). Thus, whether they are aware of it or not, all organisms are apparently engaged in a struggle for existence with every other.

Darwinian evolution can be an accurate description of life only if phenotypic traits are heritable. Offspring must tend to resemble their parents. Given that this is the case, traits that are fit will tend to spread in a population, since a fit trait enables its possessor to have more viable offspring. The stage is now set for the problem of altruism. A behavior is altruistic just in case it benefits others (whether they are members of the same species, or of other species) at the expense of the organism whose behavior it is.<sup>4</sup> It follows from this that someone who acts altruistically reduces her own fitness, while boosting that of the recipient. Altruists will therefore, on average, have fewer descendants than non-altruists. We ought to expect altruism to be a casualty of the struggle for existence.

But altruism does seem to exist, among human beings and elsewhere in the natural world: people and other animals sometimes seem willing to put themselves out for others, without hope of recompense. This is the problem of altruism: given that we are the products of evolution, how did we come to be altruistic, to whatever extent? E. O. Wilson called this “the central theoretical problem of sociobiology” (Wilson, 1975: 3). This problem is of first importance to evolutionary ethics, since it seems part of the

---

<sup>4</sup> This is a stipulative and biologized definition of a concept that has its place in ordinary moral discourse, but its main ingredients are defensible. We would not call an action altruistic if it did not (aim to) benefit someone other than the actor. If the actor believed that she would benefit from the act more than from alternatives, then it is not altruistic.

definition of morality that it will require us sometimes to give at least *some* weight to the interests of others independently of our own interests: that is, morality requires altruism of us.

Darwin's solution to the problem of altruism remained the most influential up until the 1960s. It goes by the name of **group selection**. Though it is true that individuals who behave altruistically have lower fitness than other individuals who do not, internally altruistic *groups* can outcompete internally selfish groups. That is, a group composed of individuals who have a disposition to aid one another might be expected to do better than a group composed of only selfish individuals:

although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an advancement in the standard of morality and an increase in the number of well-endowed men will certainly give an immense advantage to one tribe over another. There can be no doubt that a tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to give aid to each other and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection. (Darwin, 1871: 166)

Thus, despite the fact that altruism is individually less fit than selfishness, selfish groups would go extinct, and altruism would spread.

Group selectionist explanations were routinely invoked by biologists from Darwin's day till the 1960s. But in that decade it was subjected to seemingly devastating criticism by a new generation of evolutionary biologists, especially John Maynard Smith (1964) and George C. Williams (1966). These biologists were more mathematically sophisticated than their predecessors, and utilized precise models to test group selectionist hypotheses. They found that group selection was vulnerable to *subversion from within*. On the assumption that altruistic acts benefit recipients

more than they cost donors (so that altruism is a net benefit to the group), groups composed largely of altruists will indeed grow faster than those composed largely or exclusively of selfish individuals (just as Darwin predicted). But since selfish individuals are fitter than altruists *within* each group, the proportion of altruists in each group declines. Though (largely) altruistic groups might drive selfish groups to extinction, they are destined to become selfish themselves. So long as selfish organisms can arise within altruistic groups, whether by mutation or by immigration (and over the kind of times-spans with which we are concerned, such events are very likely), altruism will eventually be driven to extinction.

We should not conclude from this that group selection is impossible. Under the right conditions, it can certainly be a powerful force. If the proportion of altruists within groups necessarily declines, then group selection requires that groups do not persist long enough for this factor to eliminate altruism. If altruistic groups break up and reform, or establish colonies, *and* the successor groups have a higher proportion of altruists than the mother group, altruism can persist and even increase across the global population, so long as the formation of new altruistic groups occurs rapidly enough rate to outrun the effects of subversion from within. In the 1970s, most biologists believed that these conditions would be met with so rarely that group selection could not be a powerful force in evolution, and was therefore extremely unlikely to be the source of altruism. More recently, group selectionism has made something of a comeback (Sober & Wilson, 1998; Wilson & Sober, 1994). In important part, however, this comeback has consisted of reinterpreting supposedly individual selection processes as disguised instances of group selection. Though the biologists and philosophers who press this viewpoint have powerful arguments, we shall continue to treat these processes as alternatives to group selection, if only because the mechanisms at work in them are distinctive.

These supposed alternatives to group selection are *kin*

*altruism* and *reciprocal altruism*. Kin altruism is one of the triumphs of the gene selectionist perspective upon evolution. From this perspective, adaptations are for the ultimate benefit of the genes, not the individuals they “build.” If this is the case, however, then “altruistic” behavior by an organism can, under the right conditions, be a kind of genetic selfishness—in biological terms, they boost *inclusive fitness* (fitness calculated from the genetic viewpoint). Other things being equal, sexually reproducing organisms share half their genes with full siblings, and with their offspring, one quarter of their genes with nephews and nieces and grandchildren, and so on. Thus, by aiding a close relative, organisms can benefit the genes she shares with it. Genes for kin altruism can therefore spread in a population. Evolutionary biologists have been able to use kin selection to explain the apparent altruism of the social insects. As a result of their unique biology, worker bees and ants share three quarters of their genes with one another, while queens share only half their genes with their offspring. Workers therefore do better, from a genetic perspective, by aiding the queen to create more siblings of theirs than they would by having offspring of their own (Hamilton, 1972).

Kin altruism is restricted to relatives, but reciprocal altruism can cross almost any barriers, including within its ambit even members of other species. Reciprocal altruism, roughly, refers to the benefits given to a recipient in exchange for, or in the expectation of, a return (Trivers, 1971). For this reason, some thinkers argue that we ought not to consider it a species of altruism at all (Sesardic, 1995). But this is mistaken, at least from the point of view of the evolutionary origins of morality. Our concern is with how we came to possess moral dispositions and concepts, not with the moral character of the evolutionary process itself. Kin altruism deserves the name because it might lead to the development of, and be driven by, genuine sympathy for others, despite the fact that at a genetic level it is selfish. So too reciprocal altruism might lead to altruistic feelings, despite its more familiar

brand of (ultimate) selfishness.

Indeed, this is precisely what game-theoretic modelling of reciprocal altruism seems to show. Representing interactions between potential cooperators as a prisoner's dilemma brings out the point nicely. In such situations, both (or all) participants do better by cooperating than they would by mutual defection, but they are wary of cooperation because the worst possible outcome for them occurs when they cooperate with a defector. In a prisoner's dilemma, it is therefore to each party's advantage if it can convincingly demonstrate its willingness to cooperate on condition that the other does likewise. But if the interactors are merely rational calculators of utility, they cannot convince one another of their disposition to cooperate once doing so is no longer in their interests—that is, once the other party has cooperated. So it is to each interactor's advantage to be able to show that it acts upon motives other than the calculation of utility. One way out of this dilemma might consist in demonstrating that the organism calculates long-term utility, but there are at least two problems with this solution. It may be too costly to implement, from a biological point of view, since the machinery for calculating long term pay offs is complex, and it may be insufficient, in any case, since organisms will be less likely to interact with others when they know that these others might defect as soon as the payoffs from doing so cross a certain threshold.<sup>5</sup> Far better, then, for the organism to adopt the solution of manifesting the disposition to cooperate, on condition the other does, no matter what the payoffs. The disposition to altruism, now considered as a psychological state or a motive upon which organisms act, will, under the right conditions, boost inclusive fitness (Frank, 1988). So it is not true

---

<sup>5</sup> Moreover, it can be to an organism's advantage to possess dispositions to action, in certain circumstances, even though it would not be to its advantage actually to act upon those dispositions. If everyone knows that I shall stop at nothing to avenge a petty slight, then I may not be slighted at all, and that is to my advantage, though if I were slighted, and reacted in the threatened manner, I may pay a high price. For many examples of this kind, see Frank (1988).

that altruism cannot emerge from reciprocal exchanges.

A great deal of subtle and important work by evolutionary biologists (Dawkins, 1976; Smith, 1982), game theorists (Skyrms, 1996; Vanderschraaf, 1999) and philosophers (Kitcher, 1993, 1998; Sober, 1994; Sober & Wilson, 1998) has gone into elaborating these explanations for how altruism might emerge from the mechanisms of natural selection, even though those mechanisms ultimately reward “selfish” behavior. But these different theorists often have different targets in mind, when they seek to explain altruism. To avoid confusion, we need to adopt an important distinction Sesardic makes, between *psychological* altruism (altruism<sub>p</sub>) and *evolutionary* altruism (altruism<sub>e</sub>). An organism acts altruistically<sub>p</sub> if it acts with the intention of benefiting others at some cost to itself, whereas it acts altruistically<sub>e</sub> if it actually boosts the fitness of others relative to itself. Separating these kinds of altruism is essential if we are to understand just what claims different theorists are making.

Thus kin altruism is primarily an explanation of altruism<sub>e</sub>. It seeks to explain how helping behavior can be selected for, via the notion of inclusive fitness. It can be agnostic on altruism<sub>p</sub>, or downright sceptical concerning its existence. Certainly some theorists have tried to interpret kin altruism as consistent with psychological egoism. If organisms typically act to aid close relatives—say, their offspring—to relieve feelings of distress of their own, then kin selection may be psychologically egoistic. But kin selection might also be the first step in a two stage argument for altruism<sub>p</sub>. This argument comes in two forms, defending a restricted and a general disposition to altruism<sub>p</sub>. On the first, altruistic dispositions toward close kin are held to be the product of kin selection. This view is most convincingly articulated by Elliott Sober, in work on his own and with David Sloan Wilson (Sober, 1994; Sober & Wilson, 1998, 2000). Essentially, Sober and Wilson argue that kin selection mechanisms will be driven by concern for the welfare of kin, rather than by egoistic desires because such concern is a more a direct solution to the design

problem under consideration, and therefore more reliable (on the plausible assumption that there are cases in which purely hedonistic desires will not motivate the organism to act in the ways which maximize its inclusive fitness). Thus, kin selection will probably result in the formation of altruistic<sub>p</sub> dispositions and desires. But these desires will be restricted, in the sense that they will have as their target only (close) kin.

Thus kin selection might explain how we come to have altruistic desires directed at our kin. But this is far from an explanation of the kind of altruism necessary for morality. The extent to which morality requires us to treat all persons equally, regardless of ties of blood or affection, is a matter of ongoing debate, of course, but there can be no doubt that a morality worthy of the name requires *some* altruistic concern for those beyond the circle of family. We must be willing, at very least, to sacrifice our trivial interests for their most important ones. Might kin selection explain this more general altruism<sub>p</sub>? Alexander Rosenberg (2000) speculates (I choose that term advisedly, to indicate his own lack of commitment to the hypothesis) that it might. If human beings have lived in relatively small groups for most of our evolutionary history (as most anthropologists believe) then we might have been selected for relatively indiscriminate altruism<sub>p</sub>. Mechanisms which would allow us to distinguish relatives from non-relatives would have a cost, a cost which would not be worth paying if we lived in bands which consisted almost entirely of kin and encounters with strangers were relatively rare. However, given enough time natural selection would evolve more discriminating mechanisms, Rosenberg suggests (He obviously believes that humans have been around long enough for such indiscriminate altruism to have been eliminated.). Certainly, it does not seem uncontroversial to say that we tend to have stronger altruistic dispositions toward kin than toward strangers, which suggests that kin selection has not been implemented by way of indiscriminate altruism<sub>p</sub> in human beings. Perhaps, however, we have evolved a disposition to aid those who exhibit some characteristic which (in the environment of

evolutionary adaptation) would have been a reliable marker for kinship. Palmer and Palmer (2002) suggest that accent and dialect might play this role. If altruistic dispositions are triggered by *any* similarity markers, than altruism<sub>p</sub> might be promoted by any information that allows us to empathize with others. But if kin altruism is triggered by specific markers—whether of kinship, or of some other property that was, in the environment of evolutionary adaptation, reliably associated with kinship—then its range will of necessity be restricted. In that case, we shall have to turn to other mechanisms to explain the origins of our apparent tendency to possess (at least some) altruistic<sub>p</sub> tendencies toward almost all human beings.

Indeed, our altruistic<sub>p</sub> dispositions even extend beyond the bounds of our own species. This may seem mysterious from an evolutionary perspective, but it is not. Since reciprocal altruism is based on exchange, we can engage in it with any kind of organism so long as we are able to benefit one another. We have seen how reciprocal altruism might require the development of a cooperative disposition. We can now express this by saying that, though it is not in any obvious sense altruistic<sub>c</sub>, it may nevertheless be the condition for the development of altruism<sub>p</sub>, and this altruism<sub>p</sub> may, in turn, motivate altruistic<sub>c</sub> acts.

We now have some idea of how the dispositions which underlie morality might have come about. We are finally able to turn to our main subject: just what kind of morality might we expect to be a product of natural selection? Do some moral theories look more or less plausible in the light of the evolutionary story we have briefly sketched? In particular, does moral realism look less plausible in its light?

### III. Evolutionary Meta-ethics

In what ways might the kinds of stories sketched by evolutionary biologists, game theorists and philosophers have the

kinds of deflationary or eliminativist implications that some philosophers fear—or welcome—in them? The answer to this question will naturally depend on what we believe morality to consist in (that is, what properties an accurate analysis of the concept would impute to it; what morality is, or would be, if there was such a thing), such that evolution could threaten it.

We might best approach the question by asking what propositions a full and complete analysis of our concept of morality would contain, which might plausibly be threatened by the evolutionary hypothesis. There are at least five relevant possibilities:

(1) Morality might commit us to the existence of Platonic moral facts, which are ontologically independent of human beings or other rational beings.

(2) Morality might commit us to the existence of moral facts which, while not ontologically independent of the existence of any rational beings, are binding on all such beings, and which therefore cannot vary across space or time.

(3) Morality might commit us to the existence of objective properties, the truth conditions of which do not essentially contain references to the subjective states of the beings upon whom they are binding.

(4) Morality might commit us to the existence of moral facts which are such that their existence ensures that it is rational for us to behave morally, in general or (more strongly) on each particular occasion.

(5) Morality might require us to give some weight to the interests of all parties affected by our acts, regardless of their relationship to us.

Some of these claims are, *prima facie* at least, much more plausible than others. Thus, how threatening to moral realism an evolutionary explanation of morality will be depends upon which—if any—of these claims it is taken to undermine.

## A. Evolutionary Expressivism

Evolutionary ethicists often take their hypotheses to undermine (3). In fact, they claim, the subjective states of the person making a moral judgment figure in its truth-conditions. The thought underlying this view is seldom explicitly developed, but the idea seems to be something like this: if morality is an adaptation, then we will very likely discover that its essential building blocks did not spring into existence with *homo sapiens*, but exist in other species as well, especially those closely related to us (Flack & de Waal, 2000). Indeed, the evolutionary hypotheses for the development of morality trace the development of the dispositions and behaviors thought to underlie it in social insects and bats, fish and monkeys (Ridley, 1996; Wilkinson, 1990). Clearly, however, these organisms are not moved to act by truth-assessable representations of the world. Instead, their proto-moral behavior must be driven by instinct or by feeling. It is this proto-morality which human beings inherit from simpler creatures, and which we go on to elaborate into complex intellectual systems. But no matter how intricate a superstructure we build upon these inherited foundations, it remains the case that morality is essentially subjective. We are motivated to act as we do by feelings, not by beliefs. To think otherwise is to insert a gap in nature, a sudden leap or saltation (to use Steven Jay Gould's useful term), where in fact there is none. It is to fail to see that we are continuous with the rest of the animal kingdom, in our morality as much as anywhere else (McShea & McShea, 1999; Waller, 1996, 1997).

Thus, evolutionary expressivists conclude, moral judgments do not really express about facts about the world outside us at all. Instead, they express our feelings, our evolved sentiments. Evolution gives crucial support to expressivism. Something very like this view is defended by Michael Ruse (1998) and Bruce N. Waller (1996, 1997).

Clearly, evolutionary expressivism has *normative* implications.

It implies that claim (5) above is false. If our moral judgments express our evolved sentiments, then the content of those judgments is limited by the range of sentiments which we have evolved to feel. If we believe, further, as Ruse and Waller both do, that evolution has selected for altruistic<sub>p</sub> dispositions only with regard to close kin and perhaps others who bear markers reliably associated with close kin in the environment of evolutionary adaptation, then we shall conclude that the range of true moral judgments of which we are capable is much narrower than it is usually taken to be. For Ruse, for instance, it is a mistake to believe that morality requires us to give much weight to the needs of those who are distant from us. Since morality is “rooted in our feelings” (Ruse, 1998: 241), but we are likely to feel much more strongly for kin than for the distant needy, morality cannot require impartiality or indiscriminate altruism of us.

However, it is a mistake to think, as Ruse and Waller sometimes seem to, that the only alternative to accepting the normative and meta-ethical positions for which they argue is to reject a substantially evolutionary explanation of morality. We can coherently deny either that (a) evolution can be expected to give rise to altruism<sub>p</sub> that is essentially limited to kin, or (b) that because emotions played a crucial role in the evolutionary history of morality, such emotions must figure in the truth-conditions of moral judgments today, or (c), on the assumption that evolution gives rise to relatively restricted altruism<sub>p</sub>, it follows that we have correlatively restricted moral obligations. Claim (a) is denied by several philosophers (e.g. Kitcher, 1998), but evaluating it would take us too far afield, into the realm of game theory. Here we shall concentrate on claims (b) and (c).

Claims (b) and (c) together entail the Spencerian claim: that morality is identical with our evolved dispositions. Claim (b) holds that the emotions which figure in the evolutionary history of morality continue to figure in its contemporary truth-conditions, and claim (c) holds that our obligations extend just as far as do these sentiments. Since the evolutionary models of morality have it

arising out of the interactions of relatively unsophisticated organisms, the implicit claim is that human rationality adds nothing significant to morality.

But why think this? The most powerful argument in its favor seems to be a redundancy argument. Proto-morality—the core of morality we share with vampire bats, cleaner fish and the other primates—is largely or entirely a subjective phenomenon, driven by instinct and desire. This fact makes the objectivist hypothesis redundant. Why postulate moral beliefs, when it is clear that the subjective core of morality is sufficient to explain moral behavior?

Indeed, we might read the evolutionary expressivist as offering crucial support to a view which has attracted a great deal of attention in recent meta-ethics: the claim that moral considerations are explanatorily irrelevant. In Harman's (1977) original version of this claim, the moral properties of actions, events or characters were held to be irrelevant to the judgments we were disposed to make concerning them. Given our dispositions, we would make the same judgments regardless of their truth. This line of argument invites the response (roughly the one advanced by Sturgeon [1985]) that we have been given no reason to think that these dispositions do not track real properties, and without such an independent argument, no reason to reject their deliverances. From this perspective, we can view the evolutionary considerations as filling the gap Sturgeon sees in Harman's argument. By providing an explanation of our moral dispositions, which shows how we (might possibly) have come by them *for reasons that have nothing to do with morality*, it casts doubt on their reliability as trackers of truth. Indeed, the evolutionary history which gave rise to our moral sentiments systematically favors (what we call) selfishness. What more evidence do we need to be convinced that they are unlikely to track *real* properties of the world that are *really* moral?

Something rather like this reinforced redundancy argument is advanced by Richard Joyce (2001). Joyce asks us to compare ourselves to John, who is certifiably paranoid. John believes that Sally is "out to get him." Now, it is possible that Sally is really out

to get John, but knowing that John is paranoid leads us to think that his claim is unlikely to be true. John would judge that Sally is out to get him, no matter how she behaved. Similarly, Joyce claims, since we have been naturally selected to think that certain actions and events are right or wrong, our judgments are not sensitive to the truth of these claims. We would continue to make them, regardless of whether they were true. Hence, like John's claims about Sally, our moral judgments are unjustified, and therefore should be considered to be (probably) false.

How should we respond to this reinforced redundancy argument? The first thing we need to note is that it is very plausible to think that our evolved dispositions do track real properties (Rottschaefer & Martinsen, 1990). Indeed, Joyce himself concedes this point, noting that the dispositions which underlie morality would not be fitness enhancing if they were not a response to real properties in the external world. So the argument goes through only if there are good reasons to doubt that these real properties are *really* moral.

Joyce has an independent argument against identifying the dispositional properties that trigger our moral sentiments with properly moral sentiments. However, since the argument is entirely conceptual, and not evolutionary, I relegate discussion of it to a footnote.<sup>6</sup> Suffice it to say that it is ultimately unconvincing.

---

<sup>6</sup> The argument consists, essentially, in an attempt to undermine (4), the claim that we have reason to act morally. Construed literally, Joyce argues, moral claims make *categorical* demands on us; they require us to behave as they prescribe, regardless of our interests or desires. Such categorical requirements are, by their nature, reason-providing: If someone ought (morally) to  $\phi$ , then she has a *reason* to  $\phi$ , regardless of her interests or desires. But, Joyce argues, we can make no sense of a reason that is independent of our interests or desires. The broadest framework of reason-giving is the framework of practical rationality. It is also the only inescapable framework, because it, and it alone, is presupposed by any demand for reasons. But the demands of morality are not the demands of practical rationality, since the demands of the latter must be understood in terms of interests and desires. The demands of the only inescapable framework there is are hypothetical. So there cannot be the kinds of categorical demands to the

However, there is apparently another route to the same goal. It lies through that much-discussed scourge of evolutionary ethics: the (so-called) naturalistic fallacy. We need to examine this alleged fallacy, both because it is widely taken to invalidate any substantively evolutionary ethics, and for its implications for the dispositional analysis of moral concepts.

## B. The Naturalistic Fallacy

The naturalistic fallacy is often cited as a supposedly decisive objection to any evolutionary analysis of morality (Lemos, 1999; Woolcock, 1999). Someone commits the naturalistic fallacy when they attempt to define goodness in natural terms. Do evolutionary ethicists commit this fallacy? It is unlikely that they do, for the simple reason that it far from clear that there is any such fallacy (Smith, 1994, 2000).

Certainly, it is a mistake to think that any *evolutionary* analysis of the meaning of moral terms can capture the implicit

---

existence of which morality commits us to. Construed literally, moral discourse commits us to a kind of reason which is incoherent, so we are in error when we use it in this way.

Joyce is, no doubt, correct in holding that moral claims are appropriately analyzed as categorically binding. It doesn't follow, however, that we must understand categoricity in terms of practical rationality. Indeed, if we did cash out moral "oughts" in this fashion, we would miss their point. We are required to act morally, quite independently of our interests or desires. The moral ought can only be understood in irreducibly moral terms (Devitt, 2002). Thus, Joyce is right in claiming that moral judgments are categorically binding, and right to suggest that this entails that each person has a reason to act morally. But that reason must itself be understood morally. We ought to refrain from torturing innocent children because it is *wrong* to act in this manner, not because we necessarily have desires or interests which will be satisfied by so doing. Moral demands can only be grasped from inside the institution of morality. Cashing out its claims in non-moral terms is not merely contingently impossible, it is *constitutive* of morality that this is so (For similar reasons, attempting to show that morality is objective on the grounds that accepting some moral system is in our evolutionary interests, as Campbell [1996] claims, misses the point of morality and therefore fails to justify it.).

commitments necessarily had by all competent language users. If this were possible, then the meaning of such terms would have had to have undergone a significant alteration since the theory of evolution was formulated in the nineteenth century, and that is an extremely implausible suggestion (Joyce, 2000). But a correct analysis of a concept doesn't have to be *a priori*. Indeed, the analyses of natural kind terms offered by the sciences are *a posteriori*, and far from obvious. It is no objection to the proposition that water is H<sub>2</sub>O that competent speakers do not necessarily intend "H<sub>2</sub>O" by "water." Since the theory of evolution is a (well confirmed) scientific hypothesis, we can expect that the kind of analyses it is capable of generating will also be *a posteriori*. Though we certainly don't *mean* "likely to enhance inclusive fitness in the environment of evolutionary adaptation" (or whatever) by "good," we might discover that actions and character traits that we are disposed to call good are *in fact* likely to enhance inclusive fitness. Nothing Moore or his supporters have said suggest that this cannot be the case.

That said, it needs to be recognized that the analyses of moral terms that evolutionary ethicists have offered have tended to be very implausible. Consider a representative analysis:

the term "evil" designates behaviours by one or more members of a group (society) that, were it generalised, would reduce the long-term fitness (i.e., over many generations) of all members of the group. (Thompson, 2002: 246)

The problem with all such analyses is not that they commit any fallacy, it is that they are unconvincing as analyses, and they are unconvincing because of the revision of morality they would force upon us. Analyses do sometimes force us to revise our concepts. Consider the analysis of our concept "fish." When this concept was correctly analyzed, we realized that not all organisms we were disposed to call fish had the features mentioned in the analysis. Whales and dolphins, for instance, breathe air and bear

their young live. So we had to exclude whales and dolphins from the class of fish. This was a cost, in a sense, because the concept we were analyzing was one which we had previously been disposed to apply to dolphins and tuna equally, but it was a cost we were prepared to pay, since we now realized that there were very significant differences between dolphin and tuna, such that by applying different natural kind terms to them, we were doing a better job of cutting the world at its joints.

However, we are not prepared to pay the analogous cost in revising our moral terms. Consider how we would have to revise our morality if Thompson's analysis were to be accepted. If all and only acts which, if generalized, would reduce fitness in the long-term are to be described as evil, then indiscriminate altruism is evil, on the plausible assumption (defended in Mackie, 1978) that indiscriminate altruism generalized provides the conditions for a rapid increase in organisms that play the "defect" strategy in prisoner's dilemmas, and therefore lead to a fall in average fitness. Now, we can coherently debate whether indiscriminate altruism is good. Perhaps it is better described as unwise. But whatever else it is, it seems radically implausible to suggest that it is evil.<sup>7</sup>

The point can be generalized: any analysis of moral terms that claims that they ought to be understood *directly* in evolutionary terms—which claims that "good" is equivalent to "highly evolved," as Spencer had, it or to having the properties that would be approved of by organisms which had evolved under ideal conditions, as Collier and Stingl (1993) suggest—fails, because it implies that certain propositions, to which we are more strongly attached than we are to any evolutionary analysis of goodness, are false. At least so long as the models for describing the evolution of

---

<sup>7</sup> Harms's teleosemantic analysis of moral concepts fails for exactly the same reason. Harms argues that moral judgments are true just in case they are playing the role for which they were selected (Harms, 2000). But many true moral judgments are fitness-reducing, not enhancing, because morality has in some way broken free from the functions which account for its existence.

morality are reasonably accurate (and these models or something very like them are accepted by all evolutionary ethicists), it is easy to think of actions and character traits that would meet the requirements of the suggested analysis, but which are not good, or which are good, yet would not be approved of. It is plausible to think that xenophobia is an adaptation which boosts (or, in the environment of evolutionary adaptation, boosted) inclusive fitness, but xenophobia is not good.

It is not because these analyses commit a fallacy that they are false. It is not an open question whether indiscriminate altruism is evil, or xenophobia good. These are closed questions, and the analyses which entail them fail for this banal reason. They are false, not fallacious. This does not imply that evolutionary considerations might not figure importantly in moral theory, perhaps even at levels other than the genetic. But it does imply that philosophers like Joyce are right in saying that we cannot identify moral properties with the properties we have evolved to feel certain kinds of dispositions toward. To put the same point in another way, it is a mistake to identify proto-morality—the animal base, as it were of morality full-blown—with morality itself. On this point at least, we ought to side with Huxley against Spencer. We shall return to this point. For the moment, we need to pick up the main thread of our journey.

### C. Evolutionary Error Theories

It is a mistake, we now know, to identify moral properties with the real properties toward which natural selection has made us sensitive. Morality is not a system of enlightened selfishness, as Spencer thought. Or, more carefully, our concept of morality is not of such a system. According to some evolutionary ethicists, we have here the basis for an error theory: nothing answers to our concept of morality. The question therefore arises how we came by such a concept. Why have we made such a glaring error?

There is, of course, an evolutionary explanation available of

our alleged error. Since morality is fitness-enhancing, it is in our interests for us to be strongly motivated to act upon its dictates. Emotional pushes and pulls will frequently be sufficient to cause us to act appropriately, but morality will have an even more powerful grip on us if it is backed up by belief. Hence, not merely the dispositions to act morally, but also the disposition to believe that morality is objective, is adaptive. Both the subjective reality of morality, and the illusion of objectivity, are the product of natural selection (Ruse, 1998). In fact, morality is nothing more than a set of evolved dispositions, a “collective illusion foisted upon us by our genes” (1998: 253).

Defending an error theory successfully depends on making good on two requirements. It demands, first, that a convincing analysis of the concept in question be given, and second that it be shown that nothing corresponds to the concept. Mackie’s error theory, for instance, analyses moral claims as committing those who make them to the existence of objective prescriptive facts: facts the recognition of which is necessarily motivating. Mackie argued that such “queer” facts are metaphysical extravagances (Mackie, 1977). The evolutionary error theory analyzes morality as demanding more of us than we can give: it claims that our concept of morality is of a set of obligations which transcend the bounds of kin and reciprocity, but our moral dispositions are confined within these bounds. Our concept of morality commits us to (5), but (5) is false.

Those philosophers who are inclined to give dispositional analyses of moral concepts, however—with one or two explicitly evolutionary exceptions—do not take the dispositions in question to be simply evolved sentiments. On the contrary, philosophers who have taken this line explicitly argue that the dispositions in question need to be cultivated and trimmed (McDowell, 1985, 1995). For those of us who set great store by the method of reflective equilibrium, such a refined dispositional analysis seems much more convincing, indeed perhaps even inescapable, than the directly evolutionary analysis. Consider our brief discussion of

directly evolutionary analyses of moral concepts. We rejected these analyses because they had wildly implausible implications. In other words, they clashed with our moral intuitions, suitably adjusted to accommodate our best moral and empirical theories. On the assumption that these intuitions are internalized, so that they are reliably expressed in our moral sentiments (an assumption which seems plausible given the way in which the responses of people and entire societies change across time), it was indeed to our dispositions to which we appealed in rejecting the directly evolutionary analyses.

The best dispositional accounts of morality therefore do not seem to be directly evolutionary. If these accounts are acceptable, then our moral judgments do not commit us to any error—not, at least, to any error identified by the evolutionary ethicists. If evolution truly undermines moral realism, we must look elsewhere for the reason.

Our discussion of the meta-ethical implication of evolutionary hypotheses concerning the origins and function of morality has so far focused on a set of issues which centre around (3), the claim that morality has truth-conditions which are independent of the subjective states of those who make them, and (5), that morality requires us to give some weight to the interests of all parties affected by our actions. We now turn to (2), the claim that moral facts cannot vary across space and time, and are therefore equally binding on all rational creatures.

#### D. Morality on Other Planets

We may not be alone in the universe. Intelligent life may have evolved on other planets. Perhaps, indeed, we are not the only moral animals in the universe. But what would alien morality look like? It might be the case, some evolutionary ethicists argue (Ruse, 1998; Waller, 1996), that the kinds of actions which we regard as obligatory are rightly held to be immoral by some aliens. If their genetic constitution were different, to ours or if their evolution

took a different path, then the illusion of objectivity under which they labor might attach to actions we regard as immoral. Surely this is sufficient to show that objective morality is an illusion? If there were an objective morality, then it would be binding upon all rational creatures (as Kant points out). But there is no such morality.

What are we to make of this argument? The contention that the contents of our morality is sensitive to the details of evolutionary history is plausible. What counts as harming and benefiting someone, most obviously, is in important part a function of their biology, which makes them vulnerable to certain dangers and in need of certain resources and opportunities. But this fact is surely not sufficient to establish the species-relativism of morality, for it is vulnerable to the same kinds of replies that are used to discredit many claims of descriptive cultural relativism. We establish any kind of interesting relativism only if we show that members of different groups have different *fundamental* obligations (Moser & Carson, 2001; Rachels, 1995). Clearly, the fact that Australians are required to drive on the left hand side of the road, while Americans are required to drive on the right, does not establish any kind of interesting relativism. Similarly, the fact—if it is a fact—that we might have been required to eat one another's faeces (to use Ruse's own example) if evolution had taken a different path is not sufficient to establish the kind of species relativism that Ruse and Waller hope to demonstrate. It might be that a high enough level of abstraction, the kinds of morality which would emerge from the demands of cooperation would be the same in all possible worlds, consisting of injunctions to treat everyone impartially, to sacrifice one's own lesser interests for the greater interests of others, and so on. At least, nothing Ruse and Waller say shows that this is not the case.

However, Waller at least has a reply to this line of thought implicit in his work. Though it might be true that mere differences in the content of morality are not sufficient to establish the kind of species-relativity that would undermine the objectivity of morality,

the *right* kind of differences would have this effect, and the right kind of differences might arise from different patterns of genetic inheritance. Biologists have convincingly explained the intensely social behavior of ants, bees and termites—in particular, their apparent readiness to sacrifice themselves for the nest—in terms of their unique system of reproduction. Worker bees, for instance, are more closely related to one another than to their queen, and more closely related to each other than we are to our children. So they boost their genetic fitness by “farming” the queen, to make more siblings of theirs, and by sacrificing themselves for the nest. Now, Waller asks us to imagine that organisms with this kind of reproductive system evolved a degree of intelligence comparable to our, and as they did so came to possess a morality which is an outgrowth (at least in important part) of kin selection, just as ours probably is. Would they not have *fundamental* moral obligations to each other quite different from ours? They would find our emphasis on the individual and her rights “not merely absurd, but morally odious” (Waller, 1996: 253). This thought experiment shows that morality is indeed species-relative, in the strong sense required to undermine moral objectivity.

Or does it? Actually, there are problems with understanding just who is the subject of moral obligations in Waller’s thought experiment. It may be that individual intelligent ants do not have obligations toward one another because there are no **individual** intelligent ants, at least not in the sense in which we understand “individual.” Eusocial organisms like ants and bees are in many ways better thought of as constituting a single super-organism, than as so many separate individuals (One of the lessons of evolutionary biology is that many biological concepts have vague boundaries. “Species” is one well-known example, but “individual” may be equally vague. Is the Portuguese Man ‘o War one animal, or a colony of four different kinds of animals? There may be no unique defensible answer to this question.).

But if eusocial organisms are not individuals in their own right, then the possibility that individual intelligent ants would reject our

concept of individual rights does not demonstrate that moral concepts are species-relative, in the strong sense Waller requires. A nest of ants might be a rights-bearer, in the sense which concerns us, and all our moral concepts might apply to it. Indeed, it is not even clear that we can intelligibly attribute intelligence, as we understand it, to *individual* eusocial organisms. Like the Borg in *Star Trek*, they might constitute a single super-organism, with a single, distributed, mind. Waller therefore faces a dilemma: either the individual ants are not individuals at all, in the sense in which we are, and therefore the fact that they lack individual rights is no surprise, or they are individuals in our sense, and we have no reason to think them incapable of understanding and appreciating individual rights. On the first disjunct, their rights are appropriately (from our point of view) constrained by their genetic structure; on the second, they transcend that structure.

We might more appropriately regard the nest as the individual, the entity which has rights and obligations. In that case, the intelligent ants' lack of individual rights is no more interesting than is the fact that my skin cells lack rights against me.

Thus, Ruse and Waller fail to establish that morality is species-relative in any fundamental sense. Though it may certainly be wrong for us to treat the members of alien species in just the same ways we are required to treat one another, this is of no more meta-ethical significance than the fact that there are plants that are nutritious food sources for some animals and poisonous for others. There may be a single set of ethical obligations, consisting in injunctions to benefit and avoid harming others, though it may also be true that it will be difficult to know how to fulfil these obligations towards being sufficiently unlike us in biological structure.

#### IV. Spencer and Huxley Redux

By now it should be apparent that a number of apparently

separate claims advanced by the deflationary evolutionary ethicists are actually interconnected. They claim (a) that a dispositional analysis of morality is correct, in the sense that though it may not answer to our concept of morality (they are divided among themselves on this question), nevertheless it captures the reality of moral phenomenon. They also hold (b) that the dispositions in question are the *direct* products of evolution, and therefore do not extend much beyond kin and those in a position to benefit us in return. Finally, they argue that (a) and (b) imply (c) that the truth-conditions of moral claims make essential reference to the subjective states of those who make them: since morality is founded on a set of dispositions we share with much simpler organisms, it is essentially a matter of feeling, and not belief.

We might call the conjunction of these three claims the neo-Spencerian position. Neo-Spencerians follow Spencer in identifying morality proper with the set of evolved dispositions, in opposition to Huxley and neo-Huxleyeans<sup>8</sup> who believe that morality is importantly different from proto-morality. We cannot, as might have thought, put the Huxley/Spencer debate behind us. It continues to play itself out among evolutionary ethicists today.

As we have seen, the neo-Spencerians take their position to be better grounded scientifically. It makes human morality continuous with other animal behaviours, refusing—parsimoniously, as they see it—to postulate additional mechanisms to do the work in humans that instincts and feelings carry out in other animals. Irreducibly moral properties are mysterious and unscientific, they might claim. At very least, the burden of proof ought to be on those who reject the neo-Spencerian view, to show that additional mechanisms are necessary. Merely insisting that preferences cannot be moral reasons (Woolcock, 1993, 1999, 2000) just begs the

---

<sup>8</sup> Contemporary Huxleyeans—those who believe that culture or human cognitive abilities significantly transform our proto-moral dispositions—include, among philosophers, Dennett (1995, 2003), McGinn (1979), Rottschaefer (1998) and Singer (1981, 1999), and among biologists Dawkins (1976) and Williams (1995).

question.

We have, I believe, already seen that we have decisive reason to reject the neo-Spencerian view. In the course of considering the deflationary evolutionary arguments, we have already taken it apart, plank by plank. Lest it seem that our arguments rest on an ignorant or superstitious affirmation of mysterious human powers over good scientific argument, I want briefly to scrutinize the scientific credentials of the neo-Spencerian position before I rehearse the arguments against it we have already sketched. How well does the neo-Spencerian view achieve its main aim, and explain the evolution of our moral sense?

The Spencerians face a major challenge in explaining how it is we come by the concept of morality at all. Although game-theoretical modelling demonstrates how our moral instincts might have evolved, the same models also lead us to expect that such instincts will not be the only products of repeated occasions for cooperation and conflict in the paleolithic. Group selectionist hypotheses are sometimes criticized on the grounds that though they might explain the evolution of genuine intragroup altruism<sub>p</sub>, they would also tend to give rise to intergroup hostility and violence (Laland, Odling-Smee, & Feldman, 2000). In fact, the same charge can be generalized. Whatever set of processes gave rise to our cooperative dispositions, they—or some other set of interactions—would also have given rise to selfish<sub>p</sub> dispositions, dispositions which lead us to favor our own lesser interests over the greater interests of others. Our entire complex human nature has evolutionary foundations, and it does not take much reflection for us to realize that that nature includes a great many a- and immoral dispositions, as well as moral ones.

The challenge for the Spencerian is therefore this: how did we come to group a certain set of dispositions together, giving only to them the name of the moral sentiments? It might be suggested that these emotions, and only these emotions, motivate cooperative behavior, and that they therefore constitute a class that can easily be identified. But the class of dispositions which motivate (or

might motivate) cooperative behavior is not co-extensive with the class of moral dispositions. We can be motivated to cooperate for selfish<sub>p</sub> reasons, and we can have moral duties that are self-regarding. Sometimes, morality even requires us to act against our sympathetic impulses, as when we are cruel to be kind, or when our sympathetic impulse is out of place because the suffering which is its trigger is deserved punishment. Indeed, it is plausible to think that the linked notions of deserved punishment and undeserved suffering just are irreducibly moral notions that cannot be cashed out in nonmoral terms.

If morality is not to be entirely mysterious, however, we must be able to give some kind of explanation as to how we came to possess these irreducibly moral concepts. Here is one suggestion, that builds upon the work of evolutionary biologists. In a famous article, Robert Trivers suggests that self-deception might have been a product of an “arms race” between cheats and cheat-detectors in iterated prisoner’s dilemmas. The thought runs as follows: given that the disposition to cooperate becomes widespread in a population, the defect strategy becomes profitable, and is maintained at some, relatively low, level by frequency-dependent selection. In this situation, most people will be disposed to cooperate with one another, but will also always be on the look-out for defectors. Cheats will therefore have to develop increasingly sophisticated means to hide their true intentions, which will prompt the development of ever better cheat-detection mechanisms. Trivers argues that cheats will be far better at hiding their intentions if they are hidden even from themselves, for then they will not have to fear giving them away inadvertently. Thus, cheats will come to believe themselves to be genuinely altruistic<sub>p</sub>. Self-deception might thus be the product of natural selection, an offshoot of the profitable ability to deceive others (Trivers, 1985).

Trivers does not seem to have noticed, however, that the self-deception whose origin he traces itself requires that we come to have the notion of morality. To see this, consider what the conscious content of our beliefs must be, if we are to engage in

such self-deception. We must believe, not merely that we are disposed to cooperate with one another, but that we are disposed to cooperate in the right *manner* for the right **reasons**: to cooperate with cooperators, in order to achieve morally permissible ends. In other words, Trivers-style self-deception requires that we possess the concept of morality, morality full-blown, not mere proto-morality. If his suggestion that self-deception is itself an evolved characteristic is correct, we must have acquired the concept of morality at the same time, or earlier, for this self-deception requires it. Thus, promoting our own selfish concerns requires that we possess the idea of morality, on this hypothesis. If it is plausible, then the Spencerian view is self-defeating: the evolved dispositions cannot substitute for morality proper; instead, they require it.

This is precisely what we should expect, I suggest, given what we know about the evolution of morality. We know that it has foundations in proto-morality, in a set of evolved dispositions which, in the environment of evolutionary adaptation, were fitness-enhancing. But we know, too, that moral judgments have in some manner floated free of this subjective base. Plausibly, our moral emotions—what we can, in retrospect alone, recognize as such—are incoherent, in the state in which we inherit them, and need to be trimmed and altered, to be brought into something approaching wide reflective equilibrium. These dispositions made us sensitive to the needs and interests of others, but did so only partially and inconsistently, only to the extent to which this sensitivity was in our own interests. At the same time, however, it gave us the belief that our moral sensitivity should not be self-interested. This laid the groundwork for a thorough-going transformation of our inherited proto-morality, a transformation effected by our cognitive abilities (themselves products of natural selection).

Thus, we transformed our moral sentiments, rejecting some and changing others, so that the resulting morality, towards which we continue to work, bears only a passing resemblance to the set of

inherited dispositions. On this view, Huxley stands vindicated. Our morality is the product of evolution, in the sense that we came to have its raw material and the powers with which we reworked this material, as a result of natural selection. It remains the case, however, that the resulting morality is able to turn against the very processes which gave birth to it. Huxley was right: morality lies not in natural selection, but, sometimes at least, in combating it.

## V. Conclusion

Neo-Spencerians hold that the evolutionary explanations of the origin of morality demonstrate that it consists, wholly or mainly, in a set of evolved dispositions, which extend little further than our kin. If this were true, we would have strong reasons to believe that a widespread concept of morality, according to which it consists (at minimum) in a set of obligations which require us to give some weight to the interests of everyone, is false. Morality would have a greatly restricted range and importance. Perhaps it would be eliminated altogether.

If neo-Spencerians were correct in their central claim, that morality is to be identified with proto-morality, then claim (5) above—that morality requires us to give some weight to (at minimum) all persons affected by our acts—would be false. But we have seen no reason simply to identify morality with proto-morality. In fact, the notion that rationality adds nothing essential to morality is highly implausible, in view of the fact that we hold all and only rational beings morally responsible. It is very plausible to think that an essential threshold is crossed when rationality is added to proto-morality (McGinn, 1979).

We have also seen that there are decisive *moral* reasons to reject the proposed analyses. They would force revisions on our concept of morality which are far greater than we are willing to countenance. Of course, the neo-Spencerians will reject our right to appeal to our concept of morality in this context. But unless

they can make good on their identification of morality with proto-morality, we have no reason to take notice of them. Morality is proto-morality pruned and refined, proto-morality with at least the more glaring irrational elements eliminated. The intuitions it yields are far more likely to be truth-preserving than are those which the neo-Spencerian insists we ought to feel.

Though the evolutionists, of all stripes, are no doubt right in insisting that morality has evolutionary origins, it cannot be identified with the set of dispositions which we have as a result of our evolutionary history. This initial set of dispositions need do no more than make us sensitive to the needs and interests of others, and give us the concept of a morality that is impartial. From that point on, the significant work is carried out by rationality, which prunes proto-morality to make it answer to the concept of morality. Our moral sentiments are gradually extended, and our self-interest trimmed in the process. There are no decisive arguments against moral realism to be found in evolution.

What relevancies, finally, have the evolutionary hypotheses for our ethical and meta-ethical theories? The project of naturalizing ethics requires that we base our theories on the facts of human nature, as we know them, and of the social and physical world. A naturalistic ethics is concerned with what we are actually like. Evolutionary hypotheses can guide us, in suggesting hypotheses for further research: look for traits that would probably have been adaptive in the environment in which our ancestors evolved. But it cannot substitute for the examination of our actual characteristics. Indeed, theories of human nature, as it actually is, are themselves important constraints upon evolutionary theorizing. We go wrong, then, if we allow our theories to be driven by evolutionary hypotheses. They ought not to constitute a separate realm of theorizing about humanity and its nature, moral or otherwise. Instead, they constitute just one small part of the jigsaw, one more piece of evidence which, together with evidence from psychology and neurobiology, history and even literature, build up the picture of human nature.

## References

- Campbell, R. (1996). Can biology make ethics objective? *Biology and Philosophy*, 11, 1: 21-31.
- Collier, J., & Stingl, M. (1993). Evolutionary naturalism and the objectivity of morality. *Biology and Philosophy*, 8, 1: 47-60.
- Darwin, C. (1871). *The descent of man, and selection in relation to sex* (Vol. 1). London: John Murray.
- Dawkins, R. (1976). *The selfish gene*. New York: Oxford University Press.
- Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life*. New York: Simon & Schuster.
- Dennett, D. C. (2003). *Freedom evolves*. London: Allen Lane.
- Devitt, M. (2002). Moral realism: A naturalistic perspective. *Croatian Journal of Philosophy*, 2, 4: 1-15.
- Flack, J. C., & de Waal, F. B. M. (2000). "Any animal whatever": Darwinian building blocks of morality in monkeys and apes. *Journal of Consciousness Studies*, 7, 1-2: 1-29.
- Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.
- Hamilton, W. D. (1972). Altruism and related phenomena, mainly in social insects. *Annual Reviews of Ecology and Systematics*, 3: 193-232.
- Harman, G. (1977). *The nature of morality: An introduction to ethics*. New York: Oxford University Press.
- Harms, W. F. (2000). Adaptation and moral realism. *Biology and Philosophy*, 15, 5: 699-712.
- Huxley, T. H. (1989). Evolution and ethics. In J. Paradis & G. C. Williams (Eds.), *Evolution and ethics: T. H. Huxley's Evolution and ethics with new essays on its Victorian and sociobiological context* (pp. 57-174). Princeton, NJ: Princeton University Press.
- Joyce, R. (2000). Darwinian ethics and error. *Biology and Philosophy*, 15, 5: 713-732.
- Joyce, R. (2001). *The myth of morality*. Cambridge, UK:

- Cambridge University Press.
- Kevles, D. J. (1995). *In the name of eugenics: Genetics and the uses of human heredity*. Cambridge, MA: Harvard University Press.
- Kitcher, P. (1993). The evolution of human altruism. *Journal of Philosophy*, 90, 10: 497-516.
- Kitcher, P. (1998). Psychological altruism, evolutionary origins, and moral rules. *Philosophical Studies*, 89, 2-3: 283-316.
- Laland, K. N., Odling-Smee, F. J., & Feldman, M. W. (2000). Group selection: A niche construction perspective. *Journal of Consciousness Studies*, 7, 1-2: 221-225.
- Lemos, J. (1999). Bridging the is/ought gap with evolutionary biology: Is this a bridge too far? *The Southern Journal of Philosophy*, 37, 4: 559-577.
- Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Harmondsworth, UK: Penguin.
- Mackie, J. L. (1978). The law of the jungle: Moral alternatives and principles of evolution. *Philosophy*, 53, 206: 455-464.
- Maynard Smith, J. (1964). Group selection and kin selection. *Nature*, 201, 4924: 1145-1147.
- Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.
- McDowell, J. (1985). Values and secondary qualities. In T. Honderich (Ed.), *Morality and objectivity: A tribute to J. L. Mackie* (pp. 110-129). London: Routledge & Kegan Paul.
- McDowell, J. (1995). Two sorts of naturalism. In R. Hursthouse, G. Lawrence, & W. Quinn (Eds.), *Virtues and reasons: Philippa Foot and moral theory, essays in honour of Philippa Foot* (pp. 149-179). Oxford, UK: Oxford University Press.
- McGinn, C. (1979). Evolution, animals, and the basis of morality. *Inquiry*, 22, 1-4: 81-99.
- McShea, R. J., & McShea, D. W. (1999). Biology and value theory. In J. Maienschein & M. Ruse (Eds.), *Biology and the foundation of ethics* (pp. 307-327). Cambridge, UK: Cambridge University Press.

- Moore, G. E. (1903). *Principia ethica*. Cambridge, UK: Cambridge University Press.
- Moser, P. K., & Carson, T. L. (2001). Introduction. In P. K. Moser & T. L. Carson (Eds.), *Moral relativism: A reader* (pp. 1-21). New York: Oxford University Press.
- Palmer, J. A., & Palmer, L. K. (2002). *Evolutionary psychology: The ultimate origins of human behavior*. Boston: Allyn & Bacon.
- Rachels, J. (1995). *The elements of moral philosophy*. New York: McGraw-Hill.
- Ridley, M. (1996). *The origins of virtue: Human instincts and the evolution of cooperation*. London: Viking.
- Rosenberg, A. (2000). The biological justification of ethics: A best-case scenario. In *Darwinism in philosophy, social science and policy* (pp. 118-136). Cambridge, UK: Cambridge University Press.
- Rottschaefer, W. A. (1998). *The biology and psychology of moral agency*. Cambridge, UK: Cambridge University Press.
- Rottschaefer, W. A., & Martinsen, D. (1990). Really taking Darwin seriously: An alternative to Michael Ruse's Darwinian metaethics. *Biology and Philosophy*, 5, 2: 149-173.
- Ruse, M. (1998). *Taking Darwin seriously: A naturalistic approach to philosophy* (2nd ed.). Amherst, NY: Prometheus Books.
- Sesardic, N. (1995). Recent work on human altruism and evolution. *Ethics*, 106, 1: 128-157.
- Singer, P. (1981). *The expanding circle: Ethics and sociobiology*. New York: Farrar, Straus and Giroux.
- Singer, P. (1999). *A Darwinian left: Politics, evolution and cooperation*. London: Weidenfeld & Nicolson.
- Skyrms, B. (1996). *Evolution of the social contract*. New York: Cambridge University Press.
- Smith, M. (1994). *The moral problem*. Oxford, UK: Blackwell.
- Smith, M. (2000). Moral realism. In H. LaFollette (Ed.), *The*

- Blackwell guide to ethical theory* (pp. 15-37). Oxford, UK: Blackwell.
- Sober, E. (1994). Did evolution make us psychological egoists? In *From a biological point of view: Essays in evolutionary philosophy* (pp. 8-27). Cambridge, UK: Cambridge University Press.
- Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: MIT Press.
- Sober, E., & Wilson, D. S. (2000). Summary of *Unto others: The evolution and psychology of unselfish behavior*. *Journal of Consciousness Studies*, 7, 1-2: 185-206.
- Spencer, H. (1883). *The data of ethics*. New York: D. Appleton.
- Sturgeon, N. (1985). Moral explanations. In D. Copp & D. Zimmerman (Eds.), *Morality, reason and truth* (pp. 49-78). Totowa, NJ: Rowman & Littlefield.
- Thompson, P. (2002). The evolutionary biology of evil. *The Monist*, 85, 2: 239-259.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 1: 35-57.
- Trivers, R. L. (1985). *Social evolution*. Menlo Park, CA: Benjamin/Cummings.
- Vanderschraaf, P. (1999). Game theory, evolution, and justice. *Philosophy and Public Affairs*, 28, 4: 325-358.
- Waller, B. N. (1996). Moral commitment without objectivity or illusion: Comments on Ruse and Woolcock. *Biology and Philosophy*, 11, 2: 245-254.
- Waller, B. N. (1997). What rationality adds to animal morality. *Biology and Philosophy*, 12, 3: 341-356.
- Wilkinson, G. S. (1990). Food sharing in vampire bats. *Scientific American*, 262, 2: 76-82.
- Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.
- Williams, G. C. (1995). Mother nature is a wicked old witch. In M. H. Nitecki & D. V. Nitecki (Eds.), *Evolutionary ethics*

(pp. 217-231). Albany, NY: State University of New York Press.

- Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, 17, 4: 585-654.
- Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Cambridge, MA: Harvard University Press.
- Woolcock, P. G. (1993). Ruse's Darwinian meta-ethics: A critique. *Biology and Philosophy*, 8, 4: 423-439.
- Woolcock, P. G. (1999). The case against evolutionary ethics today. In J. Maienschein & M. Ruse (Eds.), *Biology and the foundation of ethics* (pp. 276-306). Cambridge, UK: Cambridge University Press.
- Woolcock, P. G. (2000). Objectivity and illusion in evolutionary ethics: Comments on Waller. *Biology and Philosophy*, 15, 1: 39-60.

## 當今演化倫理學的前景

尼爾·李維

(張滌之譯)

Florey Neuroscience Institutes, University of Melbourne  
Level 2, Alan Gilbert Building, 161 Barry Street, Carlton South, Victoria, 3053  
Australia  
E-mail: neil.levy@philosophy.ox.ac.uk

### 摘 要

解釋人類起源的演化論普遍受到抗拒的原因之一，在於不少人恐懼道德演化論可能摧毀道德的根基。這是由於人們懷疑，如果道德不過是立基演化之行爲傾向，將使道德顯得虛幻不實。若干哲學家也認同這個觀點，並用道德的演化起源來攻擊道德實在論。倘若上述爲真，那麼我們將面臨一個抉擇的困境，或以科學之名否決道德，或是否決目前最受證實的人類起源之科學解釋，以拯救道德。然而，如同本文呈現的，我們無需接受演化倫理學者的主張，因為道德正像我們所理解的一般，可與演化論相容並存。

**關鍵詞：**演化、倫理學、後設倫理學、史賓塞、赫胥黎