Biology and Philosophy **18:** 567–581, 2003.
© 2003 *Kluwer Academic Publishers. Printed in the Netherlands.*

Debunking morality: evolutionary naturalism and moral error theory

HALLVARD LILLEHAMMER
King's College, Cambridge University, Cambridge, CB2 1ST, UK
(e-mail: hallvard.lillehammer@kings.cam.ac.uk)

Received 20 November 2001; accepted in revised form 22 October 2002

Key words: R.D. Alexander, Evolutionary Naturalism, Moral Error Theory, Moral Objectivity, M. Ruse, R. Wright

Abstract. The paper distinguishes three strategies by means of which empirical discoveries about the nature of morality can be used to undermine moral judgements. On the first strategy, moral judgements are shown to be unjustified in virtue of being shown to rest on ignorance or false belief. On the second strategy, moral judgements are shown to be false by being shown to entail claims inconsistent with the relevant empirical discoveries. On the third strategy, moral judgements are shown to be false in virtue of being shown to be unjustified; truth having been defined epistemologically in terms of justification. By interpreting three recent error theoretical arguments in light of these strategies, the paper evaluates the epistemological and metaphysical relevance of empirical discoveries about morality as a naturally evolved phenomenon.

Introduction

Reflection on the fact that morality is a natural phenomenon gives rise to the question how we should respond to empirical discoveries about the origins of that phenomenon and its place in nature. Given the relative openness of our moral sensibility to the discovery of new facts, it is natural to think that our moral sensibility should be responsive to the discovery of facts about that sensibility itself. What form should this responsiveness take? On one view, the discovery of facts about the nature of our moral sensibility should be factored into morality in a piecemeal fashion. On this view, while some empirical discoveries may have significant effects on our moral judgements, others may not. On the one hand, a theory claiming that humans are inherently biased towards their kin as a result of the past effects of this disposition on their inclusive fitness would raise questions about the way this bias has traditionally been conceptualised in moral terms. On the other hand, an account of our sexual instincts in terms of a biological need to reproduce is unlikely to justify the rejection of all the pleasures of sex, and a similar revelation about the biological basis of parental love is unlikely to justify torturing babies for fun. This view is necessarily open-ended, and it is likely to be partial. Yet I think it is plausible, and shall endorse it in what follows. What I shall question is a stronger view, according to which the discovery of empirical facts about the nature of our moral sensibility has the power to falsify all moral judgements. On this view,

empirical facts about the nature of our moral sensibility shows that morality is committed to some form of metaphysical illusion, which has been variously compared to a belief in spiritualism, astrology, or Macbeth's hallucination of a dagger (c.f. Mackie (1977), Ruse (1986, 1995)). If this error theoretic view is right, all our moral beliefs ought to be rejected as false or incoherent.

In this paper, I evaluate the implications of empirical discovery for the epistemology and metaphysics of moral judgements. In doing so I shall question whether all moral judgements include a commitment to claims incompatible with empirical claims about the nature of our moral sensibility. I shall not assess the veracity of the empirical claims in question, but rather ask what the consequences would be if they were true. I thereby hope to throw some light on what conditions must be met in order for empirical discoveries about the nature of morality to have a debunking effect on moral judgements.

Morality

By morality, I mean a set of judgements entailing a commitment to the normative privilege of some practical options over others regardless of an individual's contingent ends. By a practical option I mean some way in which agents can act. I define morality this inclusively for two reasons. First, I want the definition to include some judgements we may not normally call moral, such as judgements of prudence or well being. The reason for this is that some judgements of these kinds are likely to be the target of debunking explanations of the same kind as those offered for what we are more naturally inclined to call moral judgements. Second, I shall remain neutral on the question of whether some normative concepts are morally more fundamental than others. My reason for this is purely strategic. The error theoretic views I discuss below make no commitment on the issue. Nor, therefore, shall I.

The Function of Morality

When speaking of an account of morality in non-moral terms I shall speak of the function such an account assigns to morality. The function of morality in which I am interested is one assigned to morality in what I call an empirical sense. What I mean by this is that the assignment to morality of a function is an empirical hypothesis about morality as a contingent and evolved natural phenomenon. The fact that morality is an evolved natural phenomenon is to be established regardless of the commitments embedded in the content of moral judgements, and therefore regardless of whether the idea of morality as an evolved natural phenomenon is embodied in the content of moral judgements. It might be true that some notion of morality as a contingent and evolved natural phenomenon is embedded in the content of some moral judgements. If it is, this is not essential to the assignment of such a function in an empirical sense.

Nor does the empirical assignment of a function to morality need to have the

status of a normative claim in ethical theory. The claim that morality is a contingent evolved natural phenomenon may conceivably constrain ethical theory by telling us what is possible or impossible, feasible or unfeasible, but in itself it need not (although it has on occasion been intended to by its proponents) tell us how to act (c.f. Sober (1994), Woolcock (1999)).

The functions most commonly attributed to morality are those of individual or collective co-ordination (c.f. Alexander (1987), Dawkins (1976), Ruse (1986, 1995), Sheffler (1992), Wright (1994)). On the standard views, morality is construed as a co-ordinating device serving a set of basic human needs and interests. Although the attribution of function may vary from the microbiological on the one hand to the psychoanalytical on the other, the basic idea is deceptively simple. By making moral judgements which assign normative privilege to practical options, agents can become motivated to pursue certain ends which they might not otherwise be inclined to pursue, where the ends in question are ones which protect or enhance some subset of basic human needs or interests. Consider, for example, judgements about how to act. An agent who judges that she has an overriding reason to jump into the lake to save the drowning child will have a rational incentive to save the child. For insofar as she is not motivated to save the child, she will appear irrational on her own terms. Insofar as the agent has a rationally unified intentional profile, she will be motivated to save the drowning child, even if she is initially disinclined to do so. It does not follow that she is actually is motivated to save the child. But insofar as she is not, she will appear irrational on her own terms.

According to the moral error theorist, the needs or interests which moral judgements serve are explicable in non-moral terms, and might include such ends as enhancement of inclusive fitness, reproductive capacity, sex, survival, power, or pleasure. The needs in question might be either individual or collective. A disposition to make moral judgements might be thought to protect such individual needs or interests as one's own survival, or such collective needs or interests as the reproductive potential of one's kin. A disposition to judge that one ought to save the drowning child might serve an agent's own interests insofar as the child is one's own, or the interests of the agent's group insofar as the child is a human being at some arbitrary degree of relatedness to the agent. While different accounts differ on the issue of what the relevant co-ordinating functions are, they agree that the existence of such functions can be a novel empirical discovery with significant consequences for a moral self-understanding to which the facts discovered have previously been unknown.

The assignment of a co-ordinating function to morality might be taken by its proponents to have empirical explanatory power in two superficially different ways (c.f. Kitcher (1998)). First, the co-ordinating function of morality could explain its role as a dynamic part of the complex structure that is human life (c.f. Cummins (1975)). Second, the coordinating function of morality could explain why it exists, both in the sense of why it has developed, and of why it persists (c.f. Wright (1983)). While the claim that the co-ordinating function of morality accounts for its existence is naturally associated with the claim that morality is a biological product of natural selection, it need not be. In principle, an explanation of the emergence of morality as

a co-ordinating device could be given in historical, economic, or psychological terms while maintaining the crucial feature that the explication of this function is a novel empirical discovery with consequences for our moral self-understanding (c.f. Railton (2000)). This fact has occasionally been overlooked in the recent literature on evolutionary naturalism. The consequence has been a lack of appreciation that the kind of debunking strategy proposed by some evolutionary naturalist writers on the nature of morality has clear historical parallels in such diverse areas as historical materialism and psychoanalysis, for example.

A functional explanation of morality can bring individual moral judgements within its scope in different ways. At its most potent, a functional explanation of morality would target every individual judgement and explain the function that judgement serves in the context in which it is made. Thus, a judgement that the wage demands of workers are immodest might function to serve the interests of the ruling class in the context of a dispute over wages. A universally plausible functional account would be more likely work indirectly. The significant autonomy agents enjoy in forming their moral judgements allows for individual judgements promoting ends contrary to those morality as a whole functions to promote. A deluded egomaniac might form judgements about his intrinsic moral greatness thwarting rather than promoting ends of individual of social co-ordination, for example. This possibility might appear to undermine the claim that morality has a co-ordinating function. That fact that a functional account might work indirectly explains why it does not do so. It might be the coordinating function of morality that explains how it is possible for the egomaniac to make his functionally degenerate moral judgements in the first place.

What distinguishes moral judgements from other judgements such that the existence of a functional explanation for moral judgements undermines their claim to truth, whereas a functional explanation of metaphysically respectable judgements does not undermine their claim to truth? Moral error theorists need an answer to this question. After all, the judgements included in the functional explanation of morality in non-moral terms may also serve to co-ordinate. So why not also give an error-theoretic account of them? Why not think the existence of such functional explanations undermines the claim to truth of all co-ordinating judgements equally? Thus, the biologist R L. Trivers writes that 'the conventional view that natural selection favours nervous systems which produce ever more accurate images of the world must be a very naïve view of natural selection' (Trivers (1976) vi).

In response, the moral error theorist might appeal to explanatory differences between the relationship which morality bears to its functional explanation and the relationship other judgements bear to their functional explanation. For example, only for some judgements may it be true that the entities they quantify over are themselves presupposed to exist by the theory providing the functional explanation of these judgements. The error theorist might argue that the functional explanation of morality does not presuppose the existence of the moral truths moral judgements aim to pick out, but only non-moral kinds, such as the physical, chemical, biological, psychological, social, mathematical, and modal kinds presupposed in some scientifically respectable account of reality. In this sense, morality would be

accounted for as a co-ordinating device evolved by moral animals in response to a non-moral reality. The error theorist might then claim that the functional account of the judgements embodied in the functional account of morality itself presupposes at least some of the kinds presupposed in the functional explanation of morality. These kinds might then be said to genuinely exist. What distinguishes morality from the judgements accounting for its function is then that only the latter are targets of a functional explanation presupposing the very metaphysics to which they are themselves committed. In this way, the metaphysical postulates of moral judgements can be eliminated in favour of those embodied in the functional account of morality in non-moral terms. It is a moot point where the error theorist would have to draw the line. It might be thought that the existence of believers in moral facts is sufficient to import a 'space of reasons' excluded from the domain of scientific facts evolutionary naturalists believe in (c.f McDowell (1994)). Yet the existence of beliefs in moral facts is obviously inconsistent with the non-existence of beliefs. A more interesting moral error theory would be one accounting for the illusion of moral facts at least partly in terms of the function of realistically construed psychological capacities. In what follows, I shall assume that the moral error theorist can draw the line where he likes.

Debunking Morality

What are the conditions under which a functional account of morality in non-moral terms would undermine moral judgements? I shall consider three different strategies by means of which an account of morality in non-moral terms might be used to debunk moral judgements. The first strategy is epistemological. On this strategy, moral judgements are undermined as resting either on ignorance or false belief, with the consequence that we lose our justification for the judgements in question. The second strategy is purely metaphysical; on this strategy, moral judgements are shown to entail false claims about the nature of reality. The third strategy combines the epistemological and metaphysical strategies by defining truth in epistemological terms. On this strategy, moral judgements are shown to be false by being shown to be unjustified because truth has been defined in terms of justification. I shall consider each strategy with reference to recent error theoretic arguments given by R. D. Alexander, Robert Wright, and Michael Ruse.

Epistemological Debunking Strategies

There are at least two kinds of epistemological debunking strategy. The first strategy derives from the fact that the justification of judgements can be undermined when judgements depend on relevant ignorance, New empirical knowledge might affect the justification of a moral judgement on the conditions that previous ignorance was relevant to the content of the judgement and the judgement can be shown to rest on

ignorance of this knowledge. These conditions apply because justification does not entail omniscience, either about morality or the world in general.

It is clearly true that some moral judgements are made in ignorance of empirical facts about the nature of morality. For example, the relevance of genetic relatedness to the scope of altruism could be a novel discovery, as could an explanation of moral sentiments in terms of subconscious forces like the super-ego or the will to power (c.f. Sheffler (1992), Nietzsche (1967)). Yet ordinary people are not completely ignorant of the function of morality. For example, the idea of morality as a social co-ordinatiag device is embodied in common sense moral appeals to the interests of one's family, tribe, nation, or species. Furthermore, it is not plausible that all moral judgements are made conditionally on ignorance of facts about the function of morality. As already mentioned, some moral judgements seem to be already premised on the idea of morality having a co-ordinating function. In addition, discoveries about the function of morality can be consistently incorporated into our moral self-understanding without radical revision of our moral judgements. As I noted in the introduction, biological explanations of the basis of parental love have not been shown to destabilise our antecedent commitment to its value. On the other hand, it is clearly possible that some moral judgements are made conditionally upon ignorance of the function of morality. For example, certain kinds of group, tribe, or gender bias could be undermined by a discovery of their instrumental relationship to the promotion of ends with which we would not reflectively identify.

Judgements can also be undermined when they rest on false beliefs, Empirical discovery can affect the justification of a moral judgement where the false beliefs are relevant to the content of the judgement and where the judgement is made conditionally upon the holding of beliefs inconsistent with this knowledge. It is possible that some moral judgements are made conditionally upon false beliefs about the function of morality. If the claims of some Marxists are correct, large parts of morality are an ideological construct by means of which the interests of the powerful are promoted at the cost of oppression of the weak (c.f. Geuss (1989)). Furthermore, if the claims of some evolutionary naturalists are correct, moral judgements are conceptually articulated instruments for the promotion of biological interests, the effectiveness of which is better served if they do not reflect what R. D. Alexander calls the 'evolutionary significance of our acts' (Alexander, 1987:159):

> 'Righteousness is a source of motivation. We gain by thinking that we are right, and by convincing our allies and enemies . . . It provides . . . a rationale for sinking deeper into self-deception about our motives and for justifying acts that could not otherwise be justified.' (123).

According to Alexander, morality is a system of indirect reciprocity in which 'confluences of interest within groups are used to deal with conflicts of interest between groups' (19). On Alexander's view, the justifications embodied in moral judgements do not work by way of picking out genuine moral features of the world. There is 'no single source of right or wrong' (191). Morality is an introspectively

opaque co-ordinating device designed to promote the 'compromises' and 'contracts' which are the 'only real solutions to actual conflicts of interest' (191). Alexander is committed to the claim that morality rests on false beliefs about its empirical nature. Alexander's view therefore entails the possibility that at least some moral judgements should be withdrawn in light of the discovery of the real motives behind them. For example, attributions of virtue and consequent approbation on the grounds of altruism could be withdrawn in light of the discovery that what appears to be altruism is really some form of selfishness in disguise. It does not follow that all moral judgements should be withdrawn in light of the discovery of the real motives behind them. While the discovery that what appeared to be an instance of altruism is really selfishness in disguise might be depressing, the possibility remains that we would on reflection prefer the appearance of altruistic acts over blatantly selfish behaviour. Furthermore, it is not clear that ordinary people are entirely deluded about their underlying motives, even if Alexander's account of them is correct. In fact, the idea that it can be in one's interest to behave as morality requires is one of the most platitudinous ways of justifying moral courses of action to agents whose moral commitments are in doubt. It therefore does not follow from Alexander's account that all moral judgements are unjustified. Continued (if guarded) commitment to the dispositions associated with friendship, solidarity and community are consistent with a realistic view of the empirical nature of morality. This is not to say that the bulk of moral judgements is necessarily immune to epistemological debunking of the kind Alexander suggests. Suppose, contrary to currently available evidence, that the entire practice of moral judgement is a manipulative ploy imposed on humans by evil spirits, functioning in the service of their murderous ends and predictably leading to the extinction of the human race. Such a discovery would justify a radical rethink of our moral commitments. Even so, it is hard to imagine what could be meant by the claim that such a discovery would justify the abandonment of morality as a whole. There are other justifications for abstaining from genocide than its recommendation by norms imposed by evil spirits. Not every part of morality is plausibly destructive from either an individual or a collective point of view. In fact, it is hard to imagine how a purely destructive system of norms could have survived across the entire human race over any evolutionarily significant period of human history.

Epistemological debunking entails the absence of justification for the moral judgements debunked. If Alexander is right, some moral judgements are unjustified for this reason. Epistemological debunking does not entail the falsehood of the moral judgements debunked. Contrary to what his remarks suggests, it does not follow from Alexander's account that any moral judgements are false, nor is it clear from his account how the falsehood of any moral judgement would follow from the soundness of an epistemological debunking strategy alone. After all, true beliefs can be held without justification. In order for an epistemological debunking strategy to have error theoretic consequences there must be some bridge principle connecting the absence of justification and the absence of truth. Over the next three sections, I shall discuss two different ways of supplying such a principle. I shall argue that neither strategy succeeds in establishing a universal moral error theory.

Metaphysical Debunking Strategies

The success of a metaphysical debunking explanation of moral judgements depends on the existence of claims embodied within those judgements inconsistent with that explanation. Although error theorists do not always mark the distinction, such a theory could have varying scope, from a universal theory embracing all moral concepts, to a partial theory focusing on some subset of moral concepts, such as 'reason', 'value', 'good', or 'right' (c.f. Joyce (2001)). The evolutionary naturalists Wright (1994), Ruse (1986, 1995) both defend error theories with a universal flavour. Wright states his position as follows:

> '. . . belonging to a species . . . whose members justify their actions morally, we are designed to think of ourselves as good and our actions as defensible, even when these propositions are objectively dubious. The new paradigm, by exposing the biological machinery behind this illusion, makes the illusion harder to buy' (1994: 338-9).

According to Wright, '. . . there's definitely no reason to assume that existing moral codes reflect some higher truth apprehended via divine inspiration or . . . philosophical inquiry' (147). The moral truths allegedly postulated by moral judgements are nothing but a 'genetically orchestrated sophistry'; postulates which are '. . . as practically necessary as they are intellectually vacuous' (147, 353).

Ruse agrees that empirical discoveries are capable of explaining away the apparent reference of morality. In apparent contrast to Wright, however, Ruse also appears to claim that we have no choice but to remain in the grip of the illusion of moral objectivity:

> 'The Darwinian argues that morality simply does not work . . . unless we believe that it is objective. Darwinian theory shows that . . . morality is a function of (subjective) feelings; but it shows also that we have (and must have) the illusion of objectivity' (1986: 253).

According to Ruse, we have no choice but to make moral judgements in promotion of our needs and interests, in spite of the fact that these judgements are both epistemologically and metaphysically unfounded. I shall return to the question of how we should understand this claim after evaluating the prospects of the evolutionary naturalist brand of the metaphysical debunking strategy.

Truth Defined Epistemologically

One way of providing a bridge principle between justification and truth is to claim that truth is partly definable in terms of justification. On such an account, judgements are *response dependent*, in the sense that their truth is definable in terms of the beliefs of agents in epistemologically privileged circumstances (c.f. Wright

(1992)). Epistemologically constrained response dependence entails a conceptual connection between justification and truth, from which it follows that at least some forms of successful epistemological debunking are logically paramount to successful metaphysical debunking.

There are good reasons to believe that an epistemologically constrained response dependent account is true of some moral judgements, Some moral concepts carry a claim to response dependence on their sleeve. This is arguably true of some morally thick concepts, such as 'wisdom' or 'virtue' (c.f. McDowell (1984)). For the idea of a wise or virtuous person to play an intelligible role in moral discourse and education, there must be an intimate link between what wisdom and virtue are on the one hand, and what people could on reflection come to consider wise or virtuous on the other. The claim that a wise or virtuous person is therefore also admirable makes no sense without the assumption that a wise or virtuous person is also someone who, in reflectively adequate conditions, people would actually admire. The idea that all possible insight into the nature of wisdom or virtue could possibly have escaped us is as bizarre as the idea that there could be a funny joke the funniness of which it is not possible to detect (c.f. Wright (1992)). (A parallel, but less plausible, claim about colour judgements is known as 'revelation' c.f. Jackson (1998).)

There is a second set of moral judgements for which an epistemologically constrained response dependent account is plausible. This is a subset of the moral judgements including the concept 'obligation'. Suppose Jones has a moral obligation to not insult Smith behind his back. It makes no sense to claim that Jones could have such an obligation if no person, even in the most reflectively privileged circumstances, would think that Jones has such an obligation. The social and intellectual role of the concept of a moral obligation, and the link this concept has to concepts like blame, advice, promise, and agreement would be unintelligible on the assumption that the facts about our obligations escape us completely. Consider finally the example of the evil spirits discussed above. If these evil spirits were to impose a set of standards of behaviour to evaluate human action and the correctness of these standards were completely beyond the grasp of human beings, it is obscure how the standards in question could impose moral obligations humans. Perhaps they could impose moral obligations on evil spirits. Even so, the humans would need their own principles to get along, and the obligatory nature of these had better be at least partly within human grasp. This is evidence for the existence of a necessary connection between the concept of moral obligation and the standard epistemological virtues of justification, such as reflective coherence, comprehensiveness, explanatory power, stability in light of further information, and so on (c.f. Railton (1986), Wright (1992), Brower (1993), Smith (1994), Jackson (1998)). To this extent, the concept of moral obligation is arguably different from some of the concepts employed in the natural sciences, the extensions of which we can at least imagine to outrun the reach of our best efforts to grasp them, For example, it is clearly possible that our biological origins, as postulated by evolutionary naturalism, should have remained forever beyond our grasp. Not so Jones' obligation to refrain from insulting Smith behind his back, and understandably not so on an epistemologicaily constrained response dependent account of moral obligation.

The plausibility of an epistemologically constrained response dependent account of some moral judgements is enough to support a metaphysical interpretation of the epistemological debunking strategies discussed in the previous section. The overall metaphysical significance of this depends on two factors. First, it depends on the range of moral concepts for which an epistemologically constrained response dependent account is plausible. Arguably, there are moral concepts for which a response dependent account is not plausible (c.f. Lillehammer (2000)). Yet a response dependent account is plausible for other moral concepts, such as those of wisdom, virtue, and obligation. Second, the metaphysical significance of a bridge principle between moral truth and justification depends on the potential for epistemological debunking of the relevant kind. I have argued that this scope is likely to be partial. It is therefore unlikely that a metaphysical debunking strategy based on an epistemologically constrained definition of moral truth will have universal scope.

Moral Truth and Explanatory Efficacy

Another way of providing a bridge principle between justification and truth is to claim that there are reasons to believe in the existence of truths on a given subject matter only if such truths form part of the explanation of our beliefs about that subject matter (c.f. Sturgeon (1984)). Thus, given that the explanation of beliefs in evil spirits does not include any truths about actually existing evil spirits, we have no reason to believe that claims entailing the existence of evil spirits are true. Michael Ruse makes implicit use of this principle in his argument for a moral error theory:

> 'The objectivist must agree that his/her ultimate principles are . . . redundant. You would believe what you do about right or wrong, irrespective of whether or not a 'true' right or wrong existed . . . Given two worlds, identical except that one has an objective morality and the other does not, the humans therein would think and act exactly the same ways, Hence the objective foundation for morality is redundant' (Ruse, 1986: 254).

Let us call this the *redundancy argument*. According to Ruse, the fact that we make the moral judgements we do is fully explained by their co-ordinating function. Thus, if there were any objective moral truths, they would be redundant in the explanation of our moral beliefs. It is a plausible epistemological principle that if your belief in some proposition varies irrespective of whether that proposition is true, then you cannot be said to know the truth of that proposition. It follows that if there were any objective moral truths, we would have no knowledge of them. I have already argued that for some moral truths this would be absurd. So far, however, Ruse has only made an epistemological claim about knowledge, not a metaphysical claim about truth, It does not follow from the fact that your belief that P varies regardless of whether or not P is true either that P is not true or that there is no fact about whether P is true. Ruse's metaphysical claim is based on the implicit invocation of the

aforementioned bridge principle about the role of truth in the explanation of beliefs, According to that principle, the only reason we could have for believing in objective moral truths is that they form part of the explanation of why we have the moral beliefs we do. Yet according to the redundancy argument, objective moral truths form no part of the explanation of why we have the moral beliefs we do. It follows that there is no reason to believe in objective moral truths.

Ruse claims the existence of objective moral truths is inconsistent with the objective truth of an evolutionary naturalist account of morality. What does Ruse mean by 'objectivity'? The only credible interpretation of the objectivity claimed by Ruse's evolutionary naturalism is objectivity in the sense of epistemologically unconstrained response *independence*. According to Ruse, the biological and evolutionary truths explaining the existence of our moral judgements are truths obtaining independently of the beliefs, desires, and other responses of those who try to find out about them. Ruse is clearly committed to the (perfectly plausible) claim that our evolutionary origins might have remained forever unknown.

If the objectivity claimed by moral judgements were the epistemologically unconstrained and response independent objectivity of Ruse's evolutionary naturalism, then the redundancy argument would have some force. It is not obvious what reason there could be to believe in response independent moral truths if reference to a response independent non-moral reality were sufficient to explain why we have the moral beliefs we do. The same does not hold, however, if the objectivity claimed by moral judgements is the epistemologically constrained and response dependent objectivity defended for some moral judgements in the previous section. There are two reasons for this.

First, there is no valid inference from the non-existence of *response independent* moral truths to the non-existence of moral truths *as such* if moral truths are response dependent. Just as the non-occurrence of colour terms in a physical account of colour perception fails to undermine common sense colour attributions if truths about colour are explicable in terms of the dispositions of agents to regard things as coloured, so the non-occurrence of moral terms in a functional account of morality fails to undermine common sense moral judgements if moral truths are explicable in terms of the dispositions of agents to regard things in moral terms. (c.f. Rottschaefer and Martinsen (1990), Jackson (1998)). I have argued that some moral judgements are so explicable.

Second, if moral truths are response dependent, then there is a sense in which they are arguably explanatorily efficacious. On an epistemologically constrained response dependent account, moral truth is partly definable in terms of the responses of agents in conditions where they display the standard epistemological virtues (c.f. Smith (1994), Jackson (1998)). Yet ordinary people frequently change their minds, or loose their conviction, in the face of evidence that their moral judgements do not exhibit the standard epistemological virtues. It follows that epistemologically defined moral truths can play a role in the explanation of moral beliefs. Not only is this view consistent with the existence of a functional account of morality in non-moral terms, it can provide an explanatory supplement to such an account. One way in which moral judgements can co-ordinate is by reliably tracking the responses

of agents in conditions constitutive of the standard epistemological virtues. Even if moral truths are not explanatorily efficacious in this way, it does not follow that all moral judgements are false. If some moral truths are epistemologically definable in terms of justification, and if some moral judgements are relevantly justified, then those judgements are true, regardless of whether we know that they are true, and thus of the explanation why we make them. It follows that the redundancy argument fails to undermine the metaphysical commitments of the set of moral judgements for which an epistemologically constrained response dependent analysis is true. I have argued that there is good reason to believe that this set is not empty. If so, the redundancy argument fails to establish the truth of a universal moral error theory.

Two Levels of Moral Commitment

Even if moral judgements were not committed to claims falsified by a functional account of morality in non-moral terms, much philosophical theorising about morality would be. Wright is correct to question whether there are objective moral truths apprehensible by divine inspiration. And if the only possible mode of apprehension of moral truth were a quasi-perceptual mode of experience with intrinsic motivational efficacy, there would be something to be said for Mackie's notorious argument from queerness (Mackie 1977). But even if false philosophical claims abound, it does not follow that such claims are constitutive of human morality. In some places, Ruse appears to admit as much:

> 'In the case of morality, we are all part of the game, and even those of us who realise this have no desire to drop out. Thus, literally, we would not speak of an 'illusion'. Illusion in ethics means believing one has a moral obligation to be friendly to cabbages. But ethics is still what I have referred to as a 'collective illusion', in the sense that we all think it is something it really is not' (1986: 257).

And again:

> 'It seems therefore, morality is a collective illusion foisted upon us by our genes. Note, however, that the illusion lies not in the morality itself, but in its sense of objectivity' (1986: 253).

According to Ruse, the illusion of objectivity is not, after all, detrimental to all the substantial commitments embodied in common sense morality. At least some substantial moral commitments can survive reflection on the fact of illusion:

> 'Nor would I have those of us who see the illusory nature of morality's objectivity throw over moral thought' (253),

What Ruse appears to commit himself to in these remarks is a distinction between

illusions at two levels of moral commitment. On the first level, an agent might be suffering from illusion about what moral judgements to make. On the second level, an agent might be suffering from an illusion about the nature of moral judgements. Ruse thinks it is an illusion in the first sense that we should be friendly to cabbages and an illusion in the second sense that moral judgements pick out a realm of response independent moral facts. Ruse claims we all suffer from the second illusion, but not that we all suffer from the first. It follows that illusion at the second level does not entail illusion at the first.

According to Ruse, the illusion at the second level involves what has become known as a 'projective error', whereby states which do not aim to represent a response independent reality are mistaken for states which do (c.f. Mackie (1977), Blackburn (1993)). He describes the projective error in terms of a distinction between a cognitivist and a non-cognitivist account of morality:

> Although I am a non-cognitivist . . . I differ from other non-cognitivists. For someone like the emotivist, normative ethics has to be translated out as a report on feelings, perhaps combined with a bit of exhortation . . . For me, this is simply not strong enough. The way in which biology avoids this is by making moral claims seem as if they were objective . . . what I want to suggest is that - contra to the emotivists - the meaning of morality is that it is objective. Because it is not, it is in this sense that it is an illusion . . . (Ruse, 1995: 254).

While one connotation of the term 'non-cognitivism' is the denial of knowledge which Ruse defends for moral claims, there is also a sense in which it is misleading to call Ruse's view non-cognitivist. As that term is also frequently used, non-cognitivism comprises the variously sophisticated versions of emotivism which Ruse rejects (c.f. Gibbard (1990), Blackburn (1993, 1998)). Ruse's error theory would not gain any support from non-cognitivism so understood. A commitment to this latter kind of non-cognitivism would entail a double layer of protection from error for moral judgements. On this view, not only can first order moral judgements be insulated from false second order judgements about the nature of morality. According to this brand of non-cognitivism, the false appearances which generate the alleged illusion of moral objectivity are not in fact representations at all, but rather superficial aspects of non-representational moral sentiments (c.f. Blackburn (1998)). If Ruse wants to reject this form of non-cognitivism, how can he make sense of the notion of a projective error? The obvious answer is to adopt the kind of response dependent account discussed in previous sections. On this view, the projective error consists in confusing a response dependent conception of moral objectivity with a response independent conception. First order judgements could then indeed be free from second order illusion. Yet it would be implausible to claim that morality is an illusion with everyone in its grip. On this view, the illusion of moral objectivity is arguably a philosophical disease, not an affliction of morality as such. Furthermore, the illusion would now be revealed by an account of the content of moral judgements, not by producing a functional account of morality in terms of

its evolutionary origins. In order for the illusion in question to be universal, the confusion between response dependence and response independence would have to be embedded in the content of moral judgements in such a way that being so confused would count as a necessary condition of mastery for *all* moral concepts. There is no good reason to think that such an argument is forthcoming.

Conclusion

I have accepted that empirical reflection on the nature of morality can generate arguments in favour of the revision of moral judgements. I have denied that empirical reflection on the nature of morality generates arguments for a universal moral error theory. My rejection of this claim does not rule out that a universal moral error theory can be established on other grounds. Were the truth of a moral error theory to be established on other grounds, an explanation would be called for how a practice of judgement could have developed while universally committed to falsehood. A functional account of morality in non-moral terms could provide this explanation by describing how the needs of human beings are served by having false moral beliefs. If so, there would be a genuine abductive relationship between moral error theory and a functional account of morality in non-moral terms. Yet to construe this abductive relationship as one in which the truth of a universal moral error theory can be inferred from the existence of a functional account of morality in non-moral terms would be seriously misleading.

References

Alexander R.D. 1987. *The Biology of Moral Systems*. Aldine de Gruyter, New York.
Blackburn S. 1993. *Essays in Quasi-Realism*. Oxford University Press, Oxford.
Blackburn S. 1998. *Ruling Passions*. Oxford University Press, Oxford.
Brower B.W. 1993. 'Dispositional Ethical Realism'. *Ethics* 103: 221–249.
Cummins R. 1975. 'Functional Analysis'. *Journal of Philosophy* 72: 741–764.
Dawkins R. 1976. *The Selfish Gene*. Oxford University Press, Oxford.
Geuss R. 1989. *The Idea of a Critical Theory*. Cambridge University Press, Cambridge.
Gibbard A. 1990. *Wise Choices, Apt Feelings*. Oxford University Press, Oxford.
Jackson F. 1998. *From Metaphysics to Ethics*. Oxford University Press, Oxford.
Joyce R. 2001. *The Myth of Morality*. Cambridge University Press, Cambridge.
Kitcher P. 1998. 'Function and Design'. In: Hull D. and Ruse M. (eds), *The Philosophy of Biology*. Oxford University Press, Oxford, pp. 258–279.
Lillehammer H. 2000. 'Revisionary Dispositionalism and Practical Reason'. *The Journal of Ethics* 4: 173–190.
Mackie J.L. 1977. *Ethics: Inventing Right and Wrong*. Penguin, London.
McDowell J. 1984. 'Morality and Secondary Qualities'. In: Honderich T. (ed.), *Morality and Objectivity*. Routledge, London.
McDowell J. 1994. *Mind and World*. Harvard University Press, Cambridge, MA.
Nietzsche F. 1967. In: Kaufman W. (ed.), *The Will to Power*. Vintage, New York.
Sober E. 1994. 'Prospects for an Evolutionary Ethics'. In: *From a Biological Point of View*. Cambridge University Press, Cambridge, pp. 93–113.

Railton P. 1986. 'Moral Realism'. *Philosophical Review* 95: 163–207.
Railton P. 2000. 'Morality, Ideology, and Reflection; or, The Duck Sits Yet'. In: Harcourt E. (ed.), *Morality, Reflection & Ideology*. Oxford University Press, Oxford, pp. 113–147.
Rottschaefer W.A. and Martinsen D. 1990. 'Really Taking Darwin Seriously'. *Biology and Philosophy* 5: 149–173.
Ruse M. 1986. *Taking Darwin Seriously*. Basil Blackwell, Oxford.
Ruse M. 1995. *Evolutionary Naturalism*. Routledge, London.
Sheffler S. 1992. *Human Morality*. Oxford University Press, Oxford.
Smith M. 1994. *The Moral Problem*. Basil Blackwell, Oxford.
Sturgeon N. 1984. 'Moral Explanations'. In: Copp D. and Zimmermann D. (eds), *Morality, Reason, and Truth*. Rowman and Littlefield, Totowa, pp. 49–78.
Trivers R.L. 1976. 'Foreword'. In: Dawkins R. (ed.), *The Selfish Gene*. Oxford University Press, Oxford.
Woolcock P.G. 1999. 'The Case against Evolutionary Ethics Today'. In: Ruse M. and Maienschein J. (eds), *Biology and the Foundations of Ethics*. Cambridge University Press, Cambridge, pp. 276–306.
Wright C. 1992. *Truth and Objectivity*. Harvard University Press, Cambridge, MA, 1992.
Wright L. 1983. 'Functions'. *Philosophical Review* 82: 139–168.
Wright R. 1994. *The Moral Animal*. Little, Brown & Company, London.