Philosophical Explorations, Vol. 9, No. 1, March 2006

Routledge
Taylor & Francis Group

AS YOU WERE?
Moral philosophy and the aetiology of moral experience

Garrett Cullity

What is the significance of empirical work on moral judgement for moral philosophy? Although the more radical conclusions that some writers have attempted to draw from this work are overstated, few areas of moral philosophy can remain unaffected by it. The most important question it raises is in moral epistemology. Given the explanation of our moral experience, how far can we trust it? Responding to this, the view defended here emphasizes the interrelatedness of moral psychology and moral epistemology. On this view, the empirical study of moral judgement does have important implications for moral philosophy. But moral philosophy also has important implications for the empirical study of moral judgement.

Introduction

An impressive array of scholars from the sciences of human thought and behaviour—principal among them psychology, anthropology, evolutionary biology, game theory, neurology and cognitive neuroscience—is now at work on explaining the phenomenon of moral experience and judgement in human beings. There has been a recent surge of work on the proximal aetiology of moral judgement, examining the causal structure of episodes of moral judgement and the relation of our capacity for moral judgement to our other cognitive and affective capacities. This in turn informs the more speculative, but potentially more important, exercise of reconstructing its distal aetiology: explaining how our species has come to possess this capacity, by reference to the formative circumstances of our Pleistocene ancestors—showing how the way we human beings are now has resulted from the way we were.

This essay asks how much attention moral philosophy should be paying to this work. The next section distinguishes the main issues discussed by moral philosophy. I then consider those issues in turn, asking how far they are affected by empirical theorizing about two topics: first, the relationship between reason-recognition, emotion and moral judgement; and secondly, the evolutionary origins of moral dispositions. I shall argue that although the more radical conclusions that some writers have attempted to draw from this work are overstated, few areas of moral philosophy can remain unaffected by it. The most important question it raises, I shall argue, is in moral epistemology. Given the explanation of our moral experience, to what extent can we trust it? Responding to this, I shall argue for a view that emphasizes the interrelatedness of moral psychology and moral

ISSN 1386-9795 print/1741-5918 online/06/010117-15
© 2006 Taylor & Francis
DOI: 10.1080/13869790500492730

epistemology. On this view, the empirical study of moral judgement does have important implications for moral philosophy. But moral philosophy also has important implications for the empirical study of moral judgement.

Five Issues in Moral Philosophy

Moral philosophy discusses five main kinds of question.

Normative ethics asks which moral evaluations should be made and why.[1] On most views, it prominently includes questions about the rightness and wrongness of actions. But it also evaluates a range of objects other than actions: persons, character traits, states of mind, groups, institutions and states of affairs. Normative ethical theory aims to give a general account of the structure of our reasons (what I shall be calling our 'constitutive justifications') for evaluations of these various kinds, and their relationship to each other.[2]

Meta-ethics, by contrast, asks about the nature and status of ethical evaluation. Its questions are of four main kinds. *Moral psychology* asks what kind of state of mind you are in when you make an ethical evaluation, and its relationship to action or motivation to act. Is it a cognitive, belief-like state? Or does the tightness of the connection between ethical evaluation and motivation make it a non-cognitive state, akin to desire or emotion? *Moral semantics* studies the meaning of moral sentences. Are moral sentences truth-apt, so that the utterance of them makes an assertion? Or does their surface assertoric form disguise a different underlying semantic function, so that uttering them amounts to the non-truth-apt expression of a non-cognitive attitude—a command, an attitude of allegiance, or an attitude of approval or disapproval? *Moral metaphysics* asks whether there are mind-independent moral facts of which our moral evaluations give a more or less accurate report. How could such facts be related to those which the natural sciences report on?

The topic of the other main branch of meta-ethics—*moral epistemology*—is naturally described as the justification of moral judgements. However, that can be misleading, since 'the justification of a moral judgement' can mean two different things. First, this phrase can concern the content of a moral judgement—that a particular action is wrong, say—and can refer us to why that is so. If the action is wrong because of the distress it causes, then we can say that its causing distress is the reason or justification for the wrongness of the action. A justification of this kind refers to what *makes* the action wrong, so I shall call it a 'constitutive justification'. It should be contrasted with a second thing: the warrant a person has for making a moral judgement, which I shall call that person's 'epistemic justification' for the judgement. Constitutive justifications of judgement-contents are clearly different from epistemic justifications of states of judging.[3] I might be warranted in thinking that an action is wrong when it is not (perhaps I'm warranted in believing incorrectly that it causes distress): if it is not wrong, there is nothing that makes it wrong, and therefore no constitutive justification of the judgement that it is wrong. Conversely, there can be constitutive justification without epistemic justification: I might not be warranted in thinking that an action is wrong when there *is* something that makes it wrong. Moral epistemology, then, concerns the epistemic justification of states of moral judgement—moral judgings. It concerns the grounds that warrant a person in making a judgement with a given content.

Constitutive justification is the province of normative ethics, as described above.

This gives us five kinds of enquiry in moral philosophy: the questions of normative ethics, and the four different groups of questions in meta-ethics. To what extent can they be answered independently of empirical work on the structure and aetiology of moral experience?

A dominant theme of recent work on moral judgement in psychology and neuroscience has been the core role played by the emotions. The ways in which emotional impairment leads to dysfunctions in moral judgement and behaviour have been studied in psychopaths (Blair 1997; Blair et al. 2006), patients suffering damage to the prefrontal cortex (Anderson et al. 1999; Damasio, Tranel, and Damasio 1990), and autistic subjects (Hauser, Young, and Cushman forthcoming). Psychologists have also studied the structure of reasoning about trolley-problems and other standard philosophical thought-experiments in neurotypicals (Haidt 2001; Haidt et al. 1993; Hauser, Young, and Cushman forthcoming), and brain scientists have examined the localization of neural activity during judgements about such cases, or about simpler moral sentences (Greene and Haidt 2002; Moll, de Oliveirra-Souza, and Eslinger 2003). It emerges that moral judgements are influenced by emotionally laden framing effects (Sinnott-Armstrong forthcoming; Tversky and Kahneman 1981), vary with a subject's degree of emotional involvement (Greene et al. 2001), and can even be influenced by the hypnotically induced association of emotions with non-moral words (Wheatley and Haidt forthcoming). It thus seems not only that emotions are needed to make moral judgements, but that manipulating our emotions will alter them (Prinz 2006).

No one is claiming to have found evidence for the existence of a discrete mental module for morality. But an influential view is that our moral experience is standardly the product of intuitive, automatic, affective processes—processes that are pervasive in our formation of preferences and evaluations generally (Zajonc 1980), and in social cognition more particularly (Bargh 1994), to which moral judgement is tightly linked (Haidt 2001). Our intuitive moral responses are shaped under social pressures, but then are activated unreflectively. Moral reasoning is typically recruited after the formation of moral judgements, adducing grounds for evaluations already reached (Haidt 2001). Our moral experience is thus one expression of a general capacity we have for effortless classification and evaluation, with which we navigate the social world. The usefulness of this capacity is why it has evolved its way into us. And there are convincing explanations of why expressions of this capacity should include the altruistic and cooperative forms we associate with morality (Sober and Wilson 1998; Trivers 1985).

Some writers have been quick to claim that this empirical work supports conclusions in moral philosophy. In moral psychology and moral semantics, it has been taken to support noncognitivist or sentimentalist views, on which moral attitudes are seen as emotional states or dispositions towards them, and moral sentences express such states or dispositions (Greene 2003; Nichols 2004; Prinz 2006). In moral metaphysics, it has been taken to support the anti-realist denial that there are any mind-independent moral facts (Greene 2003; Joyce 2006). In moral epistemology, it has been taken to undermine the evidential status of pre-reflective moral intuitions (Sinnott-Armstrong forthcoming). And in normative ethics, it has been taken to support consequentialism, by undermining the rational credentials of the psychological mechanisms that produce nonconsequentialist judgements (Greene 2003; Sinnott-Armstrong forthcoming).

In what follows, I shall point out some of the problems with these claims. Many moral philosophers will agree with that assessment. Finding the empirical work both unsurprising and irrelevant, they will hold that, indeed, it leaves us just as we were. But I shall argue against that view too. This work does have wide significance for moral philosophy. I shall argue for a rather different view of its significance than others have done. To arrive at that view, I shall work through the issues distinguished above. My focus, in keeping with the theme of this volume, will be on the four metaethical issues. However, if (as I shall be arguing) empirical work has significant implications for moral epistemology, then it will have significant implications for normative ethics too—for it will affect which views in normative ethics we are warranted in holding.

Moral Psychology and Moral Semantics

What kind of psychological state or states is moral judgement? A forceful line of thought is this: this is a straightforwardly empirical question—one on which philosophers have always felt free to indulge in armchair speculation, but which can now be handed over to proper empirical investigation by psychologists.

That investigation is now intensively under way; and some are ready to announce its results. What we have discovered, they claim, is that moral judgements are emotional states, or dispositions towards emotional states. In combination with a noncognitivist view of emotions, this generates noncognitivism about moral judgement.[4]

To see that this is premature, though, we should begin by distinguishing four claims:

(i) Moral judgement is an emotional state.

(ii) Moral judgement is a disposition to have certain emotions.

(iii) Emotions cause moral judgements.

(iv) Moral judgements cause emotions.

Recent empirical research appears to show not just that emotional states tend to co-occur with moral judgements, but that the two are causally linked. Hypnotizing people to experience disgust upon hearing the word 'often' will induce them to judge that actions described using the word 'often' are wrong (Wheatley and Haidt forthcoming): this supports (iii). Moreover, (iv) might also be true: some empirical results seem at least to suggest this, if not to compel it. An example is the finding of Moll, de Oliveirra-Souza, and Eslinger (2003) that the brain areas associated with emotion are activated when subjects judge 'They hung an innocent' to be wrong, but not when they judge 'Stones are made of water' to be wrong. It will be consistent to hold both (iii) and (iv), provided it is not claimed of a single token occurrence of an emotion that it both causes and is caused by the same token moral judgement.

Some surveys of the possibilities in modelling the psychology of moral judgement restrict themselves to different possible configurations of causal interaction between reasoning, emotion and moral judgement (e.g. Hauser, Young, and Cushman forthcoming). But this overlooks the possibility of constitutive rather than causal models of the relationship between these states—models such as (i) and (ii). Our understanding of the brain as a parallel distributed processor rather than a serial processor of information should make us especially alert to such possibilities.

What is the relation between the constitutive claims and the causal claims? (i) and (ii) are compatible with either the assertion or the denial of (iii) or (iv). Notice that neither constitutive claim can derive direct *support* from either causal claim. No token occurrence of an emotion can cause itself. Nor can the existence of a disposition to have an emotion be caused by that same emotion, since dispositions cannot causally explain or be explained by their own realization. To combine (i) or (ii) with (iii) or (iv), you must claim that the emotions (or dispositions towards them) that constitute moral judgements enter into causal relations with *other* emotions.

Having noticed these possibilities, however, we ought to discard the simplest constitutive model—one that identifies all moral judgements with emotional states along the lines of (i). The empirical evidence clearly rules it out. The everyday observation that there can be dispassionate moral judgements is unsurprisingly borne out by functional magnetic resonance imaging studies (Greene and Haidt 2002, 519). However, there are other, more plausible views to consider. One of them is:

(v) Moral judgements can be cognitive constituents of emotional states.

This claims that certain emotional states are partly constituted by moral judgements. Part of what it is to have the emotion of indignation or guilt, on this view, is to make (or perhaps, to be disposed to make) a moral judgement (that you have been wronged or have done wrong, respectively). The judgement is necessary for the emotion but not sufficient—thus allowing for the possibility of dispassionate moral judgements.

So: several different causal and constitutive claims seem consistent with the existing evidence from psychology and neuroscience. How do we choose between them?

Claims about moral psychology had better be empirical claims: there had better be differences between the ways they claim the world to be, if (i)–(v) are to describe genuine alternatives concerning the nature of moral attitudes. And indeed, it should be clear how further empirical observations by psychologists and neuroscientists can contribute to deciding between them. To check whether the link between emotional and moral judgement states is causal, we can investigate its disruption, examining whether the causally antecedent state can occur without its normal effect. To check whether the relation is constitutive, we can investigate whether removing the constituent state always removes the constituted state.

However, although empirical data will help us to decide between these claims, it will not do so automatically. Given a body of empirical data, we face questions about how best to describe and explain it. The data is not self-interpreting. For example, brain science supports the everyday observation that sincere moral assertion is not always accompanied by emotion. How, then, should we describe the difference between the cases where it is and those where it isn't? Should we say that moral judgement sometimes but not always causes emotion? That emotion sometimes but not always causes moral judgement? That moral judgement is a disposition to emotion, but that the disposition is not always activated? Or that moral judgement is a constituent in certain emotional states, but can occur without the other constituents? What is at issue here is, in part, whether those 'other constituents' *are* the emotion or not. And that is a classificatory question: a question about the semantics of emotion-terms.

Thus, questions about psychological explanation are not independent of questions about the semantics of psychological terms. The latter are conceptual questions; but that does not prevent them from being empirical ones too. For semantics is an empirical study. If words are arbitrary symbols, they acquire their meaning through their use. So their meaning must be ascertained by careful observation of actual usage, and convincing theorizing about that.

It might therefore seem straightforward to resolve the classificatory questions just identified. To settle whether someone emotionally disengaged from morality counts as making moral judgements, for example, we should simply survey the population and see whether most people say so. The extension that a majority of English-speakers gives to the expression 'making a moral judgement' settles its meaning in English. And indeed, some empirical studies of moral judgement seem to have proceeded on this basis (e.g. Nichols 2002, discussed by Kennett 2006).

However, this is naive. We do not settle whether whales are fish or whether spiders are insects by polling the general population. Such a poll would not settle the semantics of 'fish' or 'insect': it would only reveal people's rudimentary theories of fish and insects. Likewise, we cannot settle the semantics of moral terms by doing laboratory studies of people's rudimentary semantic theories of those terms. Rather, the semantic theorist needs to study people's actual usage of those terms, and then to produce a convincing theory that distinguishes correct from incorrect usage. We cannot settle the definition of 'morality' itself, or of any moral term, by questionnaire, for the same general reason that this will not work for biological terms. The people we are surveying might include bad speakers. They may well include people who are good speakers but bad lexicographers. And they are likely to include people who are passable lexicographers but crude biologists or philosophers.

A first conclusion to draw from this is that in moral psychology there is a two-way traffic between empirical science and philosophy. When moral psychology seeks to explain and classify states of moral judgement and their relationship to other psychological states, what it is explaining and classifying is the data gathered from everyday experience and scientific observation. Philosophy that proceeds in ignorance of this cannot be telling us what those states really are. But scientific explanation that proceeds independently of philosophy will be begging semantic questions that need to be philosophically informed.

A second conclusion will come later in this paper. Because of the interconnectedness of the questions of philosophy, this two-way traffic reaches further into moral philosophy than many people think. After turning to the other issues of moral philosophy, we will be equipped to come back and choose between the views in moral psychology surveyed above.

Moral Metaphysics

It is in moral metaphysics that the largest claims have been made for the philosophical significance of empirical research on moral judgement. Our moral judgements can be explained as the triggering of reactive dispositions that humans have evolved to navigate their social environments without effort. This, it is argued, undermines their claim to represent mind-independent moral facts (Greene 2003; Joyce 2006).

This style of argument, offering a debunking genealogy of morals, is an old one, pre-dating Marx and Nietzsche at least to Plato, who dramatizes it in the characters of Thrasymachus and Callicles. However, there is an important difference between the older genealogical arguments and this new one. The older arguments sought to unmask morality as an instrument of particular social interests: for Marx and Thrasymachus, it is a device instituted by the strong to dominate the weak; for Nietzsche and Callicles, it has been devised by the weak to restrain the strong. Such arguments seek to explain our moral culture in terms of social dynamics, and thus support a critical evaluation of it. In contrast, the argument currently being drawn from evolutionary psychology is that humans' evaluative dispositions can be explained independently of whether the world contains any facts about the goodness and badness of actions and states of affairs at all.

Whereas the older arguments simply targeted conventional moral standards, evaluating them from an alternative standpoint, the new argument would seem to target all ethical evaluation whatever.[5]

It is sometimes complained that debunking genealogical arguments commit a 'genetic fallacy'—the fallacy of inferring the falsity of a belief from an explanation of its origins. A favourite response is a *reductio*, pointing out that *every* judgement will have some genetic explanation—including the judgement that morality is undermined by its genealogy (Railton 2000). However, this can be answered by distinguishing between those genetic explanations that are and are not independent of the truth of the judgements being explained (Joyce 2006). If part of the best explanation of why we make a certain judgement is that it is true, then obviously the judgement is not undermined but supported: a genetic explanation undermines only those judgements for which this is not the case.

What are the options for moral metaphysics in response to this argument? There are five main ones. The first is to respond with a conception of moral facts that includes them within the evolutionary psychologists' genetic explanation of moral experience. The most obvious version of this is a naturalistic view—for example, one that identifies moral facts with facts about what conduces to social cooperation. The genetic explanation of cooperative dispositions as adaptive could then be rephrased by saying that a sensitivity to moral facts is adaptive. However, notice that the same general strategy seems available to someone who rejects the naturalistic identification of moral facts with natural facts. Thus, consider the view that moral facts are facts about *reasonable* terms of social cooperation, where this is held not to allow a naturalistic reduction. Someone holding this view can agree that our possessing cooperative dispositions is explained by their adaptiveness, while insisting that these dispositions amount to a responsiveness to (non-natural) moral facts.[6]

Forgoing this first option means abandoning the idea that moral experience is explained by moral facts. But that does not mean abandoning the idea that there are moral facts. For we need to be clear about the dialectical status of the genealogical argument. If our making a certain judgement can be explained without any reference to its truth, that is not itself a piece of evidence that it is false. To think so *would* be to commit a fallacy. Rather, it is an objection to one kind of argument for the truth of the judgement—namely the argument that our making the judgement is itself evidence for its truth.

The second option in response to the genealogical argument, therefore, is to claim that we have *other* reasons to believe that there are moral facts. Again, the most obvious versions of this strategy are those that identify moral facts with natural facts. Given that we have good reasons to think that natural facts exist, they supply us with good reasons for believing that moral facts exist. Even if our making moral judgements can be plausibly explained without reference to their truth, we have independent grounds for believing them to be true.[7]

These first two strategies are ways of retaining moral facts in the face of the genealogical arguments. The other options involve doing without moral facts. A third strategy agrees that independent moral facts are incredible, but maintains that moral thought, discourse and practice make perfect sense without them (Blackburn 1998; Gibbard 1990). A fourth concedes that our existing moral practice does presuppose the existence of incredible moral facts, but holds that something more defensible can be devised and recommended in its place. And a final option, reached upon the failure of all the possible variants of the first four, will be moral nihilism—the view that moral judgement must be abandoned altogether.

What is striking about this list is that it actually includes every view under serious consideration in moral metaphysics. The only view it leaves out is the kind of moral Platonism that posits a non-physical realm of values which causally impinges on us in a quasi-perceptual way. The genealogical argument undermines this view—for unless our moral experience was evidence for the existence of such entities, what other evidence could there be? But moral Platonism has manifest problems anyway, independent of any biologically informed genealogical argument—principal among them its far-fetched metaphysics, its mysterious epistemology and its difficulty in explaining the normativity of moral values.

Thus, the genealogical argument is compatible with all of the serious contenders, and is not needed to rule out the implausible ones. In moral metaphysics, that is to say, the current work in explaining the nature and origins of moral experience does leave us exactly as we were.

Moral Epistemology

Empirical investigations of moral experience leave moral metaphysics unaffected. But with moral epistemology it is a different story. Here, we face a fundamental challenge. What kind of evidential force is carried by our intuitive, non-inferential moral judgements?

Empirical work supports this challenge in two complementary ways (Joyce 2006; Sinnott-Armstrong forthcoming). One appeals to the kind of genealogy of moral judgement just discussed. As we have just seen, this does not show that there are no moral truths. But it could still undermine the idea that intuitive moral judgements are a reliable guide to knowing which moral truths there are—by showing that those judgements can be explained independently of their tendency to be true. Secondly, recent work in psychology and brain science suggests that moral evaluation recruits psychological 'mechanisms' that are implicated in manifestly *un*reliable forms of non-moral judgement. Thus, according to Jonathan Haidt (2001, 819–23), our moral reasoning, like our broader social reasoning, is systematically directed by relatedness motives (evolutionarily influential incentives to side with allies), coherence motives (to preserve one's view of oneself and the world), and mechanisms of bias (to focus on citing confirming evidence); and it is standardly employed in producing *post hoc* rationalizations of change-resistant judgements generated through a process of automatic affective response.

A first response to this work is that it is limited to an epistemically unimpressive class of moral judgements: instant moral verdicts about short descriptions of possible situations. There are few moral epistemologies that give much evidential weight to such pure knee-jerk moral reactions. Standard Rawlsian versions of coherentism work from *considered* moral judgements—those endorsed upon reflection as having been formed under conditions conducive to undistorted judgement—and sensible versions of foundationalism will not be based on anything less.

However, pointing this out does not dispose of the challenge. If I consider my own intuitive judgements and then endorse them, how does this kind of self-validation succeed in supporting them?[8] One of the themes of the work mentioned above, after all, is that reasoning about emotionally based judgements often demonstrably serves the purpose of *post hoc* rationalization rather than of seeking independent confirmation.

This challenge should be taken seriously. I want now to outline a way of addressing it. I shall begin by explaining how the corresponding challenge can be addressed for many everyday social judgements. What we learn from them will help us to see what to say about moral judgements.

We often make social judgements about other people—judgements that they are friendly or insincere, for example—and we can be warranted in doing so. Sometimes, our warrant can consist in trusting our social 'intuitions' or hunches. The fact that Fred makes me feel uneasy can be a reason for thinking that he is insincere; the fact that Betty seems friendly can be a reason for thinking that she is friendly. My track record in forming social intuitions of these kinds may make them reliable indicators of the truth of the corresponding judgements. Notice that in the first of these examples, it is an emotion (my uneasy feeling) that provides the reason. In the second, what provides my reason is my disposition to judgement itself. Betty's seeming friendly to me is my disposition to judge that she is friendly. And at least sometimes, I can be warranted in trusting such dispositions.

Such reasons are defeasible. For we know that our judgements about and emotions towards other people are fallible. We make mistakes, and moreover, we do so in systematic ways—ways that correspond to the ones Haidt documents for moral judgements and emotions. We are subject to pervasive incentives to preserve our own self-image and to side with allies. Therefore, the trust I have in my social intuitions cannot be unqualified. The fallibility of my judgements and emotions means I should try to corroborate them, in important or controversial cases. How can I do this? In two main ways. One is by giving constitutive reasons for the conclusions asserted in my social judgements—spelling out just what it is about Fred that is insincere or Betty that is friendly. The other is by checking my reactions against other people's.

However, even where I cannot articulate what it is about Betty that is friendly, that need not prevent my impression from providing me with evidence that she is friendly. After all, I have grounds for doubting whether I can articulate all the reasons to which I am capable of responding. This is a matter of everyday observation: a socially astute person can find the demeanour and gestures that will put other people at ease, but explaining what made these gestures the appropriate ones is a task for the skilled novelist rather than the skilled socialite. And unsurprisingly, this is once more backed up by brain science, which documents the ways—some of them startling—in which our information-processing and problem-solving capacities can outrun our linguistic and conceptual capacities (Ramachandran 2003; Sacks 1985). Of course, if I should be aware of evidence that Betty is unfriendly, or I have grounds for thinking that the incentives to inaccurate judgement are affecting me here, then that will undermine the extent to which my impression that she is friendly is evidence that she is. But merely being unable to articulate what I am responding to does not extinguish my warrant for trusting my social intuitions.

So our dispositions to social judgement and our emotions can both provide us with evidence warranting social judgements. Now let us ask about the relationship between these two kinds of evidence. Suppose Fred makes me uneasy, and he seems insincere to me. I have a disposition to judgement and an emotion, and each of these things gives me a (defeasible) reason for thinking him insincere. However, they need not give me *separate* reasons. If Fred seems insincere to me, that may be evidence that he is insincere. Adding to this that I feel uneasy about Fred need not supply *further* evidence that adds extra support to my judgement.[9] This poses a puzzle: if each of these states provides evidence for my judgement, how could combining them fail to provide me with stronger evidence?

Here is another puzzle. If I do manage to spell out what it is about Fred that is insincere, then that can help to justify my judgement that he is insincere. But it can also help to justify my *feeling*. My uneasiness about him can be vindicated by identifying the constitutive reason for the conclusion asserted in the corresponding social judgement. How can that be?

The best solution that I can see to this pair of puzzles is to adopt a constitutive rather than a causal model of the relationship between dispositions to judgement and emotions. My disposition to judgement can be a cognitive constituent of my emotional state. This is not to say that I cannot have an uneasy feeling about Fred without judging that he is insincere: clearly I can. But when I have an uneasy feeling with this judgement, the judgement is a constituent of my emotional state, rather than a separate state that stands in a causal relationship to it. Saying this would resolve our two puzzles, as follows. The two states will not provide separate reasons for my judgement if they are the same state. *My reaction* to Fred can be evidence for a judgement about him, if my reactions are trustworthy—and that reaction can be an emotional state with a cognitive constituent. Moreover, if the emotional state has as a constituent a disposition to judgement, that would explain how it can be supported by identifying constitutive reasons for the relevant judgement. So the second puzzle would be solved too.

In Section 3, we saw that issues of correct psychological explanation are connected to classificatory questions concerning the semantics of psychological terms, and thus to philosophy. What we are now finding is that they are connected to epistemology. The argument just presented travels from epistemology to psychology: it is an argument for thinking of the relation between certain psychological states as constitutive rather than causal.

Thus, we can answer a skeptical challenge to the evidential status of our social intuitions; and in doing so we reveal an argument for drawing conclusions about their psychology. Now let us see how this helps with the epistemology of moral judgement.

Here again, our moral 'intuitions' can be either dispositions to judgement or emotional states. An action might seem wrong to me, or I might feel averse to doing it—I might feel that it would be a shabby thing to do. Can we have good reasons to trust our moral intuitions, as we can to trust our social intuitions? There might seem to be a fundamental obstacle to extending the argument. The reliability of our social intuitions can be independently confirmed. If Betty seems friendly to me now, I'll eventually find out whether she really is friendly. I can therefore acquire inductive grounds for trusting my social impressions. In the moral case, by contrast, there is no independent way of checking my moral impressions, so the argument breaks down.

However, there is a response to this. An action's seeming morally wrong to me is its appearing to have features that make it morally wrong—features that are constitutive reasons for its wrongness. In making this judgement, I might not have a clue what those features are; but the action's seeming wrong is its seeming to me that there is something about it that makes it wrong. However, this *can* be subsequently confirmed: I can subsequently identify features that *are* constitutive reasons for its wrongness. My initial impression that the action is shabby—that there is something morally objectionable about it—might be backed up by identifying it as exploitative, or callous, or disloyal. We can back up our conclusion-asserting intuitive judgements with reason-identifying judgements.

This might seem unhelpful. After all, it simply backs up one kind of moral judgement with another. And don't we need some further ground for thinking that this is not simply a *post hoc* rationalization of our initial intuitive reaction?

To answer this, we need to make two points. The first is that it is not relevant to be raising a worry here about whether there *are* any constitutive moral reasons. Empirical research into moral judgement does not call this into question. To do so, it would need to call into question whether there are any moral truths. And we saw in the previous section that it does not do that. The only conception of moral truths against which that research tells is independently implausible. It leaves open all the conceptions of moral truths that are under active consideration.

The relevant question here is not whether there are any constitutive moral reasons, but whether our judgements are a reliable guide to the constitutive reasons there really are. How do we rebut the case for thinking that such judgements are not just *post hoc* rationalizations of automatic, entrenched reactions? The second point is that empirical research into moral judgement does not support a worry of this kind either. One way to show this is to document the empirical counter-evidence: the ways in which reasoning can directly disrupt automatic judgement formation, and can itself condition the kinds of judgements which are automatically formed (Fine 2006). The other is to point out that the blanket claim that all judgements about moral reasons are unreliable *post hoc* rationalizations is actually self-defeating. How could such a claim itself ever be warranted? It could be warranted directly if you could be warranted in identifying the considerations that really are good moral reasons, and then in finding that our judgements about moral reasons correlate with them poorly. But this requires *your* judgements about moral reasons to be reliable. On the other hand, it could be warranted indirectly if you could be warranted in believing that moral judgements rely on psychological processes that are also involved in manifestly unreliable non-moral judgements. (This would give you an inductive case for thinking that moral judgements are unreliable too.) But again, the claim that those processes are so flawed that *all* the judgements that involve them are unreliable would be self-defeating. A warrant for this claim requires a warrant for thinking that those judgements correlate poorly with the facts that they claim to report. But this again requires that the person making the claim is able to make reliable judgements about those facts. The problem is a general one. Any evidence that certain kinds of judgement are unreliable itself presupposes our capacity to distinguish good reasons from bad. Therefore this challenge *cannot* apply to all of our judgements about reasons.[10]

So: empirical research into moral judgement does not undermine the claims that there are constitutive moral reasons and that we can know what they are. Of course, to say this is not to *support* those claims. To do that would be to supply a moral epistemology, and I am not about to attempt that here.[11] My point is that it is a mistake to think that empirical research on moral judgement is any obstacle to supplying this.

But if that is true, then it does after all make sense to meet the sceptical challenge concerning intuitive, conclusion-asserting moral judgements by appealing to warranted reason-identifying moral judgements. When the epistemic credentials of the former are challenged by citing empirical work on moral judgement, it makes sense to point out that the epistemic credentials of the latter are not. The response to this challenge can therefore parallel the treatment of social judgements above. That an action strikes me as shabby or noble does not guarantee that it is: my intuitive judgements are fallible. When it seems to me that there is something wrong with an action, and this is important or controversial, I should try to articulate what makes it wrong. And if my reason-identifying judgement is warranted—I am warranted in thinking the action dishonest, say—then I will be warranted in judging that it is wrong. True, no account has been given here of what it is for reason-identifying judgements to be warranted. But if they can be—and empirical work on moral judgement is no obstacle to this—then an epistemic justification for our conclusion-asserting intuitive judgements can be derived.

What if I cannot articulate a reason? Given the sceptical challenge, my inability to articulate a constitutive reason should call into question my warrant for the intuitive judgement. However, although it will call this into question, it will not straightforwardly defeat it. For, as we have seen, I have grounds for doubting whether I can articulate all the reasons to which I am capable of responding. And my track record in correlating my moral intuitions with constitutive reasons may be good enough to give me grounds for trusting them. If so, then when my conscience troubles me about an action, that will be evidence—defeasible evidence—that there is a reason not to perform it which I have not succeeded in articulating. True, the appearance that there is a reason may be what is at fault, rather than my inability to identify it. That is something I have reason to scrutinize. But unless I have evidence that I am mistaken, I will have a warrant for being reluctant to perform the action about which I feel uneasy.

Thus, just as our social intuitions can carry evidential weight, our moral intuitions can do so too. I can find that my moral intuitions are reliable indicators of the existence of constitutive moral reasons; and when they are, it makes sense for me to trust them.

Finally, notice that our argument from the epistemology to the psychology of social judgement is paralleled for moral judgement too. The two corresponding puzzles arise: If intuitive moral judgements and moral emotions both provide me with epistemic reasons, how is it that adding one to the other need not strengthen my warrant for judgement? And how is it possible for my reason-identifying judgements to support my moral emotions? Again, the best way to solve these puzzles is to adopt a constitutive rather than a causal view of the relationship between moral judgement and moral emotion. Citing both states will not strengthen my warrant if they are the same state. And if the judgement is a constituent in the emotion, that would explain how the emotion can be supported by giving reasons for the constituent judgement. Thus, from the range of views in moral psychology surveyed in Section 3, we have an argument for endorsing:

(v) Moral judgements can be cognitive constituents of emotional states.[12]

On our way to defending this, we have also seen that we should accept a further claim about the relationship between them—this time, neither a causal nor a constitutive claim, but a claim about epistemic justification:

(vi) Emotional states can be evidence for the truth of their constituent moral judgements.

Empirical work on moral judgement *is* important to moral epistemology. It shows that our intuitive hunches need to be backed up. However, we have found that they can be. And in doing so, we have also found that moral epistemology has an important impact on moral psychology. I have not been arguing that this is a one-way relationship: of course, any theorizing about the relationship between psychological states must be anchored in empirical data. But when we do that theorizing, we need to bring our conceptual, epistemological and normative claims into harmony with each other.

Conclusion

This survey—a quick one, it is true—of the main questions of moral philosophy suggests that only in the area of moral metaphysics (actually an area in which some writers have been quickest to draw conclusions on the strength of empirical research) can moral philosophy afford to proceed in ignorance of current work on the structure and origin of moral experience and judgement. Moral philosophers should get out of their armchairs and at least venture as far as the campus library to acquaint themselves with this research. I have suggested that they should think carefully about the results once they have, and have pointed out a number of respects in which arguments drawing philosophical conclusions from such research seem premature. However, I have drawn attention to a range of ways in which empirical work has an important bearing on moral philosophy. Its most significant impact is in moral epistemology, where it presents an important challenge to the epistemic status of noninferential moral intuitions, but also helps to answer that challenge.

NOTES

[1] I use 'ethics' and 'morality' interchangeably in this essay.

[2] Thus, one obvious project of this kind—common to utilitarians, Kantians, contemporary contractualists and pluralists in the style of Ross (1930)—is to identify the basic categories of moral reasons providing fundamental moral principles from which justified evaluations can be derived. However, I would also include Jonathan Dancy's (1993, 2004) particularist view, according to which all moral reasons are context-dependent and there are no such principles, as a normative ethical theory in the sense given in the text. He is offering a general account of the structure of our reasons. Bernard Williams (1985), by contrast, is rejecting normative ethical theorizing of all of these kinds.

[3] 'Justification for moral judgement' might mean a third thing: a practical reason for engaging in the activity of judging. That is not something I discuss in this essay. However, it will be important to be aware of the need to distinguish between three of the things covered by the phrase 'moral judgement': a state of judging, the activity of arriving at that state, and the content of that state – what is judged. For further discussion of these three, plus a fourth, see Cullity (1998).

[4] Note that Prinz (2006) is a cognitivist. He holds that moral judgements are sentiments, and sentiments are dispositions to have emotions. But he also holds that they are truth-apt judgements that objects have the property of causing those emotions in us. Greene (2003, 850) apparently takes a noncognitivist view.

[5] It would also seem to apply to evaluative facts about beliefs. The evolutionary explanation of our capacity for beliefs does not proceed independently of the truth of those beliefs: it is our capacity for correctly representing the way the world independently is that is fitness-enhancing. However, this explanation does not seem to require any facts about the *goodness* of true beliefs.

[6] Note that views of either of these kinds need not be aiming to account for all moral dispositions—dispositions to connect rightness and wrongness with conceptions of purity and impurity, for example—as instances of correct judgement, any more than we need all observational beliefs to be true in order to make sense of the idea that many are.

[7] This time, I cannot see how to make plausible a non-naturalistic strategy of this type. If you did not think that our moral experience was itself evidence for the existence of non-natural moral facts, what other grounds could you have for thinking that there are any?

[8] If this is to be answered convincingly, we have to add to Rawls. For Rawls, a considered judgement is one which the person making it judges to have been made in circumstances

conducive to accuracy (Rawls 1971, 47–48). But what is required is not just that I do think that my judgement has been made in circumstances conducive to accuracy, but that I am *warranted* in thinking this.

9. It *need* not supply further evidence; but it might do so. Maybe my track record is better when my social judgements are accompanied by emotions. The point is that it need not be.

10. I am not taking myself to have proved here that it could not be true that all of our judgements about reasons are unreliable. The claim is that we could not be warranted in believing this. For a fuller discussion of this topic, see Foley (1998).

11. I describe my preferred account in Cullity (1999).

12. On this view, we should reject a sharp separation between emotions and the recognition of reasons. Notice that this is compatible with thinking that emotions are rarely caused by episodes of *reasoning* that involve conscious tokening of propositions and inferences between them. Just as I hold beliefs about the structure of the computer on my desk at times when I am not thinking conscious thoughts about it, so too can I believe that someone is distressed, and that I have a reason to help him, without tokening a sentence in my head to that effect. Reason-recognition need not involve reasoning.

REFERENCES

ANDERSON, S. R., A. BECHARA, H. DAMASIO, D. TRANEL, and A. R. DAMASIO. 1999. Impairment of social and moral behaviour related to early damage in human prefrontal cortex. *Nature Neuroscience* 2: 1032–37.

BARGH, J. 1994. The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In *Handbook of social cognition*, edited by J. R. S. Wyer and T. K. Skrull. 2nd ed. Hillsdale, N.J.: Erlbaum.

BLACKBURN, S. 1998. *Ruling passions*. Oxford: Clarendon Press.

BLAIR, JAMES, A. A. MARSH, E. FINGER, K. S. BLAIR, and J. LUO. 2006. Neuro-cognitive systems involved in morality. *Philosophical Explorations* 9 (1): 13–27.

BLAIR, R. J. R. 1997. Moral reasoning in the child with psychopathic tendencies. *Personality and Individual Differences* 22: 731–39.

CULLITY, GARRETT. 1998. Moral judgement. In *The Routledge encyclopedia of philosophy*, edited by Edward Craig. London: Routledge.

——. 1999. Virtue ethics, theory, and warrant. *Ethical Theory and Moral Practice* 2: 277–94.

DAMASIO, A. R., D. TRANEL, and H. DAMASIO. 1990. Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research* 41: 81–94.

DANCY, JONATHAN. 1993. *Moral reasons*. Oxford: Blackwell.

——. 2004. *Ethics without principles*. Oxford: Clarendon Press.

FINE, CORDELIA. 2006. Is the emotional dog wagging its rational tail, or chasing it? Unleashing reason in Haidt's social intuitionist model of moral judgment. *Philosophical Explorations* 9 (1): 83–98.

FOLEY, RICHARD. 1998. Rationality and intellectual self-trust. In *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry*, edited by Michael R. DePaul and William Ramsey. Lanham, Md.: Rowman & Littlefield.

GIBBARD, ALLAN. 1990. *Wise choices, apt feelings*. Oxford: Clarendon Press.

GREENE, JOSHUA. 2003. From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience* 4: 847–51.

GREENE, JOSHUA, and JONATHAN HAIDT. 2002. How (and where) does moral judgment work? *Trends in Cognitive Sciences* 6: 517–23.

GREENE, J. D., R. B. SOMMERVILLE, L. E. NYSTROM, J. M. DARLEY, and J. D. COHEN. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–8.

HAIDT, JONATHAN. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108: 814–34.

HAIDT, J., S. KOLLER, and M. DIAS. 1993. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 65: 613–28.

HAUSER, M. D., L. YOUNG, and F. A. CUSHMAN. Forthcoming. Reviving Rawls' linguistic analogy: Operative principles and the causal-intentional aspects of moral actions. In *Moral psychology*, edited by Walter Sinnott-Armstrong. New York: Oxford University Press.

JOYCE, RICHARD. 2006. Metaethics and the empirical sciences. *Philosophical Explorations* 9 (1): 133–48.

KENNETT, JEANETTE. 2006. Do psychopaths really threaten moral rationalism? *Philosophical Explorations* 9 (1): 69–82.

MOLL, J., R. DE OLIVEIRRA-SOUZA, and P. J. ESLINGER. 2003. Morals and the human brain: A working model. *Neuroreport* 14: 299–305.

NICHOLS, SHAUN. 2002. How psychopaths threaten moral rationalism: Is it irrational to be amoral? *The Monist* 85: 285–304.

——. 2004. *Sentimental rules: On the natural foundations of moral judgment*. New York: Oxford University Press.

PRINZ, JESSE. 2006. The emotional basis of moral judgments. *Philosophical Explorations* 9 (1): 29–43.

RAILTON, PETER. 2000. Darwinian building blocks. *Journal of Consciousness Studies* 7 (1/2): 55–60.

RAMACHANDRAN, VILAYANUR. 2003. *The emerging mind: The Reith Lectures 2003*. London: Profile.

RAWLS, J. 1971. *A theory of justice*. Cambridge, Mass.: Harvard University Press.

ROSS, W. D. 1930. *The right and the good*. Oxford: Clarendon Press.

SACKS, OLIVER. 1985. *The man who mistook his wife for a hat*. London: Duckworth.

SINNOTT-ARMSTRONG, WALTER. Forthcoming. Moral intuitionism meets empirical psychology. In *Metaethics after Moore*, edited by Terry Horgan and Mark Timmons. New York: Oxford University Press.

SOBER, E., and D. S. WILSON, 1998. *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, Mass.: Harvard University Press.

TRIVERS, ROBERT. 1985. *Social evolution*. Menlo Park, Calif.: Benjamin Cummins.

TVERSKY, AMOS, and DANIEL KAHNEMAN. 1981. The framing of decisions and the psychology of choice. *Science* 211: 453–58.

WHEATLEY, T., and J. HAIDT. 2005. Hypnotically-induced disgust makes moral judgements more severe. *Psychological Science* 16: 780–84.

WILLIAMS, BERNARD. 1985. *Ethics and the limits of philosophy*. London: Collins.

ZAJONC, R. B. 1980. Feeling and thinking: Preferences need no inferences. *American Psychologist* 35: 151–75.

Garrett Cullity, Department of Philosophy, The University of Adelaide, Adelaide, SA 5005, Australia. E-mail: garrett.cullity@adelaide.edu.au