Irrealism and the genealogy of morals

Richard Joyce

Penultimate version of the article appearing in *Ratio* 26 (2013) 351-372

Abstract

Facts about the evolutionary origins of morality may have some kind of undermining effect on morality, yet the arguments that advocate this view are varied not only in their strategies but in their conclusions. The most promising such argument is modest: it attempts to shift the burden of proof in the service of an epistemological conclusion. This paper principally focuses on two other debunking arguments. First, I outline the prospects of trying to establish an error theory on genealogical grounds. Second, I discuss how a debunking strategy can work even under the assumption that noncognitivism is true.

1. Introduction to moral debunking arguments

A genealogical debunking argument of morality takes data about the origin of moral thinking and uses them to undermine morality. The genealogy could be ontogenetic (like Freud's) or socio-historical (like Nietzsche's or Marx's), but the focus of recent attention has been the evolutionary perspective. 'Debunking' and 'undermining' are intentionally broad terms, designed to accommodate a number of different strategies and conclusions. Sharon Street's debunking argument, for example, aims to overthrow moral realism, while leaving intact the possibility of non-objective moral facts (e.g., those recognized by a constructivist) (Street 2006). Michael Ruse's earlier debunking argument often looks like it has the same aim as Street's, though on occasions he appears to try for a stronger conclusion: that all moral judgements are false (Ruse 1986, 2006, 2009). My own debunking argument has an epistemological conclusion: that all moral judgements are unjustified (Joyce 2006, 2014).

Calling all of these conclusions instances of 'debunking' is, in some sense, prejudicial. The rejection of moral realism, for example, counts as a *debunking* of morality only if one thinks that realism is somehow the natural interpretation of morality—and that is far from obvious. Any act of debunking is at the same time a *vindication* of something. For example, to show that all moral judgements are false would be to vindicate the error theoretic metaethical view. But I'll let this pass, and allow 'debunking' to remain as a usefully vague intuitive term for these arguments.

What these disparate arguments often share is a presupposition of cognitivism. Moral judgements can be all false only if moral judgements are the kind of thing that can have truth value. Moral judgements can be all unjustified (in an epistemic sense) only if moral judgements are beliefs. In other words, the noncognitivist—who holds that moral judgements (as mental states) are not beliefs and (as speech acts) are not assertions—will survey the debate over these debunking arguments with an unperturbed air.[1]

[1] While Street's argument does not presuppose cognitivism, at the same time noncognitivists needn't be worried by the argument, since for them the refutation of moral realism will be entirely welcome.

As already mentioned, the debunking argument that I have advocated (and thus, obviously, think most promising) is one with an epistemological conclusion. It is not my intention to defend or elaborate this argument further on this occasion, though it is probably best if I rehearse it briefly in order to provide a comparison class. Rather, in this paper I want to explore two different debunking avenues. First, I shall investigate what the prospects are for a debunking argument that aims to establish a moral error theory. Second, I shall question whether the noncognitivist is warranted in his/her complacency; perhaps a debunking argument against noncognitivism could be developed. My objectives are diagnostic rather than promotional, thus my conclusions regarding both these avenues will be non-committal. Given this, beginning with a brief look at a kind of debunking argument that I think likely to succeed will provide a useful backdrop.

2. Epistemological debunking

Recent years have seen a burgeoning of discussion about the evolutionary origins of the human moral faculty.[2] Part of any such nativist explanation must be an account of what it was about moral thinking that served the reproductive purposes of our ancestors. On this point hypotheses diverge, but on most accounts moral thinking was advantageous because it in some manner enhanced their cooperative tendencies. What is striking about these nativist hypotheses is that they seem entirely compatible with the error theoretic stance; they do not appear to imply or presuppose that any of our ancestors' moral judgements were *true*.

This is not so of evolutionary explanations of any kind of judgement. For example, humans quite possibly have an adaptive mechanism for distinguishing faces from other visual stimuli. But if one were to be (bizarrely!) an error theorist about faces, then the evolutionary explanation for why it might have useful for our ancestors to have this mechanism would surely fizzle. By contrast, the evolutionary hypothesis that moral thinking emerged because it strengthened social cohesion is no less plausible for the error theorist than anyone else. The best explanation of the face-identifying adaptation classifies it as a *truth-tracking* mechanism; the best explanation of the moral faculty does not classify it as a truth-tracking mechanism. This, it would seem, has epistemological consequences.

Most epistemological theories (and, I am tempted to add dogmatically, all sensible epistemological theories) hold that a belief's being justified depends on its standing in one or other specific relationship to the fact that it represents. To discover that a belief does not stand in this relation to the relevant fact is to discover that the belief lacks justification. (Whether it shows that the belief has lost its justification, or shows that it was never justified in the first place, depends on which family of epistemological theories one favours.) If the evidence were to come down in support of moral nativism, then this would seem to be confirmation that our

[2] See Alexander 1987; Irons 1996; Krebs 2005; Nichols 2005; Dwyer 2006; Machery & Mallon 2010; Mikhail 2011; Kitcher 2011.

moral beliefs have their origins in a process that is not designed for truth-tracking.[3] Note that this would not be a matter of conjuring up a far-fetched unfalsifiable skeptical hypothesis according to which our moral beliefs are bogus (like Descartes' demon); it would be the confirmation of an empirical hypothesis that appears compatible with the systematic falsehood of moral judgements. Such a confirmation, I claim, undermines the epistemic standing of moral judgements.

Justification, of course, is a relative affair. My belief that *p* may be justified while your belief that *p* is not. Perhaps at an earlier time my belief that *p* was also unjustified; perhaps in the future it will become unjustified again (if, say, I ignore mounting evidence against the belief). Thus the conclusion that all moral beliefs are unjustified should not be interpreted as making a stronger claim than is reasonable. The proposition that a belief is unjustified does not exclude the possibility that justification can be attained or reinstated in the future. The force of the epistemological debunking argument is to issue a challenge, to shift a burden of proof.

It is often claimed that the fact that skepticism (about any object of everyday belief) cannot be refuted does not thrust that skeptical stance upon believers, so long as the non-skeptical position also cannot be refuted. Thus it is claimed that the skeptic shoulders a burden of proof: it is not enough to make skepticism irrefutable, the skeptic needs positive arguments against belief. In the event that neither the skeptical nor non-skeptical position is refutable, the non-skeptic can happily carry on with his or her everyday beliefs.

Moral nativism promises to upset this picture by providing a new hypothesis about the place of moral judgements in the world (one, moreover, potentially with empirical backing). Even those who were confident that their moral beliefs are true cannot ignore the evolutionary debunking argument, inasmuch as it is incumbent upon them either to establish that the nativist hypothesis is false or to demonstrate that moral beliefs are true even according to that hypothesis. Either way, they have some work to do. To maintain confidence in moral beliefs in advance of this work is epistemically negligent; any principle that allows one to do so is gullibility dressed up as a methodology.

When I presented this argument on an earlier occasion, I made the rash decision to label it an error theoretic conclusion (Joyce 2006, p.223). I did this via suggesting that the label 'error theory' might denote a disjunction of metaethical positions: either the view that all moral judgements are false or the view that all moral judgements are unjustified. I now recant this suggestion for the following reason. Suppose all moral judgements are unjustified. This is consistent with moral judgements being true, and, moreover, objectively true; thus the claim that all moral judgements are unjustified is compatible with moral realism. But the error theory had better not be compatible with moral realism, therefore the view that all moral judgements are

[3] Note that 'truth-tracking' can be understood epistemically or evolutionarily. The latter refers to what a psychological faculty is *supposed to do* (in evolutionary terms). The former is often taken to refer to a covariation between a belief and the fact that it represents. In fact, I think epistemic truth tracking is quite difficult to spell out, and the covariation analysis runs into difficulties when beliefs concern necessary truths and necessary falsehoods. See Joyce 2014 for discussion.

unjustified had better not be sufficient for an error theory.[4] It is preferable to keep our metaethical theories separate and be clear that the conclusion to this debunking argument is epistemological in nature. The thesis that all moral judgements are unjustified lacks a label, though it is perfectly acceptable to call it a version of moral skepticism.[5]

I think the epistemological debunking argument outlined in this section has legs. But the benefits of establishing the error theory by stretching the extension of the label in the manner just described (and just renounced) are, to quote Russell, the advantages of theft. I turn now to exploring the prospects of using a debunking argument to establish the moral error theory through honest toil.

3. Error theoretic debunking

Certainly there are circumstances where learning about the origin of a belief can reveal that belief to be false. My belief that hypnosis cannot instil genuine beliefs in people is falsified if I discover that I was caused to have this belief through hypnosis. But clearly nothing so swift and sneaky as this is going to work in the case of moral judgements and moral nativism. The moral judgement that promise-breaking is wrong, say, simply doesn't imply anything about its own origins in the way that the belief about the limits of hypnosis does. Rather, we shall see, the error theoretic debunking argument depends on a principle of parsimony.

Let us start with Street's debunking argument, whose conclusion is that moral realism is probably false. She argues that the moral realist, confronted with the truth of moral nativism (we are imagining), faces a dilemma concerning the relation between our moral judgements (products of the distortions and contingencies of our evolutionary ancestry) and the supposed realm of objective moral facts. On the one hand, if there is no relation then it would be an astonishing coincidence if many of our moral judgements were even approximately true—a conclusion supposedly disagreeable to the realist. The problem with the other horn of the dilemma is that it is, according to Street, empirically dubious. I have already noted that the usual nativist hypotheses see the ancestral adaptive pay-off of having a moral faculty in terms of enhancing certain cooperative tendencies, not in terms of tracking moral truths. Street thinks this 'adaptive link hypothesis' is superior to any truth-tracking hypothesis for three reasons: it is more parsimonious, it is clearer, and it is more illuminating of the phenomenon it seeks to explain (2006, p.129). Street's irrealist conclusion might be put as follows: 'There are no objective moral facts.' Yet she doesn't deny the possibility of moral facts—they will simply be of a constructivist nature.

[4] Analogy: Ancient Greek atomists didn't have any real evidence in favour of their view; it is not unreasonable to claim that they lacked justification for their beliefs. Yet it would seem weird to be an error theorist about their atom discourse. After all, broadly speaking they got it right!

[5] Academic skepticism about morality is the denial that moral knowledge exists. If knowledge requires both truth and justification, then the error theorist's denial of truth counts as skepticism, as does the epistemological denial of justification. (See Sinnott-Armstrong 2006.)

What good, one might ask, is this to an error theorist? Let me approach this by quickly comparing Ruse's argument. Ruse maintains that being imbued with a kind of objectivity is the whole point of moral thinking, evolutionarily speaking. Morality serves its adaptive function of strengthening our motivation to cooperate by seeming to be imbued with a kind of inescapable external prescriptivity. 'It is precisely because we think that morality is more than mere subjective desires that we are led to obey it' (Ruse 1986, p.103). But, Ruse argues, this objectivity is an adaptive illusion. He argues for this latter claim via an implicit appeal to parsimony: once we have explained why morality *seems* to be objective, there is simply no call for any further explaining in terms of positing a realm of objective moral facts. At this point the conclusion to Ruse's argument looks very similar to that of Street's, reached by somewhat different means. He writes: '[M]orality 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, p.253).

However, Ruse's discussion contains elements that aren't present in Street's thinking, opening the door to the stronger error theoretic conclusion. For a start, his emphasis on the adaptive importance of the *objectivity* with which moral prescriptions are infused is not something Street mentions. A strong thread running through his argument is that moral realism is written into the phenomenology of moral experience. But he goes further, apparently moving from phenomenology to semantics: 'Ethics is subjective, but its meaning is objective' (Ruse 2006, p.22); '[W]hat I want to suggest is that…the *meaning* of morality is that it is objective' (Ruse 2009, p.507). The move from phenomenology to semantics is not something to which one can help oneself for free, but at the same time it's not unreasonable to assume that the *meaning* of a term is going to reflect our experience of the phenomena denoted by that term. If humans are designed by natural selection to experience morality as objective, then this perhaps makes more plausible the already not-ridiculous thesis that objectivity is an essential quality of morality, conceptually speaking. With this thesis operating as a bridging premise, one can get from the subconclusion 'There are no objective moral facts' to the conclusion 'There are no moral facts.' (The two propositions would stand in the same relation as 'There are no four-sided squares in the box' and 'There are no squares in the box' stand in.)

This bridging premise is a key part of this error theoretic debunking argument. Street rejects it, hence her conclusion is not error theoretic. And of course it is an extremely controversial thesis, over which much metaethical ink has been spilt. Part of the problem is that the term 'objectivity' is not well defined, and it gets used differently in different areas of philosophy. (For discussion see Joyce 2007a, 2009.) The notion that Ruse seems to have in mind is that of moral prescriptions having a kind of external authority: we feel bound to follow them because we experience them as not of our own making (unlike, say, the non-objective prescriptions of fashion).[6] Many philosophers will agree with Ruse that we tend to experience moral norms in this manner, though only some of them (a good number, to be sure) will go along with the stronger claim that this kind of objectivity is *essential* to morality, such that a normative

[6] This appears to be how Maurice Mandelbaum (1956, p.50) uses the term 'objectivity.'

framework stripped of this objectivity wouldn't even count as a 'moral' system. Those that do support the stronger semantic claim include both realists (who think that this objectivity can be satisfied) and irrealists (who think that it cannot be satisfied).

Ignoring, for a moment, the difficulty of establishing this bridging premise, let me try to reconstruct the argument that employs it. Whether this actually reflects Ruse's reasoning is not my primary concern, but I will continue to attribute it to him if only for the sake of argument. The argument turns on the application of a parsimony principle:

1. Objective moral facts aren't required to explain anything.

2. If some type of fact plays no explanatory role, then this is ground for disbelieving in this type of fact.

There are deep questions to be raised about both these premises, which I shall turn to in a moment, but initially I want to discuss them just sufficiently to motivate the need for a third premise.

In a sense, nothing is required to explain anything. What I mean by this quizzical claim is that one always has choices in how to explain any phenomenon. If the cat knocks over the vase, one can always explain the broken vase without employing the concept *cat*. Instead of using biological or zoological categories, one could (in principle) make reference to a conglomeration of organic chemicals moving about the room, or a swarm of particles and energy. Thus the concept *cat* isn't required in any explanation of anything. But this hardly means that cats are explanatorily impotent. The crucial point is that cats are reducible to entities that are described at other theoretical levels: chemistry or physics, for example. Thus, even if it were true that reference to objective moral facts isn't needed to explain anything, it wouldn't follow that objective moral facts are explanatorily impotent. For this conclusion a further premise must be added:

3. Objective moral facts aren't reducible to any facts that do have explanatory roles.

These premises yield the sub-conclusion:

4. Therefore, there is ground for disbelieving in objective moral facts (i.e., there is ground for rejecting moral realism).

We can now add the bridging premise:

5. Morality is essentially objective.

And the error theoretic conclusion follows:

6. Therefore, there is ground for disbelieving in moral facts.

Every single one of the premises is problematic. Let us start by considering premises 1 and 3 together. Ruse's argument for premise 1 is often presented via an analogy (Ruse 1986, pp.256-7, 2006, pp.22-23, 2009, pp.504-505). He refers to the spike of interest in séances in Europe in the aftermath of World War 1. Imagine a grief-stricken mother attending such a séance, during which time she comes to believe that her dead son has spoken to her from beyond the grave. We can explain everything that needs explaining about this belief by reference to psychological and sociological factors; there is no need to suppose that the belief might be *true*. Similarly (Ruse thinks), moral nativism explains everything that needs explaining about why humans judge certain actions to have objective moral status; there is no need to suppose that these judgements might be *true*.

The weakness of the analogy is brought out when we attend to premise 3. In order to suppose that the mother's belief is true, we would have to presume that the world contains supernatural forces, post-mortem consciousness, ghosts, etc.—that is, some pretty spooky ontology. It is far from obvious that this is what is required to suppose that judgements about objective morality are true. Moral naturalists (of an objectivist stripe) will often identify moral properties with naturalistic properties that we already accept in our ontological scheme. A utilitarian, for example, may identify moral goodness with happiness.[7] By contrast, any attempt to identify, say, *ghosts* with some cluster of naturalistic properties looks hopeless. In other words, the analogue of premise 3 for ghosts looks obviously true. But premise 3 as it stands for objective moral properties will be doubted by many, and therefore cannot stand without argumentative support.

Rather than return attention to the bridging premise 5, let us consider dropping all mention of objectivity, which would allow premises 5 and 6 to evaporate. The revised argument is as follows:

1*. Moral facts aren't required to explain anything.

2. If some type of fact plays no explanatory role, then this is ground for disbelieving in this type of fact.

3*. Moral facts aren't reducible to any facts that do have explanatory roles.

4*. Therefore, there is ground for disbelieving in moral facts.

[7] One may wonder what is objective about something so obviously mind-dependent as *happiness*. But this misidentifies the point. The question is whether the relational proposition 'Goodness = happiness' is true objectively (like 'Water = H_2O') or true in virtue of some human decision. (See Shafer-Landau 2007, pp.157-158.)

The stripped down argument looks a lot like one that Gilbert Harman famously uses to frame his discussion (1977). Harman doesn't endorse the argument, though; he rejects premise 3*, arguing that moral facts are reducible to facts about what reasons we have for acting, which (he thinks), properly understood, are empirical phenomena. Nor does Harman place any emphasis on moral nativism, which for Ruse is the main consideration lying behind the first premise. Harman, rather, appeals to developmental factors to explain how moral judgements might arise from non-truth-tracking mechanisms. This difference doesn't matter to our current concerns; what is significant is that moral judgements can be genealogically explained in a way that makes no reference to their being true. This supports the first premise presumably in the following manner. If moral judgements can be fully explained without reference to moral facts, then this casts immediate doubt on whether moral facts are needed to explain *anything*. (Likewise for Ruse, mutatis mutandis, concerning *objective* moral facts.) It seems to me that this move is reasonable, for what possible instance would we recognize of a moral fact playing a role in explaining phenomenon X, where this act of recognition did not involve the use of a moral judgement? Moral facts appear to have what Crispin Wright calls 'narrow cosmological role' (1992): their causal impact always involves someone's having made a judgement concerning their presence. (Cats, by contrast, have wide cosmological role, affecting the world in a myriad of judgement-independent ways: meowing, casting shadows, producing kittens, knocking over vases.) If moral explanations (e.g., 'Fred broke the promise because he's wicked') always depend on someone's having made a moral judgement, but moral judgements can always be fully explained without reference to moral facts, then the explanatory potency of moral explanantia (e.g., Fred's wickedness) is an illusion.

Whether moral facts can be reduced to facts that do have explanatory role—as 3* denies (but Harman affirms)—is a matter I don't have space to address here. Ruse (so far as I know) doesn't explicitly argue in favour of premise 3, but I have already noted that its lack of support makes the argument that I'm attributing to him vulnerable. The general format of a defence of premise 3* would be to identify some indispensable feature of moral facts that no naturalistic facts can satisfy. (I am here assuming that facts with explanatory roles must be naturalistic facts.) There are a number of promising contenders for this 'indispensable feature,' the obvious one being something to do with the categorical practical authority (the 'must-be-doneness') of moral facts. Harman reduces moral facts to facts about reasons, and thereby, arguably, satisfies a desideratum of *practical authority*—for what could have more practical authority for a person than her *reasons* for acting? On the other hand, however, Harman thinks that the only viable account of reasons is one that renders them relativistic. Yet one may argue that some quality of absolutism is an 'indispensable feature' of moral facts, and if this is correct then Harman's attempt to overthrow premise 3* must be rejected.

I don't propose to spend more time assessing the third premise, for it is premise 2 that should really be occupying our attention in evaluating the error theoretic debunking argument.

Harman's presentation of the argument does not explicitly endorse premise 2. Summing up his argument (before embarking on his rejection of 3*) he writes that 'it remains problematic whether we have any reason to suppose that there are any moral facts' (1977, p.23). Imagine it turns out that we do not have any reason to suppose that there are any moral facts. This wouldn't automatically amount to our having a reason to suppose that there are *not* any moral facts. The crucial difference is between premise 2 and the weaker 2B:

2. If some type of fact plays no explanatory role, then this is ground for disbelieving in this type of fact.

2B. If some type of fact plays no explanatory role, then we have no ground for believing in this type of fact.

Premise 2B is more plausible than 2, but it is premise 2 that's required to secure the error theoretic conclusion. Premise 2B, by comparison, looks like it will feed into an epistemological debunking argument. One cannot derive 2 from 2B without violating the adage 'Absence of evidence is not evidence of absence.'

But the adage is not to be taken as gospel, for there are certainly circumstances where absence of evidence *is* evidence of absence: most obviously, conditions in which one could reasonably expect to have evidence (see Sober 2009a, p.64). For example, if there were a leopard hiding in this room somewhere, it would be reasonable for me to expect to encounter some evidence of the fact; the absence of any such evidence provides evidence of a leopard's absence.

The key question, then, is whether these kinds of circumstances obtain for the case of moral facts. Is it reasonable for one to expect that if there were moral facts we would have evidence of them? I find this a very difficult question to answer, and I suspect that different philosophers will give different reactions. There does seem to be something unsettling about the idea of a realm of moral facts for which we have no evidence at all, such that our actual moral judgements might be, for all we know, wildly mistaken. Such an idea is a corollary of an ultra-realist conception of morality, and yet I suspect it is one at which even most so-called realists will balk. (Recall that this was one of the horns of Street's dilemma against the realist.)

Similarly, if we had some independent information about the probability of there being moral facts, then we might be able to support the stronger conclusion. Suppose we knew that moral facts were improbable, but took our moral judgements nevertheless to provide some support for their obtaining. The discovery that these moral judgements stem from a non-truth-tracking source would undermine this support, thus putting us back in the position of judging moral facts improbable. (This is not exactly *disbelief*, of course, yet framing the issue in Bayesian terms of degrees of belief is probably how the more nuanced presentation should proceed.[8]) Yet assessing the prior probability of moral facts obtaining is also a very difficult question regarding which there will be nothing remotely like a consensus among philosophers (see Brosnan 2011,

[8] See Sober 2009b, p.129.

p.55). So this route seems even less propitious for the error theorist than that sketched in the previous paragraph.

A more promising way of supporting the stronger premise 2 is via the endorsement of some methodological principle that underwrites it. Methodological empiricism, for example, will typically demand the *banishment* of any putative entity that fails to connect appropriately with perceptual input. Empiricism will often urge disbelief, rather than the withholding of belief, for any item that fails the test. (Recall Hume's directive that any book that doesn't pass empiricist muster must be 'committed to the flames.') Even without specifying any particular version of empiricism, we can be confident that explanatory impotence will count as a failure, since such impotence implies an inability to figure in any perceptual process.

This last route seems to me the most plausible way of defending premise 2, though on the face of it seems rather dogmatic: basically, one just embraces a methodological principle that demands (or at least permits) disbelief in explanatorily impotent entities. Presumably, though, the air of dogmatism may be dispelled by sensible considerations in favour of the methodology. The traditional school of empiricism, for example, wasn't based on a doctrinaire whim; its precepts were adopted for credible reasons. Whether premise 2 is plausible, then, will depend on an assessment of the considerations for and against the broader methodology that underwrites it.

Even if premise 2 is defensible, however, we have seen that there are many other 'if's in an error theoretic debunking argument of this sort, and the argument strays a long way from the genealogy of morals with which it began. Ultimately, moral nativism may find a place as a premise in an error theoretic debunking argument, but it will be a supporting role; the main actors will be propositions of a metaethical nature.[9]

4. Noncognitivist debunking

The two styles of debunking argument thus far discussed—epistemological and error theoretic—presuppose metaethical cognitivism: moral judgements can be deemed epistemically unjustified or deemed false only if they are the kind of thing that can have truth value. Rejecting this presupposition, it would therefore seem, is a way of sidestepping the whole debunking dialectic. But perhaps a similar kind of debunking challenge can be devised for the noncognitivist?

Simon Blackburn's quasi-realist project takes an irrealist ontology, a noncognitivist construal of moral judgements (according to which they express conative attitudes), and from this basis endeavours to earn the right to the trappings of realism: talk of beliefs, truth, assertions, facts, etc. (Blackburn 1984, 1993). It is difficult to integrate quasi-realism into many metaethical debates. Should it be assessed as an irrealist noncognitivist thesis, or as a position that supports

[9] The tentative attitude expressed here toward a genealogical debunking argument in favor of a moral error theory must not be mistaken for a tentative attitude toward the conclusion. I stand by the error theoretic metaethical position for which I have argued on other occasions (Joyce 2001, 2007b, 2011); the question under current scrutiny is whether genealogical considerations can be used to establish that view.

moral truths, beliefs, properties, etc.? In the present context, what needs to be noted is that if the quasi-realist program succeeds in vindicating talk of moral properties, beliefs, and truths, then, to whatever extent the epistemological and error theoretic debunking arguments work, they will apply to quasi-realist noncognitivism. I propose, then, to put the quasi-realist program to one side and work with a very simple and old-fashioned version of noncognitivism, according to which moral judgements as mental states are of a purely affective kind, and moral judgements as speech acts function solely to express those states. Let the states be simply some special form of *liking* and *disliking*. (I say this in order to exclude complications that would arise from treating noncognitivism as the view that moral judgements express *emotions*. The complication is that many emotions are mixtures of affective and cognitive components, and thus the epistemological or error theoretic arguments could apply to the cognitive elements.) According to this view, there are no moral truths, no moral beliefs, no moral properties, no moral assertions, no moral knowledge.

Even if affective states cannot be false or epistemically unjustified, they can be mistaken in various other ways. Hume allows two ways for passions to be ‘contrary to reason’: first, when based on a false belief about something’s existence; second, when based on a false belief about what means are necessary and sufficient to satisfy some desire (Hume [1740] 1978, p.416). It’s not obvious what Hume means by ‘contrary to reason’ in this context; he doesn’t necessarily mean that these are the only two ways in which passions can be normatively appraised (see Schafer 2008). But even if he were to mean this, he is pretty clearly mistaken. If a passion is based on a belief that is not merely false but *irrational* (in the sense, perhaps, of being maintained irresponsibly in the face of discrediting evidence) then presumably the passion inherits a more serious kind of wrongness.[10] The taphephobe suffers from an irrational fear of being buried alive, but it is plausible that this fear is based on an irrational (and not merely false) belief concerning the likelihood of this occurring. Often phobic fears are irrational in another sense: because the fear is had in the absence of appropriate beliefs. An arachnophobe feels fear in the presence of a harmless spider, while knowing that it is harmless. I might like someone while believing (sincerely and truly) that she has all the qualities that I despise in a person and no redeeming features. Here it would be completely natural to assess my liking as ‘bizarre’ and ‘irrational.’

Hume will be quick to point out that in all these cases it is not the passion *per se* that is at fault, but rather that its error derives from its relation to belief: the passion either stands in the wrong relation to a good belief, or stands in the right relation to a faulty belief. ‘[P]assions can be contrary to reason only so far as they are *accompany’d* with some judgement or opinion’ (Hume [1740] 1978, p.416). But it seems that affective states may also be subject to criticism without reference to beliefs. Consider our tendency to call imprudent attitudes ‘unreasonable.’ One’s liking of something may cause one harm. (In such a case, one might dislike one’s liking of the thing. Or one might not: one might like one’s self-harming liking, which may well bring

[10] And of course irrational beliefs are not a subset of false beliefs. An irrational belief may be true.

one further harm.)[11] It might be thought that imprudent passions are a special case of having false beliefs about the best means to satisfy one's desires—but this is plausible only on the assumption that people must always desire their own flourishing. Yet even when it is recognized that a person has sincere self-destructive devil-may-care desires, we do not cease to call his/her self-sabotaging actions and passions 'imprudent.'

Whether Hume allows this last category of evaluation isn't really my concern. That he does not is the tempting conclusion to draw from his memorable declaration: ''Tis as little contrary to reason to prefer even my own acknowledge'd lesser good to my greater, and have a more ardent affection for the former than the latter' (ibid.). But arguably even here Hume is making a claim about what the faculty of reason is capable of accomplishing, not placing a restriction on how imprudent preferences may be normatively assessed (see Schafer 2008). In any case, imprudent affective states *are* typically called 'unreasonable' and 'irrational,' and the revelation that we are dealing with someone so aberrant as to consciously *prefer* his 'lesser good to his greater' does not force us to retract the criticism. Perhaps there is nothing that could be said to such a person to change his mind; perhaps we'll go so far as to say that if these are really his preferences then he has no reason to refrain from pursuing the lesser good. Never mind; we can still legitimately criticize his preference as 'unreasonable.'

Not only are affective states subject to various kinds of criticism, but genealogical considerations will frequently form the basis for the criticism. Just as we cast doubt on someone's belief with the vernacular 'You only believe that because...,' so too do we disparage someone's attitudes by saying 'You only feel that because....'A person's irritation may be dismissed by observing that she is tired. A person's preference for a musical performance is discounted on the ground that the performer is his daughter. A feeling of disgust will be convicted of some kind of misfiring if it is revealed that it was prompted by hypnosis.

It is not immediately evident precisely what these verbs of 'dismissal,' 'discounting,' and 'being convicted of misfiring' denote.[12] Consider the last example just mentioned. Psychologists Thalia Wheatley and Jonathan Haidt (2005) hypnotized subjects to feel a pang of disgust upon hearing a given mundane word, like 'often' or 'take.' The subjects were then presented with vignettes and asked to morally assess one of the characters therein, named 'Dan.' Those who had been hypnotized and given the trigger word were much more inclined to assess Dan negatively, even when no form of transgression had been described. Upon being questioned, they confabulated grounds for their condemnation, or simply said things like 'It just seems so weird and disgusting,' or 'I don't know [why it's wrong], it just is.'

It is quite clear that we will dismiss disgust that has been prompted in this fashion. Perhaps we dismiss the disgust because of faulty beliefs that the subject holds: when she experiences disgust upon reading about Dan's actions, she might be inclined to 'trust' her negative arousal to be a response to something disgust-worthy. And this is where she has gone wrong, because,

[11] An affective state may also be criticized for the harm it brings to others, but since this kind of moral assessment is the very topic that is under scrutiny, it is best put aside.
[12] Cf. Daniel Kelly's comment on the use of the word 'problematic' (Kelly 2014).

unbeknownst to her, her negative arousal is definitely a response to a mundane word and not anything to do with Dan. Her belief that Dan's actions warrant disgust is mistaken, her belief that her emotions are a response to something that merits them is mistaken, and thus we dismiss her disgust and, further, dismiss the associated moral judgement (once we realize that the disgust is causally responsible for it). If noncognitivism is true, then the subject's moral judgement just is an expression of that disgust—or at least an expression of the dislike that the disgust prompts. The situation presented by Wheatley and Haidt would appear, then, to be a clear-cut case of genealogical debunking, even for a noncognitivist.

It might be useful to compare this with another case. Suppose you are slipped a pill that gives you a headache. It doesn't seem in this case that your headache, despite its unusual and secret origin, is (or can be) in any sense 'dismissible.' Generally one doesn't have very specific beliefs about the cause of a routine headache—perhaps a suspicion (a probabilistic belief?) that it's caused by dehydration. But in any case there is no belief about the headache's being *warranted* by its cause. Even when one firmly believes that dehydration has caused a headache, issues of *merit* just aren't apposite. Disgust is different in this respect. Disgust (unlike headaches) is an emotion, and this implies that disgust is more than just an affective state—it also involves or is accompanied by cognitive thoughts (e.g., concerning affective states being merited by certain events). This, it would seem, is what makes the difference, allowing the emotion—including the affective component of the emotion—to be in some circumstances dismissed on grounds of its genealogy.

Talk of 'dismissal' remains vague. I am certainly not saying anything about how we should treat a person whose emotions and moral judgements have been manipulated by, say, hypnosis. There may be various reasons for not pointing out to her what has happened. In the same way, were one to be a moral error theorist and hold that most people have false beliefs about morality, nothing obvious follows about how one should treat them. (Atheists generally don't go around knocking on theists' doors and teasing them.) The key point is that 'dismissal' amounts to some kind of unmistakable albeit vague *undermining*—and this is sufficient for my present purposes.

Could the kind of unusual and local effect generated in the Wheatley and Haidt experiment scale up to a more ubiquitous debunking argument? Daniel Kelly thinks so (2011, 2014). He argues that the human disgust response evolved as an adaptive mechanism for dealing with the twin threats of toxins and parasites; this psychological response was then co-opted for negotiating social norms (which would explain the connection between disgust and moral judgement revealed by Wheatley and Haidt). This genealogy forms the basis of a debunking argument:

> The emotion remains overly sensitive to cues related to its primary functions of protecting against poisons and parasites, which results in many false positives even in those domains. There is no reason to think the situation improves when disgust operates in the socio-moral domain. (2014)

Kelly concludes that disgust 'is not even remotely a reliable indicator of moral foul play...[and] feelings of disgust themselves should be given no weight in deciding whether an issue...is morally acceptable or morally problematic' (2011, p.148).

Joshua Greene develops a similar argument (2008). Certain deontological moral intuitions, he argues, are driven by emotional mechanisms that played an adaptive role in our prehistoric past but which now fire in response to morally irrelevant factors. Faced with 'trolley problem' scenarios, for example, subjects express reluctance to push a large person to his death off a footbridge in order to prevent a runaway trolley from killing five workers on the track, but are considerably less reluctant to save the five by pulling a lever to divert the trolley onto a side-track resulting in the death of a distant individual. The explanation, Greene argues, is that the former scenario triggers psychological mechanisms concerned with dealing with 'up close and personal violence,' the emotional effects of which generate 'moral intuitions' against the former action but not against the latter action. The emotion-driven moral intuitions produced by the evolved human brain pertain not only to personal violence, but to retributive tendencies, to non-harmful actions (like food taboos), and to harming specific versus undetermined individuals. In Greene's opinion, these moral intuitions, coupled with a human tendency toward 'post-hoc confabulation,' are responsible for deontological moral theory.

Greene uses these theses (for some of which he presents empirical evidence) as the basis of a debunking argument. 'There are good reasons to think,' he concludes, 'that our distinctively deontological moral intuitions (here, the ones that conflict with consequentialism) reflect the influence of morally irrelevant factors and are therefore unlikely to track the moral truth' (2008, pp.69-70). Greene's target is larger than Kelly's, but is still selective; Greene believes that consequentialist moral thinking remains undebunked. (See also Singer 2005.)

One might be tempted to go further still, and aim for a *global* genealogical debunking of affect-based moral judgements. After all, the earlier debunking strategies within a cognitivist framework had global ambitions; why not also those within a noncognitivist framework? Suppose that the special forms of *liking* and *disliking* which I am assuming lie at the heart of noncognitivism are more like disgust than like headaches: that is, they are given practical weight because they are thought to provide insight into the like-worthy and dislike-worthy contours of the world. Just as disgust prompted by the word 'take' is discredited if the person believes she is responding to something else (e.g., to someone else's transgressions), so too would an act of liking be discredited if it were discovered that the person is badly mistaken about what factors have aroused the state. Suppose that the mechanisms producing these liking and disliking responses can be given a particular kind of evolutionary explanation: they emerged because they helped bolster various cooperative motivations in our ancestors. This may reveal that we are ordinarily mistaken about what factors in the environment our affective states are responsive to: the states do not provide the touted insight into the like-worthy and dislike-worthy aspects of the world—they are not truth-tracking at all, but simply influence our motivations in ways that were once adaptive (perhaps via having a truth-tracking phenomenology). As with the case of hypnotically-induced disgust, such false beliefs may be sufficient to discredit the affective states and thus the moral judgements that express them. The upshot may be nothing so radical as the prescription that we must attempt forthwith to purge our minds of these affective states (if even we could). The conclusion may be more analogous to the epistemological burden-of-proof-shifting discussed earlier—namely, that these affective states are left with a question mark hanging over them: they are not to be accorded the benefit of the doubt, they are not to be granted any privileged role in decision making.

Thus far I have had little to say about another obvious way of negatively evaluating affective attitudes: judging them detrimental to one's welfare. Even headaches can be assessed in this fashion. We tend to think of moral judgement in general as a prudentially good thing, but this is more of an item of faith than a properly scrutinized empirical thesis. Moral judgements can also be disastrous for those making them and for those around them. (Just think of all those patriotic young men who ended up as corpses in the trenches of the Great War.) A number of philosophers have pushed the view that on the whole we would be better off in practical terms if morality were eliminated from our mental and social lives (Hinckfuss 1987; Moeller 2009; Garner 2010; Marks 2013). It is not my task to evaluate the case(s) offered, but rather reflect briefly on how *genealogical* considerations might reveal the imprudence.

The argument follows a pattern by now growing familiar. Ordinarily, we might be willing to grant our affective states (like liking and disliking) the benefit of the doubt. We know we are evolved beings, and we might vaguely presuppose that evolution has designed us reasonably well. Pain exists to motivate us to respond to bodily injury, fear exists to motivate us to avoid danger, and so forth. Therefore when we feel pain (or fear, etc.), we have ground—at least in the absence of any reason to think otherwise—for assuming that its distinctive stimulus event is present, and that it is probably prudent for us to act as the pain (or fear, etc.) moves us to act. The same may be true of the affective states lying at the heart of noncognitivism (whatever they may be): we may take ourselves to have ground—in the absence of any reason to think otherwise—for assuming that it is probably prudent to allow these feelings a significant role in guiding our decisions. But this is precisely where a more detailed genealogical explanation can have an undermining impact, for it can reveal that the circumstances that rendered these affective states adaptive on the African savannah (say) no longer hold in the modern world, or fail to hold in some particular circumstances. Genealogical evidence can act as a defeater of the benefit of the doubt we would otherwise accord an affective state—overturning the assumption of its contribution to our welfare. Genealogical evidence can thus help reveal an affective state to be imprudent.

If the preceding arguments seem all rather slapdash, it is because my goal has not been to advocate them, but rather to highlight the fact that if these genealogical debunking arguments work at all, they work just as much against metaethical noncognitivism as against cognitivist success theory. Even if the noncognitivist is correct that moral judgements are no more than expressions of liking and disliking, these moral judgements can still be undermined by data concerning their evolutionary origins. Of course, this 'undermining' won't amount to *being false* or *being epistemically unjustified*, but it cannot on this ground be dismissed as unimportant.

Conclusion

Genealogical debunking arguments are varied, not only in their premises but in their conclusions. They may or may not focus on the evolutionary perspective. Sometimes they rely on a principle of parsimony in the service of a radical ambition to establish an error theory; sometimes they attempt to shift the burden of proof in the service of a more modest epistemological conclusion. Though usually operative against the background of cognitivist presuppositions, genealogical debunking arguments can also have force within a noncognitivist framework. While there may be some convenience in lumping genealogical debunking strategies together as a family of philosophical arguments, in order to be effective any reasonable critic must discriminate among strategies and deploy counter-arguments applicable to his/her chosen target.

Department of Philosophy
Victoria University of Wellington
Wellington 6140
New Zealand
richard.joyce@vuw.ac.nz

References

Alexander R. (1987). *The Biology of Moral Systems* (Hawthorne, NY.: Aldine de Gruyter).

Blackburn, S. (1984). *Spreading the Word* (Oxford: Clarendon).

Blackburn, S. (1993). *Essays in Quasi-Realism* (Oxford: Oxford University Press).

Brosnan, K. (2011). 'Do the evolutionary origins of our moral beliefs undermine moral knowledge?' *Biology and Philosophy* 26: 51-64.

Dwyer, S.(2006).'How good is the linguistic analogy?' In P. Carruthers, S. Laurence, & S. Stich (eds.), *The Innate Mind, Volume 2: Culture and Cognition* (Oxford: Oxford University Press). 237-55.

Garner, R. (2010). 'Abolishing morality.' In R. Joyce & S. Kirchin (eds.), *A World Without Values* (Dordrecht: Springer Press). 217-233.

Greene, J. (2008). 'The secret joke of Kant's soul.' In W. Sinnott-Armstrong (ed.) *Moral Psychology, Vol. 3: The Neuroscience of Morality* (Cambridge, MA.: MIT Press). 35-79.

Harman, G. (1977). *The Nature of Morality: An Introduction to Ethics* (NY: Oxford University Press).

Hinckfuss, I. (1987). 'The moral society: Its structure and effects,' *Discussion Papers in Environmental Philosophy* 16 (Canberra: Philosophy Program [RSSS], Australian National University).

Hume, D. [1740] (1978). *A Treatise of Human Nature*. L. Selby-Bigge (ed.) (Oxford: Clarendon Press).

Irons, W. (1996). ‘Morality as an evolved adaptation.’ In J. Hurd (ed.), *Investigating the Biological Foundations of Human Morality* (Lewiston, NY: Edwin Mellen Press). 1-34.

Joyce, R. (2001). *The Myth of Morality*. (Cambridge: Cambridge University Press).

Joyce, R. (2006). *The Evolution of Morality*. (Cambridge, MA.: MIT Press).

Joyce, R. (2007a). ‘Moral anti-realism.’ Entry for *The Stanford Encyclopedia of Philosophy*.

Joyce, R. (2007b). ‘Morality, schmorality.’ In P. Bloomfield (ed.), *Morality and Self-Interest* (Oxford: Oxford University Press). 51-75.

Joyce, R. (2009). ‘Is moral projectivism empirically tractable?’ *Ethical Theory and Moral Practice* 12: 53-75.

Joyce, R. (2011). ‘The accidental error theorist.’ In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Vol. 6* (Oxford: Oxford University Press). 153-180.

Joyce, R. (2014). ‘Evolution, truth-tracking, and moral skepticism.’ In B. Reichardt (ed.), *Problems of Goodness: New Essays on Metaethics* (NY: Routledge).

Kelly, D. (2011). *Yuck! The Nature and Moral Significance of Disgust* (Cambridge, MA.: MIT Press).

Kelly, D. (2014). ‘Selective debunking arguments, folk psychology, and empirical moral psychology.’ In J. Wright & H. Sarkissian (eds.), *Advances in Experimental Moral Psychology: Affect, Character, and Commitments* (NY: Continuum Press).

Kitcher, P. (2011). *The Ethical Project* (Cambridge, MA.: Harvard University Press).

Krebs, D. (2005). ‘The evolution of morality.’ In D. Buss (ed.), *The Handbook of Evolutionary Psychology* (NJ.: John Wiley & Sons). 747-771.

Machery, E. & Mallon, R. (2010). ‘The evolution of morality.’ In J. Doris, G. Harman, S. Nichols, et al. (eds.), *The Moral Psychology Handbook* (Oxford: Oxford University Press). 3-46.

Mandelbaum, M. (1956). *The Phenomenology of Moral Experience* (Glencoe, IL.: Free Press).

Marks, J. (2013). *Ethics without Morals: In Defence of Amorality* (NY: Routledge).

Mikhail, J. (2011). *Elements of Moral Cognition: Rawls’ Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment* (Cambridge: Cambridge University Press).

Moeller, H.-G. (2009). *The Moral Fool: A Case for Amorality* (NY: Columbia University Press).

Nichols, S. (2005). ‘Innateness and moral psychology.’ In P. Carruthers, S. Laurence, & S. Stich (eds.), *The Innate Mind: Structure and Contents* (NY: Oxford University Press). 353-430.

Ruse, M. (1986). *Taking Darwin Seriously* (Oxford: Basil Blackwell).

Ruse, M. (2006). ‘Is Darwinian metaethics possible (and if it is, is it well-taken)?’ In G. Boniolo & G. de Anna (eds.), *Evolutionary Ethics and Contemporary Biology* (Cambridge: Cambridge University Press). 13-26.

Ruse, M. (2009). ‘Evolution and ethics: The sociobiological approach.’ In M. Ruse (ed.), *Philosophy After Darwin* (Princeton, NJ: Princeton University Press). 489-511.

Schafer, K. (2008). ‘Practical reasoning and practical reasons in Hume,’ *Hume Studies* 34: 189-208.

Shafer-Landau, R. (2007). 'Moral realism: Introduction.' In R. Shafer-Landau & T. Cuneo (eds.), *Foundation of Ethics* (Oxford: Blackwell). 157-162.

Singer, P. (2005). 'Ethics and intuitions,' *Journal of Ethics* 9: 331-352.

Sinnott-Armstrong, W. (2006). *Moral Skepticisms* (Oxford: Oxford University Press).

Sober, E. (2009a). 'Absence of evidence and evidence of absence: Evidential transitivity in connection with fossils, fishing, fine-tuning, and firing squads,' *Philosophical Studies* 143: 63-90.

Sober, E. (2009b).'Parsimony arguments in science and philosophy: A test case for naturalism$_p$,' *Proceedings and Addresses of the American Philosophical Association* 83: 117-155.

Street, S. (2006). 'A Darwinian dilemma for realist theories of value,' *Philosophical Studies* 127: 109-166.

Wheatley, T. & Haidt, J. (2005). 'Hypnotic disgust makes moral judgments more severe,' *Psychological Science* 16: 780-784.

Wright, C. (1992). *Truth and Objectivity* (Cambridge MA: Harvard University Press).