

# Debunking What?

Hallvard Lillehammer

## 1. The Problem

In the last few decades, there has been a revival in the philosophical attention paid to the thought that there is something essentially dubious about core aspects of ethical thought (Williams 1985, Joyce & Kirchin 2010, Garner & Joyce 2019).<sup>1</sup> Some philosophers attending to this question have concluded that ethical claims are uniformly false or incoherent (see, e.g., Joyce 2001, Olson 2014).<sup>2</sup> Some philosophers have defended an epistemological analogue of this ‘error theoretical’ conclusion, namely, that ethical claims are uniformly unjustified on account of their evolutionary, or

---

<sup>1</sup> It is a contested matter what the range of the ‘ethical thought’ in question is. I shall not pursue this matter here, beyond stating that ethical questions are, in some sense, concerned with practical questions of ‘how to live’; questions which present themselves as, in some sense, inescapable. For further discussion, see e.g. Williams (1985), Joyce (2001), and Enoch (2011).

<sup>2</sup> In this chapter, I use the term ‘ethical claim’ to refer interchangeably to: (a) the content of an ethical judgment, (b) a state of commitment to such a content, (c) the act of committing oneself to such a content, and (d) the act of verbally articulating a commitment to such a content in public speech. As will become clear in what follows, I intend this use of ‘ethical claim’ to be neutral with respect to the ‘cognitive’ status of such claims, and therefore also on the issue of whether ethical claims have contents that are, in a contested sense, “representationally robust” (see e.g. Price 2011). I adopt the same neutrality with respect to uses of the term ‘commitment’. It follows from this neutrality that whereas many of the claims made in this chapter could be accepted by various kinds of Moral Error Theory (such as (partial) Conservationism, or Fictionalism), these claims do not entail any such view.

otherwise causal, pedigree (see, e.g., Joyce 2005, Sauer 2018).<sup>3</sup> Among philosophers who have been attracted to both of these claims, some have gone on to ask if it would nevertheless make sense for us to continue living in accordance with core ethical claims, possibly even by continuing to make them, in spite of their lack of truth or justification (see, e.g., Joyce 2001, 2005, Olson 2014). At that point, a number of allegedly non- or pre-ethical considerations have been called upon in order to suggest that a continued commitment to ethical claims could be a reflectively coherent option; from appeals to expected subjective utility at one extreme to the prevention of climate change at the other (see, e.g., Joyce & Garner 2018).

Although there is much to be admired in the ingenious articulation of this combination of ideas, there is one further question that must be answered in order to take it seriously as a coherent option. This is the question of why we should think that our claims about whether or not to maintain our commitment to these allegedly untrue or unjustified ethical claims are any more likely candidates for truth or justification than the core ethical claims the lack of truth or justification of which they presuppose. To put the question in a different way: why should we think we have the capacity to think truly or justifiably about the question of whether to maintain our commitment to ethical

---

<sup>3</sup> The two claims are logically independent. Moral Error Theory is consistent with the claim that some ethical claims are justified but false. Moral Realism is consistent with the claim that people are unjustified in their beliefs about what the ethical truths are. Further complications arise from the fact that some philosophers have employed a debunking argument in order to show that ethical claims would be unjustified *if some meta-ethical theory* - such as a robust form of Non-naturalist Moral Realism - *were true*; thereby purporting to undermine that meta-ethical theory (Street 2006). In what follows, I bracket these complications.

claims if we don't have the capacity to think truly or justifiably about the question of what is good, right, or virtuous?<sup>4</sup>

In what follows, I assume for the purposes of argument that we do in fact have some capacity to think intelligently about the question of whether or not to retain our commitment to at least some ethical claims in spite of their lack of truth or justification, and therefore that there is an explanatory challenge here that can in principle be met. I approach this question in a somewhat unusual way, namely by focusing on two thoughts. The first thought is that while norms of truth and desirability often go together, that is not always the case. This is an issue that has historically been discussed under the label “the value of truth” (see, e.g., Nietzsche 1887, Williams 2002). The second thought is that no one likes to be made a fool of. (Professional philosophers arguably like to be made fools of less than most.) While each of these thoughts might strike the reader as quite familiar and perhaps even platitudinous, what is less familiar—and hardly platitudinous—is how these two thoughts are related with respect to the interpretation of debunking arguments in ethics. In what follows, I illustrate how the two thoughts are related with respect to debunking arguments in ethics by introducing a distinction between two different ways in which a given ethical claim can be the target of debunking, and which I refer to as ‘debunking *de dicto*’ and ‘debunking *de re*’. The aim of introducing this distinction is to provide a partial description of the different

---

<sup>4</sup> The idea that questions about *what is in my own best interest*, *what I have reasons to do*, or *what it is rational for me to do* are vulnerable to some of the same sources of skeptical contestation with respect to their truth or justification as the question of *what it is good, right or virtuous for me to do* is one implication of the view defended in Streumer (2017). See Lillehammer (1999) for a less precisely elaborated version of this idea. See Nagel (1971) for a discussion that sets the stage much of the subsequent debate.

ways in which facts about the causal pedigree of ethical claims can affect their standing, depending on the extent to which assumptions or beliefs about those facts are embedded within those ethical claims themselves. I then go on to distinguish three possible responses to the discovery of such facts, which I call ‘the resistance response’, ‘the retreat response’ and ‘the capitulation response’, respectively. I suggest that each of these responses could be appropriate in different circumstances, although no single one of them can be assumed to be the default response when confronted with the hitherto unknown causal pedigree of any given ethical claim.

For the purposes of this chapter, a ‘debunking argument’ is an argument that purports to undermine someone’s entitlement to a given set of claims in virtue of the fact that the endorsement by that someone of those claims has a certain causal pedigree, such as being based on ignorance, false beliefs, false presuppositions, or having come about as a result of some less than fully rational process. A debunking argument *in ethics* is an argument that purports to undermine someone’s entitlement to a given set of *ethical* claims in virtue of the fact that the endorsement by that someone of those claims has a certain causal pedigree, such as being based on ignorance, false beliefs, false presuppositions, or having come about as a result of some less than fully rational process. Although my discussion in this chapter will make particular reference to *evolutionary* debunking arguments in ethics, the target of the discussion is much broader than that and can be shown to extend to debunking arguments in moral and political philosophy more generally (see e.g. Joyce 2005, Singer 2005, Street 2006,

Greene 2008, Kahane 2011, Vavova 2015, Leibowitz & Sinclair 2016, Ruse & Richards 2017, Hanson 2017, Sauer 2018, Klenk 2019).<sup>5</sup>

## 2. Epistemic and Practical Debunking

When considered on their merits as potentially debilitating challenges to existing instances of ethical thought, debunking arguments have not only a narrowly *epistemic* aspect (such as pointing out the negative epistemic implications of having one's ethical convictions depend on dubious factual evidence), they also have a broader *practical* aspect (such as pointing out the badness, shamefulness or wrongness of letting one's ethical convictions be hostage to vested interests).<sup>6</sup> Thus, we can coherently ask, of any given ethical claim, *how much it matters* whether it is insensitive, unsafe, unreliable or unsystematic, and thereby defective along its paradigmatically epistemic dimension (cf. Lillehammer 2011, 196-200). In other words, we can coherently ask whether or not we

---

<sup>5</sup> There is considerable historical variation in the extent to which different debunking arguments in ethics have had a primarily epistemological focus on justification or knowledge, or a primarily metaphysical focus on truth or factuality. I pass over this complication here. See, e.g., Joyce (2016) for further discussion.

<sup>6</sup> Throughout this chapter I use the labels 'ethical' and 'epistemic' to contrast different aspects of what is commonly referred to as 'normativity'; where the former is commonly classified as an aspect of 'practical normativity' and the latter is commonly classified as an aspect of 'theoretical' normativity. Although I make free use of these labels in what follows, I do not offer a general account of normativity in this chapter. What matters for present purposes is that just as practical (e.g. ethical) claims can be evaluated from a theoretical (e.g. epistemic) perspective, so theoretical (e.g. epistemic) claims can be evaluated from a practical (e.g. ethical) perspective.

either ought, or can reasonably aspire, to live with a given set of insensitive, unsafe, unreliable or unsystematic ethical claims, for example because by doing so we shall be better placed to promote such ends that we, after causally informed reflection, would endorse as desirable aspects of social life. Moreover, this question of *how much it matters* is a practical question our reflection on which is naturally (if not inevitably) articulated in recognizably *ethical* terms.

It is worth pausing briefly to register two implications that flow from this fact. The first is that an ethical claim could in principle be epistemically lacking yet still be sensible to retain a commitment to, all things considered (e.g., because it is conducive to basic sanity or survival).<sup>8</sup> Second, the question whether ethical claims are vulnerable to debunking is not a matter to be decided purely on the basis of a prior decision about whether to classify ethical claims as ‘cognitive’ (and so the contents of robustly ‘representational’ states) on the one hand, as opposed to ‘non-cognitive’ (and so the contents of various ‘affective’ states), on the other.<sup>10</sup> It does not so depend because both cognitivist and non-cognitivist interpretations of ethical thought are consistent with the possibility that ethical claims are vulnerable to debunking along one dimension or other. Non-cognitivists are faced with this challenge (at least) along its practical

---

<sup>8</sup> It is a point worth noting in this context that for some people in some circumstances an unconditional commitment to norms of truthfulness and other paradigmatically epistemic virtues would be inconsistent with their own pursuit of wellbeing or self-preservation.

<sup>10</sup> The truth, of course, is that ethical claims have properties in common both with paradigmatic beliefs (e.g., their claim to authority or correctness) and with some paradigmatically affective states (e.g., their practicality or motivational aspect). So, whether we categorize them as either beliefs or affective states, they will be beliefs or affective states of a somewhat special sort. For further discussion, see, e.g., Blackburn (2008, 2011), Sinclair (2021).

dimension (e.g., interpreted as a question of what attitudes it is wise or admirable to adopt towards the causal pedigree of our ethical claims). Cognitivists are faced with the challenge along both its epistemic and its practical dimension (e.g., interpreted in terms of whether ethical beliefs are epistemically safe or sensitive on the one hand, and in terms of what we are to make of the answer to that question in our practical reasoning on the other). As paradigmatically exemplified by the writings of recent moral error theorists, for example, not only does the practical aspect of the debunking challenge remain once the epistemic aspect has been dealt with. This practical challenge is in principle open to a conservative response on error theoretical terms (see, e.g., Joyce 2001, Lillehammer 2004, Olson 2014). If the question of the plausibility of debunking arguments in ethics were reducible without remainder to a set of narrowly epistemic or ‘theoretical’ questions about the extent to which ethical claims comply with norms of truth-conducive inquiry, then the very idea of endorsing a conservative form of Moral Error Theory would be incoherent. Yet it isn’t.<sup>11</sup>

### **3. The Entanglement Objection**

It might be objected that the distinction drawn in the previous section between epistemic and practical considerations, and according to which the epistemic standing of ethical claims could be debunked while in principle leaving their practical standing intact, is based on a simple misunderstanding.

After all, it is just not true that the norms at stake in the evaluation and criticism of ethical claims (or any other practical claims for that matter) can be sharply divided

---

<sup>11</sup>See, e.g., Clarke-Doane (2020) for an argument that I take to raise a similar set of questions, although from a somewhat different direction..

into the epistemic ones, such as safety or sensitivity, on the one hand, and the practical ones, such as goodness or rightness, on the other. On the contrary, some considerations of key relevance to the evaluation and criticism of ethical claims are manifestly both epistemic and practical. For this reason, if for no other, we should think of the epistemic and the practical as necessarily ‘entangled’ (see, e.g., Putnam 1981, Cuneo 2007, Lillehammer 2007).

Consider, for example, the virtues of *consistency*, *integrity*, *honesty*, *truthfulness* and *understanding*. The extent to which an individual fails to manifest one or more of these characteristics is an issue of relevance to their epistemic standing as reasonable inquirers (see e.g. Elgin 2017). Yet the extent to which they fail to manifest one or more of these characteristics is also an issue of relevance to their practical standing as good or right-minded people. The idea that we should somehow be able to neatly distribute these considerations into one of these categories or the other is based on a basic misunderstanding of the nature of normativity.

There is more than one way to develop this objection. One way of developing it is to argue that epistemic and practical norms are constitutively part of a single normative network, or ‘web’, in which the norms in question are neither purely epistemic nor purely practical. Let’s call this ‘the entanglement thesis’. Although further work would be required to fully articulate and assess that thesis, there are plausible considerations that speak in its favor (see, e.g., Putnam 1981, Cuneo 2007, Lillehammer 2007). Nevertheless, I am not able to argue for the entanglement thesis in this chapter.

Fortunately for present purposes, it is not necessary to establish the entanglement thesis in order to object to the employment of a logically sharp ‘epistemic versus practical’ distinction in the argument discussed in the previous section. Another



way of developing that objection is to concede that epistemic and practical norms may be logically distinct, but to point out that across a wide range of cases both epistemic and practical norms favor and disfavor the same things, where this may include ways of acting, ways of judging, ways of thinking, ways of feeling, or whatnot.

Thus, even if there is a logically sharp distinction between epistemic norms on the one hand and practical norms on the other, it could still be the case that epistemic and practical norms both speak in favor of the cultivation of characteristics such as coherence, integrity, honesty, truthfulness or understanding in the genesis, development and maintenance of ethical claims. And this is all we need to get the objection off the ground. For suppose that a debunking argument against a given set of ethical claims were to show that the causal pedigree of those claims as endorsed by a given set of thinkers is such as to put into question the coherence, integrity, honesty, truthfulness or understanding of those thinkers. In that case, the debunking argument in question would succeed in putting into question both the epistemic and the practical standing of those thinkers with respect to those claims. And this will be true even if the entanglement thesis is not. And it is quite easy to think of debunking arguments in ethics where the norms in question have both a strongly epistemic and a strongly practical ‘feel’. Among the most plausible cases in question are debunking arguments targeted at ethical claims based on motivated bias, self-deception; adaptive preferences, false consciousness, or ideology. This is not to suggest that all debunking arguments in the literature have this feature. Yet some such arguments, including some evolutionary debunking arguments, arguably do.

The fact that some debunking arguments in ethics have primarily been targeted at the alleged failure of ethical claims to comply with epistemic norms does not, therefore, show that these arguments must leave the practical standing of those claims

open or intact. This further conclusion would only follow if the ethical claims in question did not have their practical standing in any way affected by their failure to comply with the epistemic norms in play. Yet if the epistemic failures exposed by these arguments are also practical failures, this is not so. In which case, the debunking arguments in question could in principle undermine someone's entitlement to the ethical claims in question along their practical dimension as well.

Even though the entanglement objection is not one that can be so easily dismissed or evaded, neither is it an objection that can be used as some kind of skeptical 'Trojan Horse' that opens up a direct path to the blanket rejection of all ethical claims that are found to have an epistemically dubious causal pedigree. First, the entanglement in question may not extend to all ethical claims (or to all aspects of those claims). Second, the insistence on rejecting the ethical claims in question would itself be a practical claim about which ethical claims we should consider ourselves entitled to in the face of our best evidence about the causal pedigree of those ethical claims. And although it should be conceded that practical reflection on this causal pedigree has the potential to motivate a systematic reconsideration of an arbitrary range of existing ethical claims, it remains an open question how threatening to the reflective coherence of those claims that potential is. I shall return to this question in Section 5. below.

#### **4. Debunking *de dicto* and *de re***

Consider a variant of H. C. Anderson's tale, "The Emperor's New Clothes," in which a person proudly walks around completely naked in an ostentatious display of their non-existent attire, and where no one is prepared to tell them that they have nothing on until an innocent child exclaims that the Emperor is naked. Let's assume that the Emperor has formed his views about his splendid appearance on faulty epistemic grounds, due

to a combination of deception, ignorance or false belief, some of which may have registered within the content of his own self-conception (e.g. ‘I look great in the infinitely fine shirt that the tailors have made for me’), and some of which may not have (e.g. the tailors have lied and cheated their way through the entire tailoring process, regardless of whether the Emperor has been directly involved or not). Among the Emperor’s subjects, we can assume that some of them are in on the scam, others are taken in by it; and others again are able see what is going on but for reasons of their own have decided to stay silent. Either way, the default assumption is that the Emperor has been made a fool of.

By analogy with the various ways in which the Emperor’s claims about the splendour of his attire can be subject to a debunking challenge, there is an analogous plurality of ways in which ethical claims can be challenged with respect to the facts about their causal pedigree. Here I shall consider two.

In the first kind of case, it is judged (by some potentially debunking observer or participant) that:

(1) A should not believe [that P *conditional on* causal pedigree C],<sup>12</sup>

---

<sup>12</sup> The phrase ‘conditional on’ merits further unpacking beyond what I provide here. For example, the associated assumption or belief in causal pedigree targeted by *de dicto* debunking need not itself be regarded by the subject itself as a *reason* for the claim targeted for debunking. If reproducing by means of IVF were a mortal sin, the fact that it says so in the Bible would not itself be a reason *why* this is so, even it were a reason to believe it is so. In other cases, the connection between the relevant reasons and assumed or believed causal pedigree could be even more tenuous. Consider someone retaining some ethical belief conditional on their conviction that they have always believed it.

where we stipulate that the information in the square brackets is reflectively available to A (either explicitly or implicitly) from within the content of their propositional attitudes, and so is ‘within the scope’ of those attitudes. For example, we might advise someone in some given context that it is inappropriate in that context to endorse the claim that P on the grounds of an assumption or belief about the existence of causal pedigree C; either because that causal pedigree does not obtain; because the plausibility of the claim that P does not actually correlate with the existence of said causal pedigree; or because we ethically disapprove of people who believe that P conditional on their belief in C. Consider, for example, some person A’s claim that: ‘Reproducing by means of IVF is a mortal sin [as we learn from The Bible]’. In this kind of case, A’s commitment to the claim that reproducing by means of IVF is a mortal sin could in principle be debunked in any of the following ways: (a) A has not, in fact, read this claim in The Bible, (b) The Bible does not contain the claim that reproducing by means of IVF is a mortal sin, (c) the plausibility of the claim that reproducing by means of IVF is a mortal sin does not correlate with whether or not it says so in The Bible, or (d) forming one’s ethical convictions on the basis that ‘it says so in the Bible’ is a bad, wrong or irresponsible way to cultivate one’s ethical sensibility, and so on. Let’s refer to this kind of debunking as ‘debunking *de dicto*’. The interest of debunking *de dicto* in the case of the Emperor depends on the extent to which the Emperor’s behavior is explained by his assumptions or beliefs about how he came to look the way he does. By analogy, the meta-ethical interest of debunking *de dicto* depends on what, if any, assumptions or beliefs about the causal pedigree of ethical claims are actually embodied in pre-theoretical ethical thought.

In a second kind of case, it is judged (by some potentially debunking observer or participant) that:

(2) A should not believe that P [*conditional on* causal pedigree C],

where we stipulate that the information in the square brackets need not be reflectively available to A (either explicitly or implicitly) from within the content of their propositional attitudes, and so is not ‘within the scope’ of those attitudes (although it will be in within the scope of the propositional attitudes of whoever is purporting debunk their judgment). Thus, we might advise someone that the causal pedigree of their making some claim is inappropriately related to their entitlement to make that claim; hence they should at least reconsider their previous commitment to that claim. Consider, again, some person A’s claim that: ‘Reproducing by means of IVF is a mortal sin’, where they do not have any firm beliefs about how they have come to think so. In this kind of case, A’s commitment to the claim that reproducing by means of IVF is a mortal sin could in principle be debunked in any of the following ways: (a) A is not aware of any reliable source for the claim that reproducing by means of IVF is a mortal sin, (b) there is no reliable source for the claim that reproducing by means of IVF is a mortal sin, (c) the actual causes of A’s believing that reproducing by means of IVF is a mortal sin do not correlate with any reliable sources for making informed judgments about the ethics of human reproduction, or (d) to form sweeping judgments about the ethics of human reproduction without careful consideration of the causal pedigree of those judgements is a bad, wrong or irresponsible way to cultivate one’s ethical sensibility, and so on. Let’s refer to such debunking as ‘debunking *de re*’. The interest of debunking *de re* in the case of the Emperor is largely independent of the extent to

which the Emperor's behavior is explained by any assumptions or beliefs on his part about how he has come to adopt it. By analogy, the meta-ethical interest of debunking *de re* is largely independent of what claims about the causal origins of ethical beliefs are embodied in pre-theoretical ethical thought.

This rough distinction between debunking *de dicto* and debunking *de re* marks a difference between debunking arguments targeted at assumptions or beliefs about the causal pedigree of ethical claims on the one hand, and facts about that causal pedigree (whether registered by the relevant subject or not) on the other. This distinction is orthogonal to the following two distinctions. First, the distinction between debunking *de dicto* and *de re* is not equivalent to the distinction between debunking being either *local* or *global*. Although debunking *de dicto* is necessarily restricted to ethical claims with which the subject associates some assumption or belief about its causal pedigree, it can be directed at a narrower or wider range of such claims. More importantly, although debunking *de re* could in principle target the maximal set of ethical claims endorsed by some subject, it could equally be targeted at a single claim within that set. Second, the distinction between debunking *de dicto* and *de re* is not equivalent to the distinction between the target claim being *undercut* (as in being undermined) on the one hand, or being *rebutted* (as in being undermined in favor of a contrary claim) on the other. Both *de dicto* and *de re* debunking could consistently take either the form of undercutting or rebutting the relevant target claim, depending on particularities of the circumstances.

The fact that some ethical claim is vulnerable to debunking *de dicto* does not entail that the same claim is vulnerable to debunking *de re*.<sup>14</sup> Consider the following scenario. Suppose I am a generally manipulative and unsympathetic colleague. If so, you should probably not believe what I say about your merits with respect to the advisability of applying for a senior promotion *because I said so*. If that's what you believe, your belief is vulnerable to debunking *de dicto*. It does not follow that you should not believe what I say about your merits with respect to the advisability of applying for a senior promotion *as a result of me saying so* (because, let's say, the credibility of my utterances on matters of career progression is unusually high, as manifested by the fact that most of our otherwise sensible colleagues tend to agree with me). Alternatively, even if you should not believe what I say about your merits with respect to the desirability of applying for a senior promotion as a result of me saying so, it does not follow that *you should not believe what I say about these merits* (e.g., that a senior promotion is long overdue). After all, you could have evidence available that rules out all conflicting hypotheses (e.g., as a result of recently having seen the official promotion criteria). In neither case is your belief vulnerable to debunking *de re*.<sup>15</sup>

It is a notable difference between many of the evolutionary debunking arguments in contemporary ethics on the one hand, and at least some paradigmatic cases

---

<sup>14</sup> The reverse claim might initially seem more plausible. Yet even here it might be possible to think of counterexamples. (Perhaps some forms of false beliefs in causal pedigree have a felicitously fortifying effect. And so on.) I pass over this issue here.

<sup>15</sup> For versions of this thought formulated in terms of the idea that some objective notion of self-evidence could pick up the slack in this way, see, e.g., Ross (1930). For a later development, see Audi (2004). See also Brosnan (2010) and de Lazari-Radek & Singer (2014).

of what are sometimes labelled ‘genealogical’ debunking arguments historically associated with Nietzsche and other critics of ‘morality’ on the other, that whereas some paradigmatic instances of the latter have been targeted at ethical claims endorsed or sustained as a result of false and allegedly deceitfully disseminated beliefs *about their origins* (e.g., by “Christianized” victims of resentful “ascetic priests” during the course of what Nietzsche calls “the slave revolt” who regard their ethical convictions as originating in the mind of a trustworthy God), paradigmatic instances of the former have been targeted at ethical claims in a way that impose few, if any, restrictions on the causal assumptions or beliefs that have either produced or sustained them (see e.g. Nietzsche 1887). In other words, whereas at least some of the genealogical debunking of at least some ethical beliefs that has paradigmatically been practiced by Nietzsche and other critics of ‘morality’ is arguably best understood as a form of debunking *de dicto*, much of the evolutionary debunking paradigmatically practiced in contemporary ethics is generally best understood as taking the form of debunking *de re* (see also Zimmermann 2018). Moreover, paradigmatic instances of the genealogical debunking of ethical beliefs practiced by Nietzsche and other critics of ‘morality’ have incorporated an unmistakable practical element that transcends the boundary of the narrowly epistemic, such as Nietzsche’s appeal to some ethical claims being “healthy”, “higher”, of “superior rank”, and the like. Indeed, it is partly this aspect of these genealogical debunking arguments, as much as their narrowly epistemic aspects, that has given them their widespread notoriety and rhetorical force.<sup>16</sup>

---

<sup>16</sup> Compare the following definition of ideology: “An ideology... is a set of beliefs, attitudes, preferences that are distorted as a result of the operation of specific relations of power; the distortion will characteristically take the form of *presenting these beliefs, desires, etc., as inherently connected with some universal interest, when in fact they are subservient to*



Assuming for the sake of argument the truth of their respective claims about the causal pedigree of ethical claims, it is reasonable to think that the Nietzschean genealogical debunker has a stronger case against their target ethical claims than does the contemporary evolutionary debunker, *all else being equal*. It would not be very surprising, therefore, to find that some proponents of evolutionary debunking arguments in ethics have been tempted to run together these different types of debunking argument by attributing to our ethical sensibility beliefs or assumptions about the origins of that sensibility which are more plausibly interpreted as the contingent accompaniments of such beliefs as found in the heads of a selected group of historical individuals (see e.g. Williams 1985).<sup>17</sup> If so, this would not be the only case in which philosophers have attributed to ethical thought *as such* a set of contentious ‘conceptual’, or ‘constitutive’, commitments about the nature and status of ethical claims which, upon further investigation, turns out to be a function of their

---

*particular interests.*” (Geuss 2008, 52. My italics). ‘Ideology critique’, understood along these lines, is naturally interpreted as partly involving a form of *de dicto* debunking of beliefs, attitudes and preferences maintained conditionally on false assumptions or beliefs about the nature and credibility of their causal pedigree (e.g. the education system, the state, or persons in authority).

<sup>17</sup> Some debunking arguments in contemporary meta-ethics quite explicitly attribute something like a manipulative intention to evolutionary forces, if only metaphorically. The following is a paradigmatic example of the kind: “It seems, therefore, that morality is a collective illusion foisted upon us by our genes” (Ruse 1986: 257). For further discussion of this tendency, see Lillehammer (2003). It is a question worthy of further investigation whether some of these arguments are themselves vulnerable to a form of ideology critique.

variable articulation and endorsement in different times and places (see e.g. Kirchin 2010, Joyce 2011, Finlay 2014).

None of this is to suggest that all ethical claims are somehow guaranteed to be safe from either epistemic or practical debunking, whether *de dicto* or *de re*. After all, it is an incontestable fact that many of the ethical claims that have historically been endorsed by human beings have been based on ignorance, false belief, or various processes of non-rational attitude formation; and where this has in some cases included false assumptions or beliefs about the causal pedigree of those claims themselves (see e.g. Greene 2013, Haidt 2012). Yet even the fact that your commitment to some ethical claim is based on a false assumption or belief about the causal pedigree of that claim itself does not entail that you have therefore been made a fool of in the manner of the Emperor in H. C. Andersen's story. On the contrary, it is always a further question what to do about this fact with respect to your previous commitment to the ethical claim in question. In particular, there is an important difference between cases where:

- (i) your commitment to some ethical claim is challenged on the basis that it has been formed as a result of a given non-rational causal process of which you were previously unaware (e.g., 'I would not have believed in filial obligations had I not had the early childhood experience of parental love'); and:
- (ii) your commitment to some ethical claim is challenged on the basis that knowledge of the fact that your commitment to that claim is a result of a given non-rational causal process is reflectively incompatible with continued commitment to that claim (e.g., 'I would never have believed that my proper place is at the bottom of the social hierarchy had I not been lied

to about my inherently inferior ‘nature’ by some of my more privileged peers’).

Of course, an emerging suspicion that you have been actively misled could sometimes alter your reflective predicament in such a way that (i) would shade into (ii). Having said that, not all cases of being actively misled are quite like that of “The Emperor’s New Clothes.” There are also less ethically problematic cases where a person is misled in a way that is genuinely in their own interest, and in such a way that the causes of coming to make some relevant ethical claim are in principle absorbable by subsequent reflection as either welcome, irrelevant or an interesting detail in the multifarious landscape of social interaction. Consider, for example, a situation in which you lure me in the direction of an unknown location where you have arranged a surprise party for me. In the case of the Emperor’s New Clothes, it might well be reasonable to respond to the discovery of the deception involved by saying: ‘If only I had known how I came to judge the way I did, I would never have let that happen!’. In the case of the Surprise Party, on the other hand, this would probably be a more unusual response (perhaps on the part of someone who really doesn’t like surprises).

## **5. Resistance, Retreat or Capitulation**

In response to a revelation of the hitherto unknown causal pedigree of some ethical claim (whether *de dicto* or *de re*), there are at least three different directions in which practical reflection might go.

In some cases, the revelation of causal pedigree could reasonably result in a mild shrugging of the shoulders. Consider, for example, the well-established fact that many of the justifications people give of their actions are based on ‘post-hoc’

rationalizations which have at most a statistically weak connection to the causally operative factors that normally explain behaviors of the relevant kind (see, e.g., Haidt 2012). Let's say that you pay me back the money you owe me and explain why you did so by saying that not paying one's debts to a colleague *just isn't done*. Let's suppose that a social psychologist then points out that the overwhelming majority of people in your situation will tend to stop paying back their debts to colleagues once it is clear to them that they can do so without significant sanction. Therefore, or so they claim, it was in all likelihood as much the non-negligible probability of sanction as your intellectual grasp of any ethical insight that made you pay off your debt. 'Fair enough', you respond. 'I guess I'm lucky that I carry this residual fear of sanctions, because as far as I'm concerned not paying one's debts to a colleague *just isn't done*.' Let's call this *the resistance response*.<sup>21</sup>

I suspect that in many cases where the previously unknown causal pedigree of ethical claims is exposed by appeal to statistics or experiment, the resistance response is a reasonable one. What this suggests, I think, is that although our epistemic standing is quite important to us, it is not (or not always) that important. Our commitment to at

---

<sup>21</sup> A reflectively robust instance of the resistance response would be a way for you to 'succeed', 'do well', or 'get things right' with respect to an ethical claim, even if there are no moral facts or truths for you to correctly "represent" by means of it (e.g. because some version of Moral Error Theory or Non-cognitivism is true). Virtually all non-realist meta-ethical theories are committed to making sense of some such vocabulary, at least some of which is likely to have a substantially ethical 'feel'(see e.g. Joyce & Garner (2018)). On a cognitivist account (such as Moral Realism), the case in question could consistently be described as one where at least some paradigmatically epistemic norms are *practically self-effacing*.

least some ethical claims is, at least sometimes, reflectively robust enough to withstand the ubiquitous fact of our limited causal knowledge and rationality.

In other cases, the revelation of the causal pedigree of ethical claims may strongly favor a substantial shift in those claims. As already noted, no one likes to be made a fool of. And the fact that one has been made a fool of, perhaps by others who stand to personally benefit at one's expense, could *itself* be a reason to revise one's ethical claims, quite independently of whether the behavior to which one was previously committed is of a kind such that, independent of its causal pedigree, one would intrinsically disapprove of it. By analogy, if the Emperor were to discover the fact that he has been exposed to public ridicule by being encouraged to walk around naked in front of his minions, he may quite sensibly reconsider his previous view of himself as having so splendidly shown off his attire. Moreover, he may decide to reconsider his previous view of himself even if he does not thereby come to judge that there is anything wrong about him walking around naked in public. The fact of his being publicly humiliated this way could be enough. Indeed, his recognition of this fact could in principle outweigh the epistemic standing of his manner of coming to behave the way he did.<sup>22</sup> Let's call this *the retreat response*.

I suspect that many cases where the previously unknown causal pedigree of ethical claims is revealed the retreat response is a reasonable one. Moreover, and as previously noted, this could be so quite independently of one's view about the intrinsic merits of the ethical claims the casual pedigree of which has now been exposed. What

---

<sup>22</sup> The significance of this kind of case for debunking arguments targeted at specific ethical claims is closely connected with the way in which pride, recognition or self-respect can be involved in being genuinely *committed* to some ethical claim or other. To pursue these connections here would take me too far afield.

this suggests, I think, is that our epistemic standing is quite important to us; sometimes to the point of having a decisive influence on how we should go on to reflectively regard our ethical claims in response to having that standing exposed as epistemically lacking. Our ethical claims are not always robust enough to withstand the potentially ubiquitous fact of our limited causal knowledge and rationality.

In a third kind of case, the revelation of its causal pedigree could genuinely favor a state of bafflement or paralysis with respect to our continued endorsement of some ethical claim. This could be the case, for example, where the facts in question reveal us to be so irrational, confused or ill-informed on the matter at hand that there is no reflectively stable way to retain our bearings when faced with the relevant facts. Let's call this *the capitulation response*.

Although some arguments in the recent literature on debunking arguments in ethics could be read so as to suggest that this is the actual situation with respect to ethical claims in general, I don't believe the capitulation response gives an accurate picture of how most of these arguments are best interpreted. As is widely recognized, even in the most extreme cases where allegedly debunking explanations have been given of particular ethical claims, whether this be in controlled experimental conditions or in the models of evolutionary psychologists, there is normally a set of implicit assumptions of a broadly ethical nature in play that serve to make those examples practically relevant and interesting. For example, the recently much publicized experiments that have been taken to show that people are inclined to judge actions, people or situations differently depending on a range of contextually specific and ethically irrelevant factors are not best interpreted as having the power to debunk all existing ethical norms of decency, reasonableness or proportionality (see, e.g., Prinz 2007, 23-32). And the sociobiological modeling that has been taken to suggest that

standard deontological responses against directly killing or otherwise harming people in one's immediate vicinity are a contingent residue of human brains designed for an ancestral environment which no longer exists (see, e.g., Singer 2005, Greene 2013) are not best interpreted as having the power to favor the wholesale abandonment of our ethical sensibility in favor of a human extinction club. On the contrary, these and similar results produced by recent work in social or evolutionary psychology are normally presented against an implicit background of shared ethical claims our reflective commitment to which is not seriously in question; at least not in that context.

I suspect there are relatively few cases where exposure of the previously unknown causal pedigree of ethical claims would be such as to make a capitulation response the uniquely most reasonable one to adopt. To take one extreme case, consider the recently much discussed thought experiment involving someone who has taken an ethical "belief pill" (see, e.g., Joyce 2005, Sinclair 2018). Suppose, in a variant of the standard belief-pill case that, instead of discovering that you have swallowed a belief pill at time  $t$ , the effectiveness of which is targeted at some particular ethical belief content  $P$ , at time  $t$ , you discover that some variably targeted ethical belief pill-like substance is widely present in the air. No doubt this would present something of a challenge to your powers of practical reasoning. Yet however challenging it might be, you will nevertheless have to work out, on any given occasion, what the best, right, or acceptable thing to do is, and so how to live in light of this fact. It cannot be assumed without further argument that the reasonably believed-in presence of such an ethical belief pill-like substance in the air must thereby prevent you from working out what to do or how to live in a more or less admirable way. For example, depending on your initial values and credences, there will be some ways of responding to this discovery (such as complete mental paralysis) that will be less advisable than others (such as

continuing to seek nourishment and company in a way that is consistent with continued survival). What this suggests is that although our epistemic standing is very important to us, its importance is not such as to cancel out everything we value, however we may actually have come to value it. Although our commitment to some ethical claims is not always robust enough to withstand the ubiquitous fact of our limited causal knowledge and rationality, it sometimes is. To draw the opposite conclusion is to think that because we invariably fail to comply with epistemic norms, we must thereby invariably fail to make sensible practical judgments about what to do or how to live. Yet even if we concede that some epistemic norms and values are coextensive with some practical norms and values, this radically skeptical conclusion about our powers of practical reasoning does not follow. More plausibly, there is a reasonably contestable space of possible variation in the extent to which different ethical claims are vulnerable to practical debunking in light of their epistemically dubious causal pedigree. In that respect, discussions of debunking arguments in meta-ethics are relevantly similar to the discussion of the ‘right’ and ‘wrong’ kinds of reasons attributed to moral agents by different kinds of ethical theory, including the so-called “one thought-too-many” problem faced by certain versions of Consequentialism (see e.g., Williams 1981).<sup>23</sup> In light of this fact, we should expect our response to different kinds of debunking arguments in ethics to depend on our wider background commitments in moral and

---

<sup>23</sup> Compare: A: ‘Why do you love me so much?’ B: ‘I don’t know. I guess there are all sorts of reasons.’ A: ‘Yes, but what are they?’; B: ‘Why do you keep asking? After all, you know that I love you more than anything else in the world!’ In the case at issue in this chapter, the “one thought too many” applies to assumptions and beliefs about the causal pedigree of ethical claims and other aspects of those claims relating to their observance of paradigmatically epistemic norms.



political philosophy, such as our views about how much it matters not only what someone brings about, what they do to bring it about, or what they intend to bring about, but also how they came to think that they ought to bring about whatever it is that they act to bring about, and so on. In other words, it is partly a practical question about what kind of people we should aspire to be, a question that will inevitably tend to have a recognizably ethical 'feel'.

## **6. The End**

Nobody likes to be made a fool of. Not by other people, and not by Mother Nature herself (if that thought makes sense). And even someone who claims not to mind having been made a fool of will nevertheless have been made into just that: a fool. No doubt this is a fate that different people could reasonably respond to in different ways, depending on the details of their circumstances and their prior values and commitments. Yet few, if any, people would regard this fate as either an epistemic or practical aspiration or ideal.

To *be made a fool of* is not the same as *being made to look like a fool*. Thus, it is clearly possible to *be made to look like a fool* without actually *being made a fool of*, for example when it is the perpetrators of deception or ideology who are subject to ignorance or false belief, as opposed to their intended targets. It has been my suggestion in this chapter that being *made to look like a fool*, as opposed to *being made a fool of*, is what certain aspects of our ethical sensibility are most vulnerable to in the face of evolutionary debunking arguments in ethics. I have accepted that in a significant range of cases such arguments do succeed in showing that with respect to some aspects of our ethical sensibility we actually could be *made a fool of* (and, in some cases, could even be *making fools of ourselves*). What I have not accepted is that such arguments imply

that there is something foolish or otherwise dubious about ethical claims as such, regardless of our assumptions or beliefs about their evolutionary or otherwise causal pedigree.<sup>26</sup>

## References

- Audi, R. 2004. *The Good in the Right*. Oxford: Oxford University Press.
- Blackburn, S. 1998. *Ruling Passions*. Oxford: Oxford University Press.
- Blackburn, S. 2011. *Practical Tortoise Raising*. Oxford: Oxford University Press.
- Brosnan, K. 2011. “Do the Evolutionary Origins of our Moral Beliefs Undermine Moral Knowledge?,” *Biology and Philosophy* 26: 51–64.
- Clarke-Doane, J. 2020. *Morality and Mathematics*. Oxford: Oxford University Press.
- Cuneo, T. 2007. *The Normative Web*. Oxford: Oxford University Press.
- De Lazari-Rakek, K & Singer, P. 2014. *The Point of View of the Universe: Sidgwick & Contemporary Ethics*. Oxford: Oxford University Press.
- Elgin, C. 2017. *True Enough*. Cambridge, MA: MIT Press.
- Enoch, D. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- Finlay, S. 2014. *A Confusion of Tongues*. Oxford: Oxford University Press.
- Garner, R. & Joyce, R. 2019. *The End of Morality: Taking Moral Abolitionism Seriously*. London: Routledge.
- Geuss, R. 2008. *Philosophy and Real Politics*. Princeton: Princeton University Press.

---

<sup>26</sup> I am grateful to Diego Machuca, Michael Klenk, Neil Sinclair and research students at the Thursday Seminar at Birkbeck for comments and questions that have led to the improvement of this paper.

- Greene, J. 2007. "The Secret Joke of Kant's Soul." In W. Sinnott-Armstrong (ed.), *Moral Psychology. Volume 3: The Neuroscience of Morality*, 35–79. Cambridge, MA: The MIT Press.
- Greene, J. 2013. *Moral Tribes*. London: Atlantic Books.
- Haidt, J. 2012. *The Righteous Mind*. London: Penguin Books.
- Hanson, L. 2017. "The Real Problem with Evolutionary Debunking Arguments," *The Philosophical Quarterly* 67: 508–533.
- Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, R. 2005. *The Evolution of Morality*. Cambridge, MA: The MIT Press.
- Joyce, R. & Kirchin, S., (eds.). 2010. *A World without Values*. Dordrecht: Springer.
- Joyce, R. 2011. "The Error in the Error of the Error Theory," *Australasian Journal of Philosophy* 89: 519–534.
- Joyce, R. 2016. "Reply: Confessions of a Modest Debunker." In U. Leibowitz & N. Sinclair (eds.), *Explanation in Ethics and Mathematics*, 124–145. Oxford: Oxford University Press.
- Joyce, R & Garner R. (eds.) 2018. *The End of Morality: Taking Moral Abolitionism Seriously*. New York: Routledge.
- Kahane, G. 2011. "Evolutionary Debunking Arguments," *Noûs* 45: 103–125.
- Kirchin, S. 2010. "A Tension on the Moral Error Theory." In Joyce & Kirchin 2010, 167–182.
- Klenk, M. (ed.). 2019. *Higher-Order Evidence and Moral Epistemology*. London: Routledge.
- Leibowitz, U. & Sinclair, N. (eds.). 2016. *Explanation in Ethics and Mathematics*. Oxford: Oxford University Press.

- Lillehammer, H. 1999. "Normative Antirealism," *The Southern Journal of Philosophy* 37: 201–225.
- Lillehammer, H. 2003. "Debunking Morality: Evolutionary Naturalism and Moral Error Theory," *Biology and Philosophy* 18: 567–581.
- Lillehammer, H. 2004. "Moral Error Theory," *Proceedings of the Aristotelian Society* 104: 93–109.
- Lillehammer, H. 2007. *Companions in Guilt*. Basingstoke: Palgrave Macmillan.
- Lillehammer, H. 2011. "The Epistemology of Ethical Intuitions," *Philosophy* 86: 175–200.
- Lillehammer, H. 2013. "A Distinction without a Difference? Good Advice for Moral Error Theorists," *Ratio* 26: 373–390.
- Lord, E. 2018. *The Importance of Being Rational*. Oxford: Oxford University Press.
- Nagel, T. 1971. *The Possibility of Altruism*. Princeton: Princeton University Press.
- Nietzsche, F. 1887. *On the Genealogy of Morality*. Various Editions.
- Olson, J. 2014. *Moral Error Theory*. Oxford: Oxford University Press.
- Putnam, H. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Prinz, J. 2007. *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Ross, W. D. 1930. *The Right and the Good*. Oxford: Oxford University Press.
- Ruse, M. 1986. *Taking Darwin Seriously*. Oxford: Basil Blackwell.
- Ruse, M. & Richards, R. (eds.). 2017. *The Cambridge Handbook of Evolutionary Ethics*. Cambridge: Cambridge University Press.
- Sauer, H. 2018. *Debunking Arguments in Ethics*. Cambridge: Cambridge University Press.

- Sinclair, N. 2018. "Belief Pills and the Possibility of Moral Epistemology," *Oxford Studies in Metaethics* 13, 98-122.
- Sinclair, N. 2021. *Practical Expressivism*. Oxford: Oxford University Press.
- Singer, P. 2005. "Ethics and Intuitions," *The Journal of Ethics* 9: 331–352.
- Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127: 109–166.
- Streumer, B. 2017. *Unbelievable Errors*. Oxford: Oxford University Press.
- Vavova, K. 2015. "Evolutionary Debunking of Moral Realism," *Philosophy Compass* 10: 104–116.
- Williams, B. 1985. *Ethics and the Limits of Philosophy*. London: Fontana.
- Williams, B. 2002. *Truth and Truthfulness*. Princeton: Princeton University Press.
- Zimmerman, A. 2018. "Veneer Theory," In D. Machuca (ed.), *Moral Skepticism: New Essays*, 2012. New York: Routledge.