Unreliable Emotions and Ethical Knowledge

James Hutton (j.hutton@tudelft.nl)

Forthcoming in *Philosophical Quarterly*, https://doi.org/10.1093/pq/pqaf011 (early view)

Abstract

How is ethical knowledge possible? One promising answer is Moral Empiricism: we can acquire ethical knowledge through emotional experiences. But Moral Empiricism faces a serious problem. Our emotions are unreliable guides to ethics, frequently failing to fit the ethical status of their objects, so the habit of basing ethical beliefs on one's emotions seems too unreliable to yield knowledge. I develop a new, virtue-epistemic solution to this problem, with practical implications for how we approach ethical decision-making. By exploiting a frequently overlooked connection between reliability and defeaters, I argue that an agent can have a reliable belief-forming habit despite having unreliable emotions. The upshot is that emotion-based ethical knowledge is possible even for people whose emotions are unreliable, but only if we cultivate the skill of noticing and responding to signs that a given emotion is unfitting.

Keywords

moral epistemology – emotion – ethical knowledge – Moral Empiricism – virtue epistemology – reliability

1. Introduction

How is ethical knowledge possible? One answer is Moral Empiricism: we can acquire ethical knowledge through emotional experiences. Many philosophers find this view attractive.[1] Moral Empiricism identifies a noninferential route to ethical knowledge, thus avoiding the threat of an infinite regress of moral justification. But, unlike other accounts of foundational ethical knowledge, it does so without positing any mysterious psychological capacities. On the contrary, Moral Empiricism invokes the familiar suite of emotions that punctuate our ethical lives—emotions like guilt, indignation, gratitude, and contempt—with which we're all acquainted. It thus fits nicely with leading empirical accounts of moral psychology according to which emotions are the main drivers of 'intuitive' ethical belief-formation.[2] So, to many of us, Moral Empiricism looks like the best candidate for a nonsceptical moral epistemology that coheres with what we know about the human mind.

But there's a drawback. A belief-forming habit can only generate knowledge if it is reliable. However, as critics have insisted, our emotions are unreliable guides to ethical truth.[3] As I'll illustrate later, people's emotions are frequently biased or unduly swayed by situational factors. Consequently, our emotions frequently fail to fit the ethical status of their objects. It follows that if an agent were to accept all the evaluative impressions her emotions convey, she

---

[1] See Cuneo (2006), Roeser (2011), Roberts (2013: 38–112), Dancy (2014), McGregor (2015), Tappolet (2016), Milona (2016; 2017; 2023), Furtak (2018) and Hutton (2022; 2023). Historical precursors include Shaftesbury ([1711] 2001), Hutcheson ([1725] 2008), and Brentano ([1874] 1969). See Hutton (2022) for my argument that this *emotional* form of Moral Empiricism is superior to rival attempts to ground ethical knowledge in experience.

[2] For overviews, see Haidt (2012,:1–127), Greene (2013: 28–143), and Woodward (2016). This picture has been attacked by May (2018), but see Kurth (2019) and Kauppinen (2022: sec. 2.1) for convincing rebuttals.

[3] This problem is pressed by Sinnott-Armstrong (1991: 91), Szigeti (2013), Pelser (2014: 114–6), and Brady (2014: 98–101), and it's frequently raised in Q&As. Of course, there are other objections one could raise against Moral Empiricism (see Hutton 2022: 593–6), but they'll have to wait for another day.

would form lots of false beliefs. For this reason, the habit of basing ethical beliefs on emotions seems doomed to be seriously unreliable and not a means of acquiring ethical knowledge. Call this the *unreliability problem*.

In what follows, I develop a new solution to the unreliability problem. By exploiting a frequently overlooked connection between reliability and defeaters, I show how to uphold Moral Empiricism while taking seriously the unreliability of our emotions. With the right kind of attentiveness to defeaters, an agent's habit of forming emotion-based ethical beliefs will be reliable, even though her emotions are unreliable (in a sense clarified below). The upshot is that emotion-based ethical knowledge is possible even for people whose emotions are unreliable, but only if we cultivate specific skills of attention and suspension of belief.

If accepted, this result has important implications, both for the practice of ethical decision-making and for metaethics. On the practical front: if I'm right, then emotional experiences should be given a central place in our endeavours to figure out how we should live. But in order to make good use of our emotions, we must cultivate the skills of noticing signs that a given emotion is unfitting that I detail below. On the metaethical front, my argument bolsters a non-sceptical, empiricist-friendly account of how ethical knowledge is possible. I thus hope to contribute to the broader project of showing how to resist ethical nihilism without resorting to coherentism or relativism, but also without invoking rational intuition or synthetic *a priori* knowledge.

I'm not the first to suggest that emotions are subject to defeaters and that agents ought to look out for these.[4] However, as far as I can tell, I'm the first to show how attentiveness to these defeaters enhances the reliability of agents' belief-forming habits and thereby makes

[4] See Lacewing (2005), Deonna (2006: 38–41), and Milona (2016: 903–5). Compare Huemer (2008) on defeaters for putative rational intuitions.

emotion-based ethical knowledge possible. This latter point is the key to reconciling Moral Empiricism with a clear-eyed acceptance of our emotions' unreliability.[5]

Overview: Section 2 fleshes out a framework for thinking about emotion-based ethical knowledge by describing an agent with idealized emotions. Section 3 details how we fall short of this ideal, which yields a precise statement of the unreliability problem. Section 4 uses an observation about defeaters to show how, in principle, agents can acquire ethical knowledge on the basis of unreliable emotions. Section 5 identifies a range of signs of unreliability that are accessible for many agents. In doing so, it paints a psychologically realistic picture of a person who acquires ethical knowledge through her emotional experiences. Section 6 sums up the significance and the limitations of the resulting nonideal account of emotion-based ethical knowledge.

2. Preliminaries: Ideal Agents and Emotion-Based Ethical Knowledge

Moral Empiricism claims that people can acquire ethical knowledge on the basis of their emotions. Informally, the unreliability problem can be put like this: emotions seem too unreliable to count as a source of knowledge.[6] But we'll need to set up the problem more precisely to make progress. For this purpose, it will prove useful to begin with an idealized case in which an agent with perfect emotions acquires emotion-based ethical knowledge.

---

[5] Existing responses either deny that emotions are unreliable enough to generate a problem (Roeser 2011: 156); deny that knowledge requires a reliable belief-forming process (Tappolet 2016: 170–3); or apply only to agents with exceptionally reliable emotional dispositions (Cuneo 2006: 82–5; Pelser 2014: 116). I don't find those responses satisfying, so I think Moral Empiricists urgently need the new approach I develop here. Alternatively, my argument can be read as complementing those responses by showing how to reconcile Moral Empiricism with different psychological and epistemological premises.

[6] I paraphrase Pelser (2014: 114).

The agent in question is the figure of the sage (*shèng*) as envisaged in Confucian ethics. In *Analects* 2.4, Kongzi describes what his mind was like once he achieved sagehood: 'At seventy, I could follow my heart's desires without overstepping the bounds of propriety' (Confucius [c. 479BC] 2003: 9). At this final stage of ethical development, the sage's 'heart' becomes perfectly attuned to the contours of the ethical landscape. Interpreting a bit, we can think of the sage as having emotional dispositions that are perfectly aligned with the demands of ethics.[7] One way of cashing this out is to say that the sage's ethical emotions are always *fitting*. Let me explain this terminology. Each token emotion is directed at some object, which we call its *target*. Whether an emotion is fitting depends on what its target is like. Specifically, each type of emotion is paired with some evaluative property, and a token emotion is fitting if and only if (henceforth 'iff') it is directed at a target that instantiates the corresponding evaluative property. For example,

- An agent's guilt is fitting iff the deed about which she feels guilty is a wrongdoing for which she is culpable.
- An agent's admiration is fitting iff the thing she admires is excellent.
- An agent's indignation is fitting iff the deed she is indignant about is wrongful.

As these examples illustrate, some types of emotion are paired with ethical properties (whereas others are paired with non-ethical evaluative properties). Let's call such emotions *ethical emotions*.[8] Applying this framework to the sage: to say that the sage's heart is perfectly

[7] Compare Mengzi's ([c. 300BC] 2008) and Wang Yangming's ([1572] 2009) elaborations of the Confucian ideal of sagehood. Some of Aristotle's remarks about virtuous agents can be interpreted in the same way (e.g. *Nicomachean Ethics*, 1106b; Aristotle [c. 330BC] 2000: 30).

[8] Examples of emotions paired with nonethical evaluative properties might include aesthetic emotions such as amusement and prudential emotions such as fear. My argument is neutral regarding the boundary between the ethical and the nonethical; if the reader thinks any of my examples fall on the nonethical side, she can substitute

attuned to the bounds of ethical propriety is to say that, whenever she experiences an ethical emotion, she does so towards a target that exemplifies the corresponding ethical property.

Now suppose that every time the sage experiences an ethical emotion, she forms the belief, based on that emotion, that its target instantiates the corresponding ethical property. E.g. when she experiences indignation towards an action, she forms the belief that that action is wrongful. Because the sage's emotions are all fitting, the ethical beliefs she forms in this way will all be true: this habit of basing ethical beliefs on emotions will be a perfectly reliable belief-forming habit. Plausibly, this means that ethical beliefs formed in this way amount to ethical knowledge. This will be the case, for instance, if we accept a virtue reliabilist account of knowledge.[9]

Of course, most (if not all) real human beings are not sages; this is what generates the unreliability problem. But before pressing on to this issue, let me note the minimal assumptions required by the foregoing account of how emotion-based ethical knowledge is possible for the sage. Firstly, the account is pretty neutral about the nature of emotions: it is compatible with any theory on which (at least some) ethical emotions are intentional,

---

others. My preference is for a broad conception of the ethical, covering the whole sphere of *goods* and *shoulds* relating to how we should live. This contrasts with the narrower sphere some authors call the 'moral' (e.g. Williams 1986; D'Arms and Jacobson 2023), hence my avoidance of the more loaded term 'moral emotions'.

[9] See Sosa (1991; 2017), discussed further below. This suggestion is bolstered by various views according to which the content of emotions enables them to justify evaluative beliefs. For instance, many theorists argue that experiencing an emotion towards *x* involves it seeming to one that *x* instantiates the corresponding evaluative property (e.g. Roberts 1988: 190–5; Mitchell 2021: 30–69) and that this makes the emotion a suitable epistemic basis for the corresponding evaluative belief (Tolhurst 1990: 85–6; Kauppinen 2013: 375–7). Since the issue of whether the content of emotion supports Moral Empiricism is orthogonal to the issue of reliability, I won't say anything more about it in this article. For further discussion, see Deonna and Teroni (2012: 118–25), Brogaard and Chudnoff (2016), Mitchell (2017), and Harrison (2021).

nondoxastic states with ethical fittingness conditions, e.g. perceptual theories, feeling-towards theories, attitudinal theories.[10] Secondly, the account is pretty neutral about the ontology of value: realists (whether naturalist or nonnaturalist; robust or relaxed) can obviously accept my talk of fitting emotions; reliable habits of ethical belief-formation; and ethical knowledge. But so can sophisticated quasi- or anti-realists. For example, if ethical discourse supports a minimalist truth predicate, then the quasi-realist can easily earn the right to talk about some emotions being fitting and others unfitting; some habits of forming ethical (quasi-)beliefs being reliable and others unreliable; some habits resulting in ethical (quasi-)knowledge, others not.[11] Thirdly, nothing about the sage's capacities clashes with a naturalistic understanding of the human mind. We can give an unmysterious explanation of how the sage's emotional dispositions operate: the ethical properties of objects supervene on their nonethical properties, and our emotions are differentially responsive to nonethical information about their targets. So long as the sage's emotional dispositions respond to nonethical information in ways that match the patterns of determination linking the nonethical base properties and the supervenient ethical properties, her emotions will fit the ethical properties of their targets. Indeed, this is exactly how we should explain our own limited capacity to respond to objects with emotions that match their ethical properties.[12] The sage doesn't have any mysterious

[10] For these theories of emotion, see respectively Tappolet (2016), Mitchell (2021), and Deonna and Teroni (2012). I should note that, *pace* Müller (2019), I'm assuming that emotions do not require a pre-emotional awareness of the corresponding evaluative property (see Mitchell 2019a for discussion).

[11] See Blackburn (1996: 86–8) and Sinnott-Armstrong (2011: 289) on quasi-realist reliability and ethical knowledge. For a sophisticated antirealist account of a different kind, compatible with everything I say in this article, see Wiggins (1991).

[12] Compare McBrayer's (2010) account of how moral perception could reliably track the ethical properties of its objects.

psychological capacities, which ordinary human beings lack; it's just, as we'll see in a moment, that the sage's emotional dispositions are unrealistically perfect.

3. Our Emotions Are Unreliable

We've seen how a sage can obtain ethical knowledge through her emotional experiences. But most, if not all, of us fall short of this ideal of sagehood. Unlike the sage, we experience unfitting ethical emotions, and this is far from a rare occurrence. I'll now spell out why this is the case, which leads to a precise statement of the unreliability problem.

One reason we experience unfitting emotions is that, unlike the sage, our patterns of emotional sensitivity don't perfectly match the patterns of nonethical-to-ethical determination that make up the ethical landscape. Someone raised in a homophobic society will likely grow up to experience unfitting negative emotions towards gay people. For instance, they will tend to experience indignation towards public displays of affection between gay people, even though these acts aren't wrongful. Since we all grew up in societies that are ethically flawed to one degree or another, it's reasonable to suppose that none of us have acquired emotional sensitivities that match the ethical landscape perfectly—a supposition which is further supported by evidence of how prevalent implicit bias is in our cultures.[13]

Another reason we experience unfitting emotions is that our emotional dispositions are inherently noisy. Whether you experience indignation towards a certain remark will partly depend on the remark's ethically relevant features (e.g. the content, context, and

[13] See, e.g. Brownstein (2018: 2) on the prevalence of implicit bias (though see Machery (2022) for some potential problems with this research paradigm). Baron *et al.* (2014: Experiment 1) illustrates the impact of implicit bias on ethical judgments. It's common among psychologists and philosophers to construe implicit bias as a partly 'affective' phenomenon (Banaji and Heiphetz 2010; Gendler 2011; Brownstein 2018). Since the 'affect' in question is directed at an external object, it's plausible that being implicitly biased against a group involves being disposed to experience unfitting negative emotions towards them.

consequences of what was said); but your emotional response will also frequently be influenced by situational factors that are irrelevant to the remark's ethical status. Such factors include your recent run of positive or negative emotions: if you're having a bad day, you're more likely to be in an irritable mood and become indignant about an innocuous remark. Other situational factors that affect our ethical emotions seem to include being in the grips of a hangover;[14] being bombarded with irritating sounds;[15] and perhaps even smelling foul odours.[16] Plausibly, all these factors work by changing our moods,[17] as a result of which our emotional dispositions get 'temporarily put "out of tune"'.[18] While it's possible, on occasion, for these situational factors to put an agent in just the right mood to respond with a fitting emotion to the situation at hand, the general tendency of noise in any system of human judgment is to reduce its reliability.[19] Consequently, the influence of these situational factors makes unfitting emotions more prevalent.

These two effects—viz. shortcomings due to imperfect patterns of ethical sensitivity and noise due to situational influences—work cumulatively to reduce the reliability of our

---

[14] See Fjær (2015) and Milton *et al.* (2019). All empirical evidence cited here should be treated as provisional, owing to the replication crisis in social psychology.

[15] See Seidel and Prinz (2013); compare Mathews and Canon (1975).

[16] See Schnall *et al.* (2008, Experiment 1) and Inbar *et al.* (2012). The validity of the 'incidental disgust' research programme has been questioned (e.g. May 2014), but the effect of disgusting smells in particular seems to hold up. As one influential meta-analysis puts it, 'gustatory and olfactory disgust inductions exert a reliable, small- to medium-sized effect on moral judgments' (Landy and Goodwin 2015: 529).

[17] On the influence of recent emotions on mood, see Kontaris *et al.* (2020); on hangovers' influence on mood, see McKinney and Coyle (2006) and Penning *et al.* (2012); on sound's influence on mood, see Smith *et al.* (1997); on odour's influence on mood, see Herz (2009) and Kontaris *et al.* (2020, 7).

[18] This phrase is from Goldie (2004: 257).

[19] See Kahneman *et al.* (2021: chap. 2): 'In noisy systems, errors do not cancel out. They add up.'

emotions.[20] Therefore, it's hard to deny that most, if not all, humans experience unfitting ethical emotions fairly frequently. If, like the sage described above, you or I were to form the corresponding ethical belief every time we experienced an ethical emotion, we would form lots of false beliefs. This means that our habit of forming emotion-based ethical beliefs wouldn't be reliable and consequently wouldn't be a means of acquiring ethical knowledge. This, in detail, is the unreliability problem.

At this point, one possible move for would-be defenders of Moral Empiricism is to claim that, for all that's been said, there might still be real-life sages among us. There are interpersonal differences in how well people's emotions match the ethical landscape. Not all societies are equally plagued by bias. And in any case, individuals' emotional dispositions can be reshaped through habituation, psychotherapy, and ethical self-cultivation. So, there might be some exceptional people whose hearts match the bounds of propriety perfectly.[21]

But this way of defending Moral Empiricism seems perilous at best. It's a tall order to make a convincing case that any real-life agent's ethical emotions are perfect (*pace* traditional views about Kongzi and other sages). Theorists in both psychology and philosophy of emotion present the influence of mood as a ubiquitous, structural feature of humans' emotional dispositions,[22] so the burden of proof lies with anyone who wants to claim that some people's emotions are immune to influence by situational factors. It thus seems unwise to tie the fate of Moral Empiricism to the existence of sages. Moreover, it's obvious that most

[20] In epistemology, 'reliability' refers to the proportion of true beliefs a belief-forming process produces. We can extend the notion to emotions as follows: let the reliability of an emotional disposition be the proportion of fitting emotions the disposition produces. Thus, to say that an agent's emotions (or, more precisely, emotional dispositions) are unreliable is to say that the proportion of fitting emotions they produce is lower than some relevant threshold.

[21] Compare Cuneo (2006: 82–85) and Pelser (2014: 116).

[22] E.g. Deonna and Teroni (2012: 104–5), Kontaris *et al.* (2020).

of us aren't sages, so Moral Empiricism loses much of its interest and appeal if it credits only sages with first-hand (i.e. non-testimonial) ethical knowledge. So, can Moral Empiricism be reconciled with a clear-eyed view of the unreliability of our emotions? I believe it can.

4. Reliable Beliefs Based on Unreliable Emotions

From here on, I'll develop my solution to the unreliability problem. First, I'll describe how, in principle, an agent whose emotions are unreliable can nevertheless have a reliable habit of forming emotion-based ethical beliefs. Then I'll use reliabilist virtue epistemology to argue that many beliefs formed in this way amount to ethical knowledge. Finally, in the next section, I'll argue that this isn't just possible in principle, but feasible for humans.

As mentioned, my solution exploits a frequently overlooked connection between reliability and defeaters. It's commonplace, across all domains of belief, that agents are capable of treating certain cues as defeaters. One treats a cue as a defeater when one responds to it by refraining from forming a belief one would otherwise have formed. (One also treats a cue as a defeater when one responds to it by relinquishing an already held belief, but our focus is on belief-formation rather than belief-maintenance.) Although the idea of responding to defeaters is commonplace, it's less commonly observed that treating cues as defeaters can increase the reliability of one's belief-forming habits. Let's illustrate this important fact with a simple numerical model.

NIGHT VISION: Amina and Saira both have good eyesight under normal conditions, but bad night vision. 99% of their visual experiences are veridical during daytime, but only 50% of them are veridical during nighttime. (Let's assume that Amina and Saira both have an equal number of visual experiences in daytime and nighttime however visual experiences are to be individuated.) Every time Amina has a visual experience, she believes what she sees. In contrast, Saira believes her visual experiences during daytime but withholds judgment when

it's nighttime—she treats the presence of nighttime as a defeater. How do things go for each agent in the course of experiencing a typical run of 1,000 visual experiences? Amina has 495 veridical and 5 nonveridical visual experiences during daytime, and 250 veridical and 250 nonveridical during nighttime. So, she forms 1,000 beliefs of which 745 are true: her belief-forming habit is 74.5% reliable. Saira has the same mix of veridical and nonveridical experiences, but she treats the presence of nighttime as a defeater, so she forms 500 beliefs of which 495 are true: her belief-forming habit is 99% reliable.

By treating this cue as a defeater, Saira filters out all the visual experiences that occur during nighttime. The proportion of veridical experiences among the undefeated ones is higher than that in her total pool of experiences. Consequently, Saira's belief-forming habit is 99% reliable even though the experiences on which the beliefs are based are, in a clear sense, only 74.5% reliable. The takeaway: if there are cues that are more likely to accompany a visual experience if it is non-veridical than if it is veridical and the agent treats these cues as defeaters, then her belief-forming habit will be more reliable than it would otherwise have been. Her vision-based belief-forming habit becomes reliable, even though her vision remains, in a clear sense, unreliable (see fig. 1).

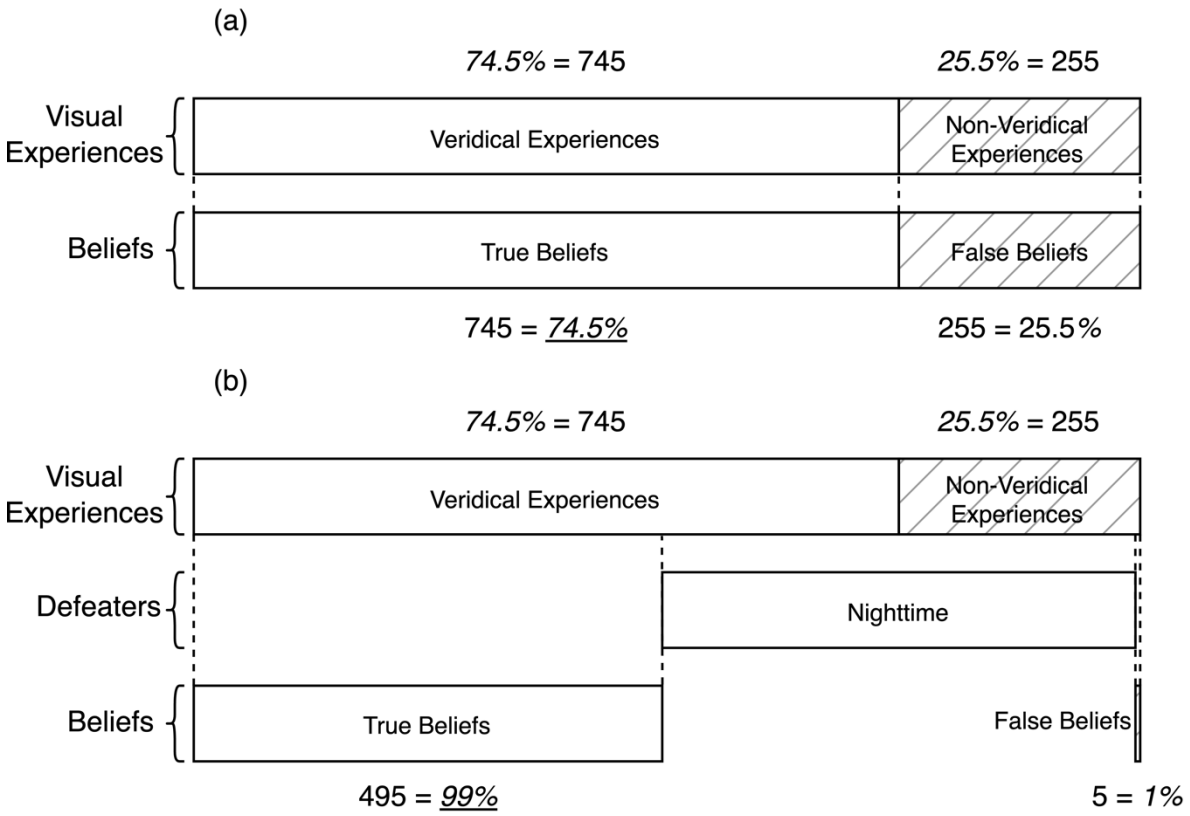

Figure 1. Amina's habit of forming vision-based beliefs while treating the presence of night-time as a

This result has a clear corollary for emotion-based beliefs: if there are cues that are more likely to accompany an emotion if it is unfitting than if it is fitting and the agent treats those cues as defeaters, then she will form a lower proportion of false emotion-based beliefs than she otherwise would have. This improves the reliability of her habit of forming emotion-based beliefs, while holding fixed the reliability of her emotional dispositions. It follows that, in principle, an agent who has unreliable emotions can have a reliable habit of forming emotion-based ethical beliefs. Specifically, this will be the case when:

(i) a sufficient proportion of the agent's unfitting emotions are accompanied by accessible cues,

(ii) which she treats as defeaters, such that

(iii) the proportion of fitting emotions among her pool of undefeated emotions equals or exceeds the proportion of true beliefs a belief-forming habit must produce in order to count as reliable.[23, 24]

(By an 'accessible' cue, I mean one it is psychologically possible for the agent to treat as a defeater. Plausibly, this means any cue it is psychologically possible for the agent to attend to.)

Let's use the phrase *moderately virtuous agent* to denote any agent whose unfitting emotions are adequately covered by accessible cues (in the sense that there is some set of accessible cues such that if she were to treat them as defeaters, her belief-forming habit would be reliable). For a moderately virtuous agent who possesses the attentional skill of noticing the relevant cues and the doxastic skill of suspending judgment when they are present, her habit of basing ethical beliefs on emotions will be reliable.

It's highly plausible that the emotion-based beliefs of a moderately virtuous agent with these skills amount to ethical knowledge, even if her emotions are pretty unreliable. Indeed, this agent's habit of forming emotion-based beliefs (by definition) meets the same threshold of reliability that applies to canonical ways of acquiring knowledge such as perception and testimony. Obviously, when we form beliefs based on, e.g. visual experience or testimony, we sometimes go wrong. But epistemologists have developed accounts of how, when all goes well, beliefs formed in these fallible ways can amount to knowledge. Many accounts are available,

[23] What proportion of true beliefs is required for reliability? Epistemologists tend not to give a precise figure, with some arguing that the threshold varies with context or subject matter. I'll follow the practice of leaving the threshold undefined. The structure of my argument is unaffected by the choice of threshold (though Moral Empiricism is easier to defend the lower it is).

[24] The sage's capacity for ethical knowledge turns out to be an edge case of this more general formula. Since all the sage's emotions are fitting, it's trivially true that a sufficient proportion of her undefeated emotions are fitting to meet the threshold of reliability.

but one attractive account is *virtue reliabilism*. On this view, so long as the belief-forming habit meets a certain threshold of reliability (which epistemologists typically do not quantify explicitly), it counts as an *epistemic virtue*. Whenever an agent forms a true belief via one of these reliable habits, and the fact that she reached a true belief rather than a false one is best explained by the fact that the habit in question is a reliable one, the belief amounts to knowledge. Crucially, this account of how to reconcile knowledge with fallibility can be applied to the moderately virtuous agent's emotion-based beliefs: by treating the relevant cues as defeaters, the agent transforms her habit of forming ethical beliefs based on emotions into an epistemic virtue. The result is that, when an agent's emotion-based ethical belief is an 'apt' manifestation of this epistemic virtue, it is an item of *knowledge*.[25]

This is a highly significant result. Due to her responsiveness to defeaters, the moderately virtuous agent's habit of forming emotion-based ethical beliefs is reliable even though her emotions are unreliable, and reliable belief-forming habits generate knowledge. Therefore, Moral Empiricism is compatible with our emotions' being unreliable. Q.E.D.

But nothing I've said so far indicates that any of *us* are moderately virtuous agents. Thus, although I have now shown how *in principle* an agent with unreliable emotions can acquire emotion-based ethical knowledge, more needs to be said to show that *we* can do so. Thus, my next task is to identify a range of cues that are accessible for us and that correlate with unfittingness.

5. Defeaters for Ethical Emotions

I have demonstrated that if enough of an agent's unfitting ethical emotions are accompanied by accessible cues correlating with unfittingness, then she can achieve emotion-based ethical

[25] See Sosa (1991; 2017). A similar story could be told for other putative necessary conditions on knowledge that involve reliability or counterfactual robustness. See also the discussion of justification in fn. 9.

knowledge by treating those cues as defeaters. But are there sufficient cues, correlating with unfittingness, that are accessible *for us*? This is an empirical matter and there will be variation from person to person. There is no guarantee that everyone's emotions will be adequately covered by accessible cues. Nonetheless, I believe there are a wide range of accessible cues that, for many of us, will be more likely to accompany unfitting than fitting emotions.

5.1 Negative Metaemotions

First, consider metaemotions. A metaemotion is an emotion, the target of which is another emotion.[26] For example, I might feel indignant about a remark someone has made, but also feel embarrassed about getting indignant. Treating negatively valenced metaemotions as defeaters for first-order ethical emotions is something most of us probably do already: if you feel embarrassed to find yourself getting indignant, you're less likely to endorse the evaluative impression the indignation conveys. Moreover, it's plausible that, for many of us, negative metaemotions are more likely to accompany unfitting first-order emotions than fitting ones. Admittedly, there is no reason to expect our metaemotions to be *more* reliable than our first-order emotions, but there is no reason to think they are *less* reliable either. In the same way that our first-order emotions intelligently combine multiple streams of information to produce holistic evaluative impressions,[27] our metaemotions are likely to be sensitive to a range of factors that bear on the appropriateness of the first-order emotion. So, if we think our emotions are fitting more often than not, we should probably expect our metaemotions to be fitting more often than not. This means that negative metaemotions will be more likely to accompany an emotion if it is unfitting than if it is fitting, which is all that's required for them

[26] See Howard (2017) and Mitchell (2019c). Using ideas from psychoanalysis, Lacewing (2005) proposes treating some metaemotions as defeaters for emotions (though he doesn't connect this with the unreliability problem).

[27] See Allman and Woodward (2008) and Railton (2014).

to be reliability-boosting defeaters. Indeed, somewhat counterintuitively, someone with fairly unreliable emotions and similarly unreliable metaemotions can get a significant boost in the reliability of her belief-forming habit by treating the latter as defeaters. Let's explore this with a simple numerical model.

AMINA'S METAEMOTIONS: Amina has fairly unreliable ethical emotions: 80% of her ethical emotions are fitting but 20% are unfitting. Some of her ethical emotions, one in five to be precise, are accompanied by negative metaemotions. These metaemotions are fairly unreliable too: 80% of them accompany unfitting first-order emotions, while 20% of them are misleading, in the sense that they are directed at fitting first-order emotions. Amina treats her negative metaemotions as defeaters. The upshot is that, for every 1,000 first-order ethical emotions Amina experiences, 800 are fitting, of which 40 are accompanied by negative metaemotions, with the result that she forms 760 true beliefs; and 200 are unfitting, of which 160 are accompanied by negative metaemotions, with the result that she forms 40 false beliefs. So, Amina forms 800 emotion-based beliefs, of which 760 are true: her belief-forming habit is 95% reliable.

By treating negative metaemotions as defeaters, Amina filters out all the emotions that are accompanied by this cue. The remaining pool of undefeated emotions is much more reliable than her total pool of emotions. Amina's metaemotions are no more reliable than her first-order emotions—both levels of emotion get things wrong one in five times. Nevertheless, adopting the habit of treating negative metaemotions as defeaters, Amina increases the

reliability of her emotion-based beliefs from 80% to 95% (see fig. 2).

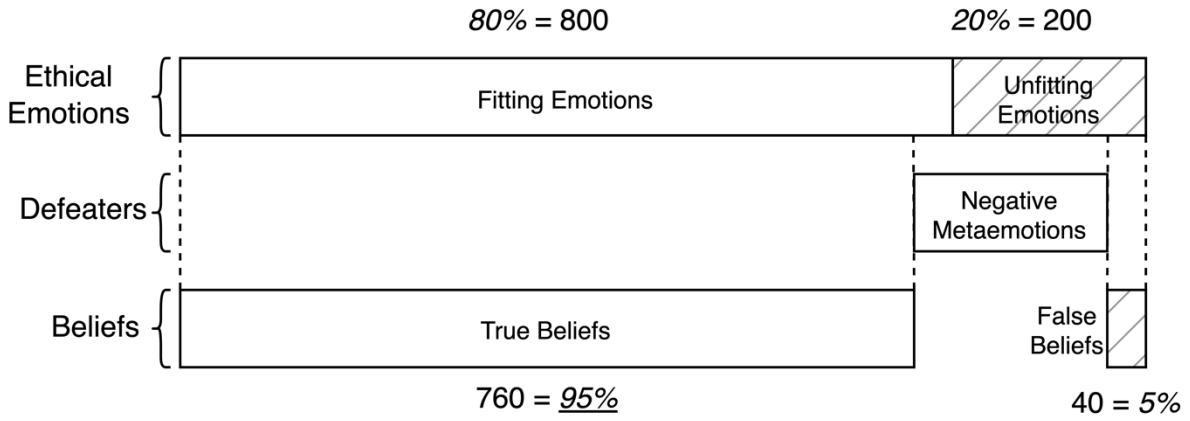

Figure 2. Amina's habit of forming emotion-based ethical beliefs, while treating negative metaemotions as defeaters.

5.2 Epistemic Feelings

Secondly, consider epistemic feelings, such as feelings of uncertainty, hesitancy, or doubt.[28] For example, one might feel indignant about someone's remark but simultaneously feel a sense of uncertainty about whether the remark really constitutes a wrongdoing. As with negative metaemotions, it seems like common sense to treat negative epistemic feelings as defeaters, and most of us probably do this already. It's also plausible that, for most agents, a negative epistemic feeling will be a fallible but useful sign that one should withhold judgment; it's reasonable to expect epistemic feelings to be roughly as reliable as our emotions. Once again, there are no guarantees here. Agents will differ in the extent to which their unfitting ethical emotions are accompanied by feelings of uncertainty, etc., and some agents may have thoroughly misleading epistemic feelings. Nevertheless, agents whose epistemic feelings are at least somewhat reliable will be able to increase the reliability of their belief-forming habits still further by treating them as defeaters.

What is the cumulative effect of treating both negative metaemotions and epistemic feelings as defeaters? So long as (i) some emotions that *aren't* accompanied by metaemotions

[28] See, e.g. Arango-Muñoz (2014) and Carruthers (2017).

are accompanied by epistemic feelings and (ii) these epistemic feelings are more likely to accompany an emotion if it is unfitting than if it is fitting, then treating epistemic feelings as defeaters will provide a further boost in reliability. To illustrate, let's make our numerical model more elaborate.

AMINA'S METAEMOTIONS AND EPISTEMIC FEELINGS: Amina's psychological profile is as described in AMINA'S METAEMOTIONS, with the addition that she has moderately reliable epistemic feelings. One in five of her first-order emotions is accompanied by a negative epistemic feeling. 80% of those epistemic feelings accompany unfitting emotions while 20% of them misleadingly accompany fitting emotions. There is a lot of overlap between Amina's epistemic feelings and her metaemotions, but the correlation isn't perfect: one in five of her negative epistemic feelings accompanies an emotion towards which she doesn't experience a negative metaemotion. Amina treats both negative metaemotions and negative epistemic feelings as defeaters. The upshot is that, for every 1,000 first-order emotions she experiences, 800 are fitting, of which 40 are accompanied by negative metaemotions, 40 are accompanied by negative epistemic feelings, and 48 are accompanied by either a negative metaemotion or a negative epistemic feeling. So, she forms 752 true beliefs. Meanwhile, of her 200 unfitting ethical emotions, 160 are accompanied by negative metaemotions, 160 are accompanied by negative epistemic feelings, and 192 are accompanied by either a negative metaemotion or a negative epistemic feeling. So, she forms 8 false beliefs. In sum, Amina forms 760 emotion-based beliefs, of which 752 are true: her belief-forming habit is 98.9% reliable.

By attending to her negative epistemic feelings, Amina winnows out more unfitting emotions. She also erroneously winnows out some fitting emotions, but since her epistemic feelings are more likely to accompany an emotion if it is unfitting than if it is fitting, the effect is to increase the proportion of true beliefs formed (see fig. 3).

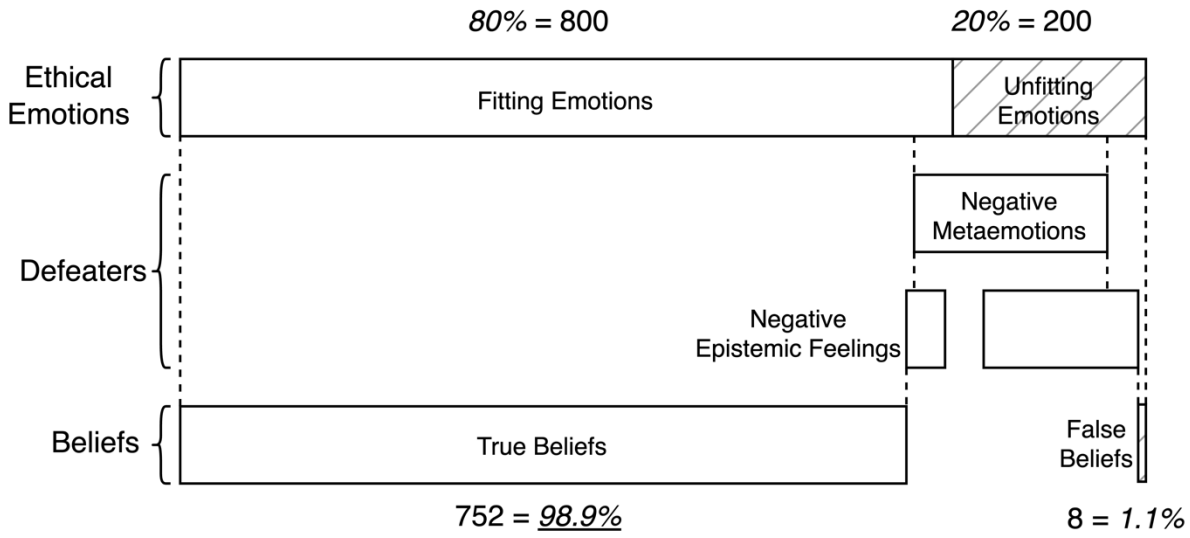

Figure 3. Amina's habit of forming emotion-based ethical beliefs, while treating negative metaemotions and epistemic feelings as defeaters.

With these highly fallible signals of unfittingness layered one on top of the other, our agent's reliability creeps up still further. This cumulative effect of fallible defences has been summed up nicely in another context with the image of 'slices of Swiss cheese being layered on top of one another, until there [are] no holes you [can] see through.'[29] It isn't a problem that each slice has lots of holes in it; as we add more layers, the gaps close.

Arguably, the agent in our model has now reached the point where 'apt' manifestations of her belief-forming habit amount to knowledge. Compare a gardener who misidentifies flowers one or two times in 100. Wouldn't we say that, when all goes well, she *knows* that this flower is a fuchsia? If so, and if Amina's psychological profile is attainable for real human beings, then we have reached a very significant point in our discussion: this is a psychologically realistic account of an agent who is far from sage-like, but who *knows* that certain things are wrong (etc.) on the basis of her emotions. But there are still other relevant cues we can point to, meaning there are still other ways to reach the threshold of being a moderately virtuous agent.

[29] See Lewis (2021: 76). The original context is pandemic mitigation.

5.3 Rationally Unintelligible Moods

The next kind of cue is provided by moods that aren't rationally intelligible. Sometimes we experience our moods as intelligible responses to how things are going.[30] Lixin is backpacking in a foreign country, and he has experienced all kinds of mishaps—he's been mugged; people have tricked him out of money; strangers have been rude and dismissive. This puts him in a certain mood: he feels ill at ease. But as well as feeling ill at ease, Lixin is aware of various features of his environment in virtue of which it makes sense for him to feel ill at ease, namely the aforementioned mishaps. Consequently, Lixin experiences his mood as an intelligible response to the hostility of his present environment. Contrast this with cases in which one's mood is *not* experienced as rationally intelligible. When I just wake up feeling unaccountably irritable or find myself growing irritable due to the incessant whine of my neighbour's lawnmower, there is nothing I'm aware of in virtue of which it makes sense for me to approach my environment with a diffuse sense of irritation. Lixin's mood is rationally intelligible; mine is rationally unintelligible.

In many cases, a rationally unintelligible mood will serve as a fallible but useful indicator that one's emotions are less likely than usual to be fitting. So long as the pattern of mishaps that have made Lixin feel ill at ease are manifestations of a genuine tendency in his environment (rather than just a run of bad luck), the shift in emotional dispositions brought on by his mood will increase his tendency to experience fitting emotions. For instance, he will be more likely to view a hostel roommate's ambiguous behaviour with suspicion, right at a time when he finds himself in an environment where suspicion is more likely to be fitting. In contrast, my rationally unintelligible mood, brought on by the sound of a lawnmower or some other situational factor, will simply put me out of tune with the ethical landscape. I'm more

[30] See Mitchell (2019b: 127–32). Here, I have in mind what Mitchell calls the 'normative sense' of intelligibility, as opposed to 'merely causal' intelligibility.

disposed to experience indignation, but there's nothing about my environment that makes genuinely wrongful actions more likely. In other words, the unintelligible mood is one in which I'm at an increased risk of having an unfitting emotion. It follows that, so long as one's sense of the intelligibility of one's mood is moderately reliable, the presence of a rationally unintelligible mood will serve as a useful though fallible signal that one's emotions are more likely than usual to be unfitting.

It's part of folk wisdom about the limitations of emotion that certain moods put our emotions temporarily out of tune, and thus render us ill-equipped for making ethical judgments or other important evaluative decisions. So, it's plausible that many of us are already in the habit of opting to sleep on it or mull things over more when we notice that we are in moods that don't make rational sense from the inside. On the other hand, some of us might need to develop the attentional skill of noticing such moods in order to make full use of this kind of cue, e.g. through familiar forms of self-cultivation such as mindfulness meditation. Either way, unintelligible moods add to the growing list of cues that it's feasible and advisable to treat as defeaters.

I won't run any more numbers, but as long as, (i) Amina's sense of being in an unintelligible mood is more likely to accompany an unfitting than a fitting emotion and (ii) it accompanies some unfitting emotions that aren't accompanied by negative metaemotions or epistemic feelings, she will be able to increase the reliability of her belief-forming habit still further by treating it as a defeater.

5.4 Conflicting Beliefs

Another important kind of cue is provided by clashes with existing beliefs. One kind of case is exemplified by, e.g. feeling indignant about some piece of conduct while seeing no relevant

difference between this and another piece of conduct you believe to be innocuous.[31] Another kind of case is exemplified by, e.g. feeling indignant about some piece of conduct while believing that your emotions aren't reliable in this context because the agent belongs to a group you suspect you're emotionally biased against.[32]

Once again, it's commonsense to treat such clashes as defeaters and there's reason to think most humans are skilled at noticing inconsistencies between the verdicts suggested by their emotions and their background beliefs.[33] Moreover, by the time we reach adulthood, most of us possess a large stock of ethical beliefs about particular cases, plus some generalizations about the *prima facie* ethical properties of various types of action and situation. As long as one's background ethical beliefs are moderately accurate, an emotion will be more likely to conflict with them if it is unfitting than if it is fitting. So, for most of us, the ability to treat these clashes as defeaters will enhance the reliability of our emotion-based beliefs. Of course, there are no guarantees here. An obvious exception would be anyone who has been brought up to believe a pervasively defective moral ideology; here as elsewhere, bad ideology can impede one's ability to acquire knowledge.

We might wonder where an agent could have got this background web of true ethical beliefs against which to check her emotions. On a modest version of Moral Empiricism, according to which emotion is just one of several routes to noninferential ethical knowledge, one suggestion would be that these true ethical beliefs trace back to some non-emotional source, such as rational intuition. More interestingly, on an ambitious version of Moral Empiricism according to which the whole superstructure of ethical knowledge rests on emotions, an agent's background ethical beliefs would stem from past emotions, plus chains of

[31] See Campbell and Kumar (2012).

[32] Compare Milona (2016: 903–5).

[33] For a general account of how we maintain coherence in thought, see Thagard (2000: 15–40).

reasoning and/or testimony tracing back to emotion. On this latter picture, a kind of bootstrapping would be possible, in which an initial batch of moderately unreliable ethical beliefs enables one to filter out some of one's unfitting emotions, leading to a less unreliable second batch of ethical beliefs, and so on in a virtuous circle (virtuous so long as, from the first batch onwards, the beliefs clash more often with unfitting emotions than with fitting ones).

5.5 Social Cues

Lastly, let me note another rich source of defeaters that is normally available to us when we're forming ethical judgments: feedback from others. We often make ethical judgments and decisions in contexts of social interaction, and we're highly attuned to our interlocutors' reactions to what we are saying and to the emotions we express. Consequently, in addition to the introspectable cues described up to this point, we often have access to a range of social cues. These social cues include the emotions your interlocutor expresses (verbally and nonverbally) towards the situation under consideration; the metaemotions she expresses towards your emotions; the epistemic feelings she expresses towards the thoughts you put forward; her sense of whether your present mood is rationally intelligible; and her beliefs about the fittingness of your emotions and the tenability of the ethical judgments you're considering. So long as your interlocutor's emotions, metaemotions, etc. are moderately reliable, treating these signals as defeaters will bring a further boost in reliability.

Having access to your interlocutor's responses is valuable even if your interlocutor's emotions, metaemotions, etc., are no more reliable than your own. Your interlocutor simply has a different perspective on the situation from you, both literally and figuratively, so there will be cases in which something about the situation leads you to experience an undefeated unfitting emotion but doesn't have the same effect on your interlocutor. Suppose that for any given unfitting emotion, you and your interlocutor each have an independent 95% chance of experiencing one of the previously mentioned defeaters. In that case, the chance that at least

one of you experiences such a defeater rises to 99.75%. In this situation, if you treat your interlocutor's conflicting emotions, metaemotions, etc. as defeaters as well as your own, you will hardly form any false emotion-based beliefs.

The dynamics of such emotional-ethical exchanges are a ripe topic for future research.[34] This is a situation in which we make epistemic use of non-linguistic as well as linguistic information from others. Moreover, if treating your interlocutor's reactions as defeaters is necessary for reaching the threshold of reliability, then we have an interesting new case in which the acquisition of knowledge—not just its testimonial transmission—is an inherently social endeavour.[35] For our purposes, the key takeaway is that it is very plausible that cues from one's interlocutor make it dramatically more feasible to elevate the reliability of one's belief-forming habit past the threshold required for knowledge.

This concludes my (provisional, incomplete) survey of the cues that, plausibly, enable us to improve the reliability of our emotion-based beliefs if we treat them as defeaters. In the final section, I'll sum up the nonideal version of Moral Empiricism that has emerged, along with its limitations.

6. Conclusion: Nonideal Moral Empiricism

The unreliability problem for Moral Empiricism was this: most, if not all, human beings have unreliable ethical emotions. Consequently, if every time we experience an ethical emotion, we

[34] The resulting socialized version of Moral Empiricism aligns with the emotion-centric methodology championed by feminist advocates of consciousness raising in the 1960s. As Kathie Sarachild put it in her address to the First National Women's Liberation Conference in 1968: 'We assume that our feelings are telling us something from which we can learn... that our feelings mean something worth analysing... that our feelings are saying something political. [...] In our groups, let's share our feelings and pool them. [...] Our feelings will lead us to ideas and then to actions' (1969: 78).

[35] Compare Levy's (2007) notion of 'radically socialized knowledge'.

form the corresponding ethical belief, we will form lots of false beliefs. Moreover, any true beliefs we happen to form in this way will not amount to knowledge, because of the overall unreliability of the belief-forming habit. My response has been to accept that our emotions are unreliable, but to exploit the possibility of developing more nuanced epistemic habits. I have shown that, by attending to cues that correlate with our emotions' being unfitting, we can increase the reliability of our emotion-based beliefs. Given the wide range of cues I have identified, from metaemotions to social feedback, it is highly plausible that many ordinary agents can elevate their habit of forming emotion-based ethical beliefs into a reliabilist epistemic virtue. The upshot is that many of the manifestations of this way of forming beliefs will be items of noninferential ethical knowledge—a highly significant result for moral epistemology.

While this account constitutes an important vindication of Moral Empiricism, it is not without limitations. Firstly, the epistemological model I have developed only applies to agents who succeed in developing the attentional skill of noticing the cues in question when they are present and the doxastic skill of treating them as defeaters. Although I've given reasons for thinking that this requirement can be met by ordinary human beings and that it is far less demanding than the ideal of sagehood, I make no claim that all agents will meet it. The emotion-based ethical beliefs of agents who do not regularly treat the relevant cues as defeaters will not amount to knowledge. Secondly, the reliability of an agent's belief-forming habit is contingent on the availability and accessibility *for that agent* of a range of cues that correlate with unfittingness. There will certainly be agents whose unfitting emotions are insufficiently covered by accessible cues. Such agents will be incapable of acquiring emotion-based ethical knowledge, no matter how skilful they are at attending to defeaters. Finally, the model assumes a baseline level of emotional reliability. This will not be met in extreme cases where an agent's emotions are severely misaligned with the ethical truths due to wide-ranging

biases and/or habituation into a seriously defective pattern of ethical-emotional sensitivity. Again, my conclusion is that Moral Empiricists must bite the bullet and concede that such agents cannot acquire emotion-based ethical knowledge.

I should note that, if there were compelling arguments for certain radically revisionary first-order views such as act-utilitarianism, or metaethical views such as nihilism, then we would probably have to conclude that the ethical emotions of every human being are systematically misleading to an extent that is beyond repair by attention to defeaters.[36] In this article, I have said nothing to defend Moral Empiricism against the kind of radical unreliability entailed by such views. All I can say here is that, by defending Moral Empiricism against the (nonradical) unreliability problem, I have indirectly strengthened the case against those revisionary views. For example, the moral epistemology I've defended gives us reason to take seriously the emotional intuitions that conflict with act-utilitarianism, and it undermines the core epistemological arguments for nihilism. However, it is a task for another day to defend Moral Empiricism against substantive arguments meant to show that our emotions are radically unreliable, e.g. evolutionary debunking arguments, and I make no claim to have defended Moral Empiricism against such challenges in this article.

Despite these limitations, the nonideal form of Moral Empiricism developed here is, I contend, philosophically significant and practically important. I have (partially) vindicated an account of how noninferential ethical knowledge is attainable for ordinary human beings, an account which takes our embodied, emotional encounters with value and disvalue as its starting point. The account also offers action-guiding advice for how to approach ethical decision-making: our efforts to reach ethical decisions should start from our emotions, but these emotions must be filtered using a range of signs of unfittingness, including negative metaemotions, unintelligible moods, and clashes with existing beliefs. Moreover, if my

[36] See Singer (2005).

observations about social feedback are correct, then emotion-based ethical knowledge is inherently communal in nature. The account thus points to a collaborative and inclusive approach to ethical inquiry, with different agents working interdependently to mutually enhance the reliability of their emotion-based ethical beliefs. By reconciling the requirements for ethical knowledge with the actual state of our emotional dispositions, I hope to have provided a practical and realistic account of the epistemological foundations of ethical inquiry.[37]

TU Delft, The Netherlands

[37] This article has been evolving for quite a few years and has benefitted enormously from feedback from audiences at Cambridge, UCL, CEU, Sheffield, Edinburgh, and Utrecht, as well as written comments from Alix Cohen, James Laing, Maxime LePoutre, Michael Milona, Norbert Paulo, and Paulina Sliwa. Thanks also to the many anonymous reviewers who gave recommendations for improvement. I gratefully acknowledge funding from the AHRC (doctoral award) and the Leverhulme Trust (ECF-2020-289).

References

Allman, J. and J. Woodward (2008) 'What Are Moral Intuitions and Why Should We Care about Them? A Neurobiological Perspective', *Philosophical Issues*, 18: 164–85.

Arango-Muñoz, S. (2014) 'The Nature of Epistemic Feelings', *Philosophical Psychology*, 27/2: 193–211. https://doi.org/10.1080/09515089.2012.732002.

Aristotle ([c. 330BC]2000) *Nicomachean Ethics*, ed. R. Crisp. Cambridge: Cambridge University Press.

Banaji, M. and L. Heiphetz (2010) 'Attitudes', in S. Fiske, D. Gilbert, and G. Lindzey (eds) *Handbook of Social Psychology*, 348–88. New York: Wiley.

Baron, A. S., Y. Dunham, M. Banaji, and S. Carey (2014) 'Constraints on the Acquisition of Social Category Concepts', *Journal of Cognition and Development*, 15/2: 238–48. https://doi.org/10.1080/15248372.2012.742902.

Blackburn, S. (1996) 'Securing the Nots: Moral Epistemology for the Quasi-Realist', in M. Timmons and W. Sinnott-Armstrong (eds) *Moral Knowledge? New Readings in Moral Epistemology*, 82–100. Oxford: Oxford University Press.

Brady, M. (2014) *Emotional Insight: The Epistemic Role of Emotional Experience*. Oxford: Oxford University Press.

Brentano, F. ([1874]1969) *Vom Ursprung Sittlicher Erkenntnis*. Hamburg: Felix Meiner.

Brogaard, B. and E. Chudnoff (2016) 'Against Emotional Dogmatism', *Philosophical Issues*, 26/1: 59–77. https://doi.org/10.1111/phis.12076.

Brownstein, M. (2018) *The Implicit Mind: Cognitive Architecture, the Self and Ethics*. Oxford: Oxford University Press.

Campbell, R. and V. Kumar (2012) 'Moral Reasoning on the Ground', *Ethics*, 122/2: 273–312. https://doi.org/10.1086/663980.

Carruthers, P. (2017) 'Are Epistemic Emotions Metacognitive?', *Philosophical Psychology*, 30/1–2: 58–78. https://doi.org/10.1080/09515089.2016.1262536.

Confucius ([c. 479BC]2003) *Analects, with Selections from Traditional Commentaries*, trans. E. Slingerland. Indianapolis, IN: Hackett.

Cuneo, T. (2006) 'Signs of Value: Reid on the Evidential Role of Feelings in Moral Judgement', *British Journal for the History of Philosophy*, 14/1: 69–91. https://doi.org/10.1080/09608780500449164.

Dancy, J. (2014) 'Intuition and Emotion', *Ethics*, 124/4: 787–812. https://doi.org/10.1086/675879.

D'Arms, J. and D. Jacobson (2023) *Rational Sentimentalism*. New York: Oxford University Press.

Deonna, J. A. (2006) 'Emotion, Perception and Perspective', *Dialectica*, 60/1: 29–46. https://doi.org/10.1111/j.1746-8361.2005.01031.x.

Deonna, J. A. and F. Teroni (2012) *The Emotions: A Philosophical Introduction*. London: Routledge.

Fjær, E. G. (2015) 'Moral Emotions the Day After Drinking', *Contemporary Drug Problems*, 42/4: 299–303. https://doi.org/10.1177/0091450915604988.

Furtak, R. (2018) *Knowing Emotions: Truthfulness and Recognition in Affective Experience*. Oxford: Oxford University Press.

Gendler, T. (2011) 'On the Epistemic Costs of Implicit Bias', *Philosophical Studies*, 156/1: 33–63.

Goldie, P. (2004) 'Emotion, Reason, and Virtue', in D. Evans and P. Cruse (eds) *Emotion, Evolution, and Rationality*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof.

Greene, J. (2013) *Moral Tribes: Emotion, Reason, and the Gap between Us and Them*. New York: Penguin.

Haidt, J. (2012) *The Righteous Mind*. London: Penguin.

Harrison, E. (2021) 'The Prospects of Emotional Dogmatism', *Philosophical Studies*, 178/8: 2535–45. https://doi.org/10.1007/s11098-020-01561-5

Herz, R. (2009) 'Aromatherapy Facts and Fictions: A Scientific Analysis of Olfactory Effects on Mood, Physiology and Behavior', *International Journal of Neuroscience*, 119/2: 263–70. https://doi.org/10.1080/00207450802333953.

Howard, S. (2017) 'Metaemotional Intentionality', *Pacific Philosophical Quarterly*, 98/3: 406–18. https://doi.org/10.1111/papq.12093.

Huemer, M. (2008) 'Revisionary Intuitionism', *Social Philosophy and Policy*, 25/1: 368–72. https://doi.org/10.1017/S026505250808014X.

Hutcheson, F. ([1725]2008) *An Inquiry into the Original of Our Ideas of Beauty and Virtue*. Indianapolis: Liberty Fund.

Hutton, J. (2022) 'Moral Experience: Perception or Emotion?', *Ethics*, 132/3: 570–7. https://doi.org/10.1086/718079.

——— (2023) 'What Attentional Moral Perception Cannot Do but Emotions Can', in R. Cowan (ed.) *Special Issue on Moral Perception, Philosophies*, 8/6: 106. Basel: MDPI. https://doi.org/10.3390/philosophies8060106

Inbar, Y., D. Pizarro, and P. Bloom (2012) 'Disgusting Smells Cause Decreased Liking of Gay Men', *Emotion*, 12/1: 23–7. https://doi.org/10.1037/a0023984.

Kahneman, D., O. Sibony, and C. Sunstein (2021) *Noise: A Flaw in Human Judgment*. London: William Collins.

Kauppinen, A. (2013) 'A Humean Theory of Moral Intuition', *Canadian Journal of Philosophy*, 43/3: 360–1.

——— (2021) 'Moral Sentimentalism', in E. N. Zalta (ed.) *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/moral-sentimentalism/ accessed 13 Jan 2025.

Kontaris, I., B. East, and D. Wilson (2020) 'Behavioral and Neurobiological Convergence of Odor, Mood and Emotion: A Review', *Frontiers in Behavioral Science*, 14. https://doi.org/10.3389/fnbeh.2020.00035.

Kurth, C. (2019) 'What Sentimentalists Should Say about Emotion', *Behavioral and Brain Sciences*, 42: e158. https://doi.org/10.1017/S0140525X18002601.

Lacewing, M. (2005) 'Emotional Self-Awareness and Ethical Deliberation', *Ratio*, 18/1: 65–71. https://doi.org/10.1111/j.1467-9329.2005.00271.x.

Landy, J. and G. Goodwin (2015) 'Does Incidental Disgust Amplify Moral Judgment? A Meta-Analytic Review of Experimental Evidence', *Perspectives on Psychological Science*, 10/4: 518–26.

Levy, N. (2007) 'Radically Socialized Knowledge and Conspiracy Theories', *Episteme*, 4/2: 181–2. https://doi.org/10.3366/epi.2007.4.2.181.

Lewis, M. (2021) *The Premonition: A Pandemic Story*. London: Penguin.

Machery, E. (2022) 'Anomalies in Implicit Attitudes Research', *WIREs Cognitive Science*, 13/1: e1569. https://doi.org/10.1002/wcs.1569.

Mathews, K. E. and L. K. Canon (1975) 'Environmental Noise Level as a Determinant of Helping Behavior', *Journal of Personality and Social Psychology*, 32/4: 571–7. https://doi.org/10.1037/0022-3514.32.4.571.

May, J. (2014) 'Does Disgust Influence Moral Judgment?', *Australasian Journal of Philosophy*, 92/1: 125–31. https://doi.org/10.1080/00048402.2013.797476.

——— (2018) *Regard for Reason in the Moral Mind*. Oxford: Oxford University Press.

McBrayer, J. (2010) 'Moral Perception and the Causal Objection', *Ratio*, 23/3: 291–7. https://doi.org/10.1111/j.1467-9329.2010.00468.x.

McGregor, R. (2015) 'Making Sense of Moral Perception', *Ethical Theory and Moral Practice*, 18/4: 745–58. https://doi.org/10.1007/s10677-015-9601-9.

McKinney, A. and K. Coyle (2006) 'Alcohol Hangover Effects on Measures of Affect the Morning after a Normal Night's Drinking', *Alcohol and Alcoholism*, 41/1: 54–60. https://doi.org/10.1093/alcalc/agh226.

Mengzi ([c. 300BC]2008) *Mengzi: With Selections from Traditional Commentaries*, trans. B. W. Van Norden. Indianapolis: Hackett.

Milona, M. (2016) 'Taking the Perceptual Analogy Seriously', *Ethical Theory and Moral Practice*, 19/4: 897–905. https://doi.org/10.1007/s10677-016-9716-7.

——— (2017) 'Intellect versus Affect: Finding Leverage in an Old Debate', *Philosophical Studies*, 174/9: 2251–6. https://doi.org/10.1007/s11098-016-0797-x.

——— (2023) 'Armchair Evaluative Knowledge and Sentimental Perceptualism', in R. Cowan (ed.) *Special Issue on Moral Perception, Philosophies*, 8/3: 51. Basel: MDPI. https://doi.org/10.3390/philosophies8030051

Milton, I. J., E. Sillence, and M. Mitchell (2019) 'Exploring the Emotional Experiences of Alcohol Hangover Syndrome in Healthy UK-Based Adults', *Drugs: Education, Prevention and Policy*, 27/3: 248–60. https://doi.org/10.1080/09687637.2019.1654431.

Mitchell, J. (2017) 'The Epistemology of Emotional Experience', *Dialectica*, 71/1: 57–64.

——— (2019a) 'Pre-Emotional Awareness and the Content-Priority View', *The Philosophical Quarterly*, 69/277: 771–4. https://doi.org/10.1093/pq/pqz018.

——— (2019b) 'The Intentionality and Intelligibility of Moods', *European Journal of Philosophy*, 27/1: 118–25. https://doi.org/10.1111/ejop.12385.

——— (2019c) 'Understanding Meta-Emotions: Prospects for a Perceptualist Account', *Canadian Journal of Philosophy*, 1–9. https://doi.org/10.1017/can.2019.47.

——— (2021) *Emotion as Feeling Towards Value: A Theory of Emotional Experience*. Oxford: Oxford University Press.

Müller, J. M. (2019) *The World-Directedness of Emotional Feeling*. Cham: Palgrave Macmillan.
Pelser, A. (2014) 'Emotion, Evaluative Perception, and Epistemic Justification', in S. Roeser and C. Todd (eds) *Emotion and Value*, 107–23. Oxford: Oxford University Press.
Penning, R., A. McKinney, and J. Verster (2012) 'Alcohol Hangover Symptoms and Their Contribution to the Overall Hangover Severity', *Alcohol and Alcoholism*, 47/3: 248–52. https://doi.org/10.1093/alcalc/ags029.
Railton, P. (2014) 'The Affective Dog and Its Rational Tale: Intuition and Attunement', *Ethics*, 124/4: 813–9. https://doi.org/10.1086/675876.
Roberts, R. (1988) 'What an Emotion Is: A Sketch', *The Philosophical Review*, 97/2: 183–9.
——— (2013) *Emotions in the Moral Life*. Cambridge: Cambridge University Press.
Roeser, S. (2011) *Moral Emotions and Intuitions*. Basingstoke: Palgrave Macmillan.
Sarachild, K. (1969) 'A Program for Feminist "Consciousness Raising"', in S. Firestone (ed.) *Notes from the Second Year: Women's Liberation. Major Writings of the Radical Feminists*, 78–80. New York, NY: Radical Feminism.
Schnall, S., J. Haidt, G. Clore, and A. Jordan (2008) 'Disgust as Embodied Moral Judgment', *Personality & Social Psychology Bulletin*, 34/8: 1096–110.
Seidel, A. and J. Prinz (2013) 'Sound Morality: Irritating and Icky Noises Amplify Judgments in Divergent Moral Domains', *Cognition*, 127/1: 1–5. https://doi.org/10.1016/j.cognition.2012.11.004.
Shaftesbury, A. A. C., Earl of ([1711]2001) *Characteristicks of Men, Manners, Opinions, Times*. Indianapolis: Liberty Fund.
Singer, P. (2005) 'Ethics and Intuitions', *The Journal of Ethics*, 9/3–4: 331–2.
Sinnott-Armstrong, W. (1991) 'Moral Experience and Justification', *Southern Journal of Philosophy*, 29/S1: 89–96. https://doi.org/10.1111/j.2041-6962.1991.tb00614.x.
——— (2011) 'Emotion and Reliability in Moral Psychology', *Emotion Review*, 3/3: 288–9. https://doi.org/10.1177/1754073911402382.
Smith, A., H. Whitney, M. Thomas, K. Perry, and P. Brockman (1997) 'Effects of Caffeine and Noise on Mood, Performance and Cardiovascular Functioning', *Human Psychopharmacology: Clinical and Experimental*, 12/1: 27–33.
Sosa, E. (1991) *Knowledge in Perspective*. Cambridge: Cambridge University Press.
——— (2017) *Epistemology*. Princeton: Princeton University Press.
Szigeti, A. (2013) 'No Need to Get Emotional? Emotions and Heuristics', *Ethical Theory and Moral Practice*, 16/4: 845–2. https://doi.org/10.1007/s10677-012-9386-z.
Tappolet, C. (2016) *Emotions, Values, and Agency*. Oxford: Oxford University Press.
Thagard, P. (2000) *Coherence in Thought and Action*. Cambridge, MA: MIT Press.
Tolhurst, W. (1990) 'On the Epistemic Value of Moral Experience', *Southern Journal of Philosophy*, 29/Supplement: 67–71.
Wang, Y. ([1572]2009) 'A Record for Practice', in P. Ivanhoe (ed.) *Readings from the Lu-Wang School of Neo-Confucianism*, trans. P. Ivanhoe, 131–9. Indianapolis: Hackett.
Wiggins, D. (1991) 'A Sensible Subjectivism', in *Needs, Values, Truth (2nd Edition)*, 185–214. Oxford: Blackwell.
Williams, B. (1986) *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.
Woodward, J. (2016) 'Emotion versus Cognition in Moral Decision-Making: A Dubious Dichotomy', in S. M. Liao (ed.) *Moral Brains: The Neuroscience of Morality*, 87–116. Oxford: Oxford University Press.