

Evolutionary Debunking, Self-Defeat and All the Evidence¹

Silvan Wittwer (Harvard University)

[Note: this is the penultimate draft of a chapter forthcoming in *Higher-Order Evidence and Moral Epistemology*, Michael Klenk (ed.), New York: Routledge. Please cite the published version.]

1. Introduction

Robust moral realism is the meta-ethical view that there are mind-independent moral facts which are irreducible to natural facts and *in principle* knowable.² Recently, this view has been criticized heavily on evolutionary grounds. So-called *evolutionary debunkers* argue that becoming aware of the evolutionary origins of our moral beliefs, realistically construed, undermines their justification.³ While there are various ways of spelling out the underlying epistemological details, an appeal to *higher-order evidence of error* has recently garnered increased philosophical attention.⁴ More precisely, Tomas Bogardus (2016), Andreas Mogensen (ms, 2017) and – at least on one plausible reconstruction⁵ – Sharon Street (2005) argue that evolutionary considerations provide such evidence, add the view – often called *Conciliationism*⁶ – that higher-order evidence of error defeats justification, and thus conclude that evolutionary considerations defeat the justification of our moral beliefs.

In response, moral realists such as Katia Vavova (2015) have objected that this evolutionary debunking argument is *self-defeating*.⁷ To see how that threat materializes, note that Conciliationism is typically motivated by *Independence*, the principle that we should assess higher-order evidence of error with respect to *p* *independently* of our original first-order evidence, beliefs or reasoning in support of *p*.⁸ To respect this principle when assessing *evolutionary* higher-order evidence of error, the moral realist would need to set aside *all* her moral evidence, beliefs and reasoning. But doing so risks setting aside *too much* to know whether she is genuinely mistaken about morality or not. So, the evolutionary debunking argument above seems to defeat itself.

The literature lacks any discussion of whether evolutionary debunkers can handle this self-defeat objection.⁹ My overall aim in this paper is to argue that they cannot, thus filling that lacuna – and vindicating Vavova’s worry. To achieve my aim, I proceed in two steps. First, I propose a novel, *prima facie* promising strategy for avoiding self-defeat. Then, I show that evolutionary debunkers face insuperable difficulties trying to successfully implement that strategy. As a result, the evolutionary debunking argument from higher-order evidence of error fails.

What does this strategy (that constitutes the first step of my argument) look like? It consists in realizing that Conciliationism isn’t the only *prima facie* plausible view on the epistemic significance of higher-order evidence of error. In light of that, it would be premature to conclude that the evolutionary debunking arguments from higher-order evidence defeats itself. Rather, there are independently compelling alternatives available – and it remains an issue of active debate which we should endorse. Further, some of these alternatives appear to have distinctive features that would help avoid self-defeat. In particular, some views (A) reject *Independence* and therefore allow first-order moral evidence into the picture, yet (B) still promise to yield the verdict that evolutionary higher-order evidence can defeat the justification of first-order moral beliefs. For example, Thomas Kelly’s (2010) *Total Evidence View* fits that bill. According to it, your total evidence determines whether your belief that p is justified or not. The total evidence includes both your first-order evidence (in support of p) and the second-order evidence of peer disagreement (regarding p). Therefore, Kelly’s (2010) view clearly satisfies (A): it rejects *Independence*, affording our first-order moral evidence, beliefs and reasoning a role in determining whether a given belief is justified or not. That amounts to a promising start, giving evolutionary debunkers enough reason to explore whether their argument avoids self-defeat if they accept the Total Evidence View instead of Conciliationism.

However, whether Kelly's (2010) view satisfies (B) is more complicated, as I show in the second step of my argument. To establish (B), evolutionary debunkers must argue that the total evidence available to the robust moral realist, consisting of her first-order moral evidence and the evolutionary higher-order evidence of error, defeats the justification of her moral beliefs. But, on reflection, evolutionary debunkers cannot discharge this argumentative burden. The exact reason for their failure depends on the kind of higher-order evidence of error that evolutionary considerations allegedly provide. Debunkers such as Street (2006), for whom evolutionary considerations supply evidence of moral unreliability, struggle with *evidential weight*. In contrast, debunkers such as Bogardus (2016) and Mogensen (ms, 2017), who construe evolutionary considerations as evidence of moral peer disagreement, are committed to a pair of inconsistent assumptions about *evolutionary counterparts*. Either way, evolutionary debunkers who appeal to the notion of higher-order evidence of error struggle to implement (B) of my proposed strategy. By implication, their arguments cannot avoid self-defeat.

Here is my plan. §2 presents the evolutionary debunking argument from higher-order evidence of error in more detail, while §3 turns towards characterizing the self-defeat objection. Then, §4 unveils my two-part strategy for dealing with self-defeat. The remainder of the paper explores whether evolutionary debunkers can make good on that strategy. §5 argues that the first part is easy to pull off, but both §6 raises concerns about the implementation of the second part, having to do with evidential weight and whether evolutionary counterparts qualify as epistemic peers. §7 concludes.

2. Evolutionary Debunking and Higher-Order Evidence

Recently, some philosophers have argued that evolutionary theory debunks beliefs in robust mind-independent moral facts. To make their case, some of these *evolutionary debunkers* rely on the notion of higher-order evidence of error. Despite differences in detail, the arguments put forth by Tomas

Bogardus (2016), Andreas Mogensen (ms, 2017) and – at least on Vavova’s (2015) compelling reconstruction – Sharon Street (2005) share the following structure:

- (1) Higher-order evidence of error about your beliefs that p, q, etc. defeats their justification.¹⁰
- (2) Evolutionary considerations provide moral realists with higher-order evidence of error about their moral beliefs.
- (3) Therefore, evolutionary considerations defeat the justification of the moral realists’ moral beliefs.

This argument, if sound, forces moral realists into moral scepticism: even if robust moral facts exist, our beliefs about them aren’t justified. To begin with, the first premise introduces the notion of *higher-order evidence of error*. Such evidence of error indicates that one suffers from some epistemic malfunction.¹¹ Here are two familiar examples:

Offside Call: In my spare time, I enjoy attending football games with my best friend Julian. We are both equally good at spotting whether a forward is offside or not. Last Sunday, though, we disagreed: while Julian judged that our forward started from an offside position and the resulting goal was thus irregular, it seemed to me that our forward timed his run well and was onside.¹²

Hypoxia: While climbing the Dufour peak in the Swiss Alps, the weather suddenly turns near the summit. I stop briefly to calculate whether there is enough time to reach the peak and start the climb down before the snow storm hits. After going over my calculations several times, I am rather confident that I should be able to make it. However, I suddenly remember that, given the high altitude, I am very likely to suffer from mild hypoxia (or

lack of oxygen), which undetectably impairs one's reasoning, leading to stupid, yet fatal mistakes.¹³

In the first case, I receive evidence of *peer disagreement*: Julian and I are equally good at making offside calls, but disagree about whether our forward was offside or not this time. Since we cannot both be right (but are equally good at making offside calls), one of us must be in error, which might very well be me. In contrast, the second case features evidence of *unreliability*: hypoxia makes it very likely that my reasoning (about time, in this case) is mistaken.¹⁴

The first premise doesn't just introduce higher-order evidence of error, but it also articulates a view about its epistemic significance. *Conciliationism* says that higher-order evidence of error defeats the *justification* of relevant first-order beliefs.¹⁵ Originally, this view was defended in the context of peer disagreement, recommending that one *conciliate* (hence its name) upon receiving evidence of peer disagreement such as in **Offside Call**.¹⁶ However, it can easily be generalized, resulting in a view that says that *any kind* of higher-order evidence of error defeats justification.¹⁷ For instance, it would say that, in **Hypoxia**, learning of my likely reason-distortion defeats the justification of my belief that I have enough time to reach the Dufour Peak and return safely. So, Conciliationism is a moderate form of *skepticism*: according to it, higher-order evidence of error defeats justification, yet only the justification of those first-order beliefs that we have such evidence about.¹⁸

Conciliationism is *prima facie* attractive.¹⁹ First, it accommodates our *intuitions*: in both **Offside Call** and **Hypoxia**, it seems *intuitively* appropriate to revise our beliefs in light of the higher-order evidence of error – which Conciliationism respects. Second, the view also plausibly *explains* our intuitions. For instance, given that Julian and I are equally like to get offside calls right and that there must be a mistake either on his or my part in **Offside Call**, we both have reason to think that we have made a mistake, which may well be enough to defeat the justification of our relevant

beliefs. Similarly, given evidence of my medical condition in **Hypoxia**, it is significantly more likely that I have made a justification-defeating mistake in my time management. Third, Conciliationism follows from a seemingly plausible principle for correctly evaluating evidence. According to *Independence*, we should assess higher-order evidence of error with respect to *p* *independently* of our original first-order evidence, beliefs or reasoning in support of *p*.²⁰ To see why that seems plausible, reconsider **Offside Call**: once I learn that Julian, an epistemic peer, disagrees with me, it would be intuitively *wrong* or *irrational* to dismiss his judgment by depending on my initial perceptual seeming that the forward wasn't offside. (Similarly, for **Hypoxia**: sticking with my original reasoning would be epistemically problematic in the face of (significant risk of) altitude-induced distortion of reasoning.) But once we accept *Independence*, Conciliationism straightforwardly follows: if it is rational to bracket one's first-order evidence, etc. and thus only the higher-order evidence of error matters, it will defeat the justification of our relevant first-order beliefs.²¹

The second premise states that evolutionary considerations provide moral realists with higher-order evidence of error about their moral beliefs.²² But what *kind* of evidence exactly? There are at least two answers in the literature, both of which work for the argument.

On the one hand, according to Sharon Street (2005), moral realists face a scenario similar to **Hypoxia**. For her, evolutionary considerations provide moral realists with evidence of moral *unreliability*.²³ After all, when moral realists reflect on the evolutionary origin of our moral beliefs, they must realize that evolution selects for *adaptive*, not true, moral beliefs.²⁴ For instance, suppose you believe that you have special moral obligations to your family, based on a corresponding moral intuition. But then you realize that we evolved to survive, not to track mind-independent moral truths, and that it is therefore likely that your belief is false. So, moral realists have good reason to think that their moral beliefs have been *unreliably* formed: they are the upshot of a process that was

not designed to get at moral truth. In that way, evolutionary considerations provide moral realists with evidence of moral unreliability.

On the other hand, for Bogardus (2016) and Mogensen (ms, 2017), moral realists find themselves in a situation similar to **Offside Call**.²⁵ According to them, evolutionary considerations amount to evidence of *possible peer disagreement*: evidence that moral realists *could* disagree with their evolutionary counterparts about fundamental moral matters (such as the wrongness of incest or slavery, our obligations towards our children, that ethnicity doesn't matter to moral standing, etc.). After all, moral realists must realize that, had humans evolved differently, they would hold different moral beliefs now. Suppose moral realists believe that incest is morally wrong, based on a corresponding moral intuition. Their evolutionary counterparts might disagree: since incest did *not* hamper *their* reproductive fitness, they don't believe that it is morally impermissible. Rather, they believe that it is perfectly morally alright, based on *their* corresponding moral intuition. In that manner, evolutionary considerations amount to evidence of possible peer disagreement.

With both premises in place, we are now able to secure the skeptical conclusion. After all, combining the claim that evolutionary considerations provide moral realists with higher-order evidence of error about their moral beliefs with the view that such evidence defeats justification yields the conclusion that evolutionary considerations defeat the justification of our moral beliefs, realistically construed. Therefore, evolutionary considerations seem to saddle moral realists with an uncomfortable commitment to moral skepticism. Even if robust moral facts exist, we do not form justified beliefs about them, given the higher-order evidence of error that evolutionary considerations provide.

3. The Threat of Self-Defeat

One of the most powerful objections to the evolutionary debunking argument above is that it threatens to be *self-defeating*. To see how that threat materializes, recall that Conciliationism is

typically motivated by *Independence*, the principle that we should assess higher-order evidence of error with respect to p *independently* of our original first-order evidence, beliefs or reasoning in support of p . To respect this principle when assessing *evolutionary* higher-order evidence of error, the moral realist would need to set aside *all* her moral evidence, beliefs and reasoning. Why? Because the debunker's idea is that evolutionary considerations call into question *all* her moral beliefs. But doing so risks setting aside too much to know whether she is mistaken about morality or not. As Katia Vavova (2015: 89-93), who develops this objection most clearly and forcefully, writes: 'we cannot determine if we are likely to be mistaken about morality if we can make no assumptions at all about what morality is like' (Vavova 2015: 92).²⁶ After all, to see whether true moral and adaptive moral beliefs do indeed come apart, as the evolutionary debunker has it, we need to know something about the contents of both of those sets. If we don't know what morality is, how can we know whether we fall short of it? Or, if morality could be about anything, we have no reason to think that mind-independent moral truths and adaptive moral beliefs don't coincide or overlap. So, the evolutionary debunking argument above seems to defeat itself.

For illustration, consider an analogy with perception: to evaluate whether my perceptual beliefs about mid-sized objects in my immediate environment (e.g. tables, chairs, desk lamps, water bottles, coffee mugs) are indeed unreliably formed, I need to make some assumptions about the contents of my perception. For instance, I need to know very roughly what a chair is to make sure that my perceptual belief that there is a chair right in front of me is false. Similarly, Vavova points out: 'I cannot show that I am not hopeless at understanding right and wrong without being allowed to make some assumptions about what is right and wrong' (ibid.). For instance, consider my moral belief that racism is wrong. At the same time, I am aware of evolutionary explanations of racism: it is adaptive to be suspicious of those who look different to me. Here, the adaptive and true moral beliefs come apart. However, importantly, to draw that distinction, I must already assume some moral truths, including that racism is morally wrong.²⁷

It is important to appreciate how *general* this objection is. More precisely, the worry does *not* depend on a *specific* interpretation of premise 2, the claim that evolutionary considerations provide higher-order evidence of error. Vavova (2015) criticizes a version of the evolutionary debunking argument according to which thinking about the evolutionary origins of their moral beliefs provides moral realists with evidence of *unreliability*. But that, in my view, artificially restricts the scope of the objection, since we can easily *generalize* it to *any* evolutionary debunking argument from higher-order evidence of error, regardless of the *kind* of evolutionary higher-order evidence. After all, any such argument subscribes to the problematic commitments that the objection capitalizes on: evolution as providing higher-order evidence of error, Conciliationism and, especially, *Independence*. For instance, suppose that evolutionary considerations amount to evidence of peer disagreement. Once the moral realist receives such evidence, she must once again set aside all her moral evidence, beliefs and reasoning to respect *Independence*. But doing so would make it impossible for her to assess whether she, as opposed to her evolutionary counterpart, is more likely mistaken about morality. Again, the evolutionary debunker ends up with *self-defeat*. So, the threat is perfectly general – and therefore relevant to any evolutionary debunker relying on the notion of higher-order evidence of error.²⁸

Of course, not all evolutionary debunking arguments appeal to (higher-order) evidence of error. Rather, some influential recent debunkers have argued that evolutionary considerations defeat the justification of our moral beliefs because they show that those beliefs are insensitive or unsafe.²⁹ For evolutionary debunking arguments developed along such lines, Vavova's self-defeat worry might not arise. However, such arguments face their own issues. For example, Justin Clarke-Doane (2015, 2016, 2017, forthcoming) has convincingly responded that our moral beliefs are *modally robust*, assuming their truth and defeasible justification.³⁰ After all, our beliefs are sensitive to the robust moral facts by default: since such facts hold necessarily, they could not possibly change. So,

any true moral belief must be sensitive as well. Similarly, our moral beliefs are safe from error: they are true (as we assume) and could not easily have evolved differently. So, our moral beliefs could not easily be false.³¹ But if the prospects of such modal evolutionary debunking arguments look dim, defending the evolutionary debunking argument from higher-order evidence against the threat of self-defeat becomes all the more *important*.³²

4. Avoiding Self-Defeat: A Strategy

As mentioned in the introduction, the literature doesn't feature any discussion of how evolutionary debunkers could or should respond to this self-defeat objection. My aim in what follows is to address that shortcoming. I shall propose and subsequently evaluate a two-pronged strategy. Its key insight is this: Conciliationism isn't the only *prima facie* plausible view on the epistemic significance of higher-order evidence of error. Rather, independently compelling alternatives are subject to active debate in the literature. And, importantly, some of these alternatives share distinctive features that promise to help avoid self-defeat. So, evolutionary debunkers should reject Conciliationism (or premise (1) of their argument). Instead, they should explore the prospects of a view that (A) rejects *Independence* and therefore allows first-order moral evidence into the picture, yet (B) still yields the verdict that evolutionary higher-order evidence defeats the justification of first-order moral beliefs.³³ In this section, I shall briefly motivate both prongs, while the next section introduces Thomas Kelly's (2010) *Total Evidence View* as a candidate framework that clearly satisfies (A). The final section will then focus on whether evolutionary debunkers could make good on (B) within Kelly's framework, discussing evolutionary evidence of moral unreliability (§6.1) before delving into evolutionary evidence of moral peer disagreement (§6.2).³⁴

To wit, three claims constitute the evolutionary debunking argument: Conciliationism about higher-order evidence of error (including *Independence*), the claim that evolutionary considerations provide the moral realist with such evidence, and the skeptical conclusion. To avoid self-defeat,

evolutionary debunkers must give up one of them. They cannot give up the last two. Without the distinctive claim about the evidential import of evolution (or, more generally, etiology), their argument would cease to be an *evolutionary* (or, more generally, etiological) one. And without the skeptical conclusion, their argument wouldn't be a *debunking* one. But Conciliationism and especially *Independence* appear to be the culprits: only if the moral realist is forced to assess the evolutionary evidence *independently* of all her first-order moral evidence does the argument threaten to defeat itself. Therefore, evolutionary debunkers should reject Conciliationism due to its commitment to *Independence*, while holding on to the claim that evolution provides higher-order evidence of error as well as the skeptical conclusion.

But giving up on Conciliationism (or premise (1)) won't be enough, of course. Rather, evolutionary debunkers also need a replacement, a view that licenses the inference from the claim that evolutionary considerations amount to higher-order evidence of error to the skeptical conclusion. More precisely, evolutionary debunkers must defend (or, at least, sketch) a view with two distinctive features: it must (A) reject *Independence* and thus allow first-order (moral) evidence into the picture, yet (B) still yield the verdict that (evolutionary) higher-order evidence defeats the justification of first-order (moral) beliefs. To successfully deal with the self-defeat objection, the view appealed to in premise (1) must satisfy both requirements. Are such views available?

5. The Total Evidence View

Fortunately for the evolutionary debunker, there are independently plausible views on *peer disagreement* that promise to fit the bill. For instance, take the *Total Evidence View*, as developed by Thomas Kelly (2010: 33-44).³⁵ This view states that your *total evidence* determines whether your belief that *p* is justified or not. The total evidence includes both your first-order evidence (in support of *p*) and the second-order evidence of peer disagreement (regarding *p*).

Sometimes, the total evidence justifies my original belief and thus makes it reasonable for me to stick to my guns. Suppose that I disagree once again with Julian about a football matter. While I insist that there was no delay of game, Julian claims that there was – and that, rather absurdly, the goalkeeper, who isn't on a yellow card, should be immediately sent off for it.³⁶ Here, the total evidence, consisting of my perceptual experience and knowledge of the football rule book, justifies my belief that there was no delay of game.

But on other occasions, the total evidence may defeat the justification of my original belief – and therefore make it reasonable for me to change my outlook. Suppose that I am a skeptic about other minds, based on the relevant class of arguments. But then I find out that an overwhelming majority of professional philosophers disagree with me, having independently arrived at their view.³⁷ Here, the total evidence, consisting of the arguments, my considered judgment as well as all the considered judgments of my peers, seems to justify the common-sense view that there are indeed other minds. So, it would be reasonable for me to conciliate – or even accept the common-sense view.³⁸

The Total Evidence View clearly satisfies the first requirement or (A) above. Unlike Conciliationism, this view rejects *Independence*: it affords our first-order evidence, beliefs and reasoning a role in determining whether a given belief is justified or not. In our context, that means that the robust moral realist no longer needs to set aside all her first-order moral evidence, beliefs and reasoning when assessing evolutionary higher-order evidence of error. And since she is not forced to do that, she does not risk setting aside too much to know whether she is mistaken or not. Therefore, the Total Evidence View looks like a suitable candidate for premise (1) of the debunkers' argument.

But what about the second requirement or (B): does the view yield the verdict that evolutionary higher-order evidence of error defeats the justification of first-order moral beliefs? Here, matters get more complicated. To establish that, evolutionary debunkers would need to argue that the total evidence available to the robust moral realist, consisting of her first-order moral evidence and the evolutionary higher-order evidence of error, does indeed defeat the justification of her moral beliefs. Can they do so?

In what follows, I shall answer that question negatively. On reflection, the total evidence available to the robust moral realist does not defeat the justification of her moral beliefs. By implication, evolutionary debunkers cannot satisfy the second requirement or (B) of my strategy above. And since they cannot do that, their argument defeats itself. However, as I shall argue, the exact reason for why evolutionary debunkers fail to establish defeat depends on the kind of higher-order evidence that evolutionary considerations allegedly provide. As I shall discuss in §6.1, debunkers such as Street (2006), for whom evolutionary considerations supply evidence of moral unreliability, struggle with *evidential weight*. In contrast, I make the case in §6.2 that the evolutionary disagreement argument developed by Bogardus (2016) and Mogensen (ms, 2017) rests on a pair of inconsistent assumptions.

6. What does the Total Evidence Defeat?

6.1. Evolution and Moral Unreliability

Suppose Street (2005) is right: evolutionary considerations provide evidence of moral unreliability. Then, the total evidence available to the moral realist consists of both that evidence and our first-order moral evidence. Does *that* evidence defeat the justification of our moral beliefs? The answer seems to depend on the *weight* of the respective bodies of evidence. If the evolutionary evidence of moral unreliability *outweighed* our first-order moral evidence, our moral beliefs would no longer

be justified. If it didn't, our moral beliefs would remain undefeated. But how should we decide which of those two conditionals holds?

To start with, we might appeal to brute intuition. After all, we seem to have widely shared intuitions about whether the first- or higher-order evidence has greater weight in some extreme cases. For instance, it is intuitive to think that the higher-order evidence of error outweighs our first-order evidence in Hypoxia. Conversely, it is natural to think that our first-order perceptual evidence outweighs the higher-order evidence of error gained by finding out that a very popular headache remedy caused hallucinations in 1 in 1 million test subjects. And there is some support for this line of thinking in the literature on peer disagreement. For instance, Kelly (2010) seems to advocate a form of epistemic particularism about the matter. When considering the question of whether the first- or higher-order evidence plays a greater role in fixing the reasonability of what to believe, he writes that '...the question of which counts for more – peer opinion, or the evidence on which the peers base their opinion? – is not, I think, a good question when it is posed at such a high level of abstraction' (34). Rather, we have to examine cases and our intuitive verdicts about them individually. Similarly, Errol Lord (2014) sketches a test for evidential weight in the context of peer disagreement that seems driven by brute intuition. To determine whether one's original reasons are strong or weighty enough to ground a permission to dismiss peer disagreement, we should ask: 'do [those original reasons] put you in a position to think your peer is crazy or otherwise epistemically suspect?' (Lord 2014: 376, fn. 15; emphasis mine). And perhaps, his proposal can be generalized to evidence of unreliability: if your first-order evidence makes the source of the evidence of unreliability seem epistemically suspect, the former outweighs the latter.

But I don't think that an appeal to brute intuition will help the debunker. Unlike our widely shared intuitions about evidential weight in extreme cases, our intuitions about the weight of evolutionary evidence of moral unreliability vis-à-vis our first-order moral evidence strike me much more moot.

To see that, it suffices to point to the persistent disagreement in the literature about the epistemic import of evolutionary biology for robustly moral belief – which seems at least partially fueled by conflicting intuitions about evidential weight. While evolutionary debunkers share the intuition that evolutionary evidence is weightier than moral evidence and therefore undermines robustly moral beliefs, robustly moral realists tend to lack the intuition – or explain it away as irrelevant. If that is correct, we have reached a dialectical stalemate, without making any progress in the matter at hand. So, an appeal to brute intuition won't help evolutionary debunkers such as Street (2006) to establish the claim that evolutionary evidence of moral unreliability outweighs our first-order moral evidence.

Instead, the evolutionary debunker might look towards a more theoretical (or formal) notion of evidential weight. Unfortunately, there is remarkably little literature on how to measure and compare the weight of evidence – and no literature at all on how to apply such ideas to evolutionary debunking. Still, there are some suggestions worth exploring. For instance, take James Joyce (2005: 162-5): for him, the balance of total evidence favors whatever ordered sequence of propositions contains an estimated higher number of truths.³⁹ Following his proposal, we estimate whether the first-order moral evidence or the evolutionary evidence of moral unreliability contains a higher number of truths. If we can reasonably expect the first-order moral evidence to feature more truths, the balance of total evidence favors the moral propositions making up that evidence. That would be bad news for evolutionary debunkers. Conversely, if we can reasonably expect the higher-order evidence of error to feature more truths, the balance of total evidence favors the propositions making up the higher-order evidence of error. That would be good news for debunkers such as Street.

However, I doubt that evolutionary debunkers like Street (2006) can avail themselves readily of the resources that Joyce's (2005) framework offers. First, we might worry generally that this

probabilistic notion of evidential weight cannot successfully model interactions between bodies of evidence at different levels. After all, when weighing evidence from different levels, the estimates won't be independent. Rather, the estimate of how many truths the first-order evidence contains will depend to some extent on our estimate of how many truths the higher-order evidence (of error) contains. That might complicate the formation of reasonable expectations. Further, that dependency might be especially pertinent in the context of evolutionary debunking. As Kevin Brosnan (2011: 55) points out, it is impossible to estimate of how likely it is that moral truths obtain prior to evolutionary influence. After all, (almost) everybody accepts that our moral beliefs evolved – and that we thus cannot assess their truth prior to or independently of evolution. But if that is correct, how can we estimate how many truths the first-order moral evidence contains? Second, the specific application of Joyce's account to evolutionary debunking might prove problematic also in another way. In particular, we might find it hard to estimate the number of basic moral truths. After all, there is plenty of disagreement about what they are and which of them are properly basic, even amongst robust moral realists. So, how are we supposed to count the number of basic robustly moral truths? Given these significant issues, evolutionary debunkers most likely cannot borrow Joyce's way of measuring the balance of total evidence. But if they cannot do that, they once again won't be able to make good on the claim that the evolutionary evidence of moral unreliability outweighs our first-order moral evidence.

In short, debunkers who construe evolutionary considerations as evidence of moral unreliability struggle at the first hurdle. To meet the second requirement or (B) of the strategy outlined in §4 and therefore avoid self-defeat, they must establish that the evolutionary evidence of moral unreliability outweighs our first-order moral evidence. But doing that, in turn, requires a plausible theoretical notion of evidential weight. Even though it might not be impossible to find such a notion, our discussion (and the dearth of literature on the subject) suggests that these evolutionary

debunkers have their work cut out for themselves.⁴⁰ It seems fair to conclude that any evolutionary debunking argument based on evidence of moral unreliability must defeat itself.

6.2. Evolution and Moral Disagreement

Suppose Bogardus (2016) and Mogensen (ms, 2017) are right: evolutionary considerations amount to evidence of possible moral peer disagreement. Does the total evidence available to the moral realist in *that* case defeat the justification of our moral beliefs?

To answer that question affirmatively, evolutionary debunkers may proceed in two steps. First, they remind us of Kelly's (2010: 42-44) diagnosis of **Offside Call**. There, Kelly argues, the total evidence available to me defeats the justification of my belief that the scorer was onside.⁴¹ Why that? Initially, Julian and I have different first-order perceptual evidence: while it appears to me that the forward was onside when he started his run, the opposite seems to be the case to Julian. This evidence justifies our initial perceptual beliefs, respectively. But once we become aware of our perceptual disagreement, we *pool* our first-order evidence and add the higher-order evidence of peer disagreement. The resulting *total* evidence neither supports my belief that the goal was scored onside – nor Julian's belief to the contrary. Instead, we have a situation of *evidential symmetry* between myself and Julian. Therefore, neither his nor my initial belief is justified anymore. That is how the Total Evidence View diagnoses cases with the structure of **Offside Call**.

In a second step, evolutionary debunkers can argue that possible moral peer disagreements with your evolutionary counterparts share the structure of **Offside Call**. Initially, we also have two different bodies of first-order evidence: while you have the moral intuition that you have special obligations to your family members, your evolutionary counterpart has the contrary moral intuition. Again, this evidence justifies our respective initial moral beliefs. But once you discover the moral peer disagreement and pool your first-order moral evidence, the *total evidence* – just as

above – supports neither your belief that you have special moral obligations to your family members nor your evolutionary counterpart’s belief to the contrary. Therefore, the justification of both your and your counterpart’s initial moral beliefs have been defeated. And, of course, the same reasoning could be employed for any other basic moral intuition that seems *prima facie* compelling and could be subject to disagreement based on divergent evolutionary histories.⁴²

By following these two steps, evolutionary debunkers who construe evolutionary considerations as evidence of disagreement could argue that the total evidence available to the robust moral realist defeats the justification of her moral beliefs. Achieving that – without even requiring a theoretical notion of evidential weight – would be no mean feat. Rather, it would show that they make good on the second requirement or (B) of my strategy outlined in §4: that the Total Evidence View yields the verdict that evolutionary higher-order evidence defeats the justification of first-order moral beliefs. As a result, these evolutionary debunkers would avoid self-defeat.

To make this move work, however, evolutionary considerations would have to amount to evidence of possible moral peer disagreement.⁴³ (We cannot merely suppose that they do, as we have done up to this point.) In more detail, our evolutionary counterparts must not just disagree with us about fundamental moral matters, but also count as our epistemic and – in a sense to be explained presently – metaphysical peers. However, I doubt that entities can simultaneously satisfy both of those criteria. But if they cannot, evolutionary considerations cannot plausibly amount to evidence of moral peer disagreement. That, in turn, means that the evolutionary debunking argument from moral peer disagreement doesn’t get off the ground.⁴⁴

To see how this worry arises, note that any evolutionary debunking argument based on moral peer disagreement rests on two crucial assumptions. First, it assumes that our counterparts have an alternative evolutionary history – and that this explains why they hold radically different moral

beliefs or disagree with us about fundamental moral matters. To illustrate, consider an example introduced in §2 above. Suppose robust moral realists believe that incest is morally wrong and that their evolutionary counterparts morally disagree. What explains their disagreement is the difference in evolutionary past: while incest hampered our ancestors' reproductive fitness, it did – by stipulation – not do so for our evolutionary counterparts.⁴⁵ As a result, they don't believe that incest is morally impermissible. Importantly, the assumption that differences in evolutionary trajectory explain differences in fundamental moral outlook is indispensable to the argument above: without it, the argument would cease to be distinctively evolutionary. Instead, it would just be a generic argument from moral peer disagreement against robust moral realism.

Second, any evolutionary debunking argument based on moral peer disagreement presupposes that our evolutionary counterparts count as our peers. More precisely, those counterparts must be comparable to us both epistemically and metaphysically. Epistemically, they must be equally well placed to us: their evidence must be similarly strong and their intellectual abilities comparable.⁴⁶ If their evidence was lacking or significantly impoverished and/or their reasoning capacities were impaired, we couldn't consider them our epistemic equals. That we are epistemic peers is important. For if our evolutionary counterparts weren't epistemic peers, their disagreement with us about fundamental moral matters wouldn't defeat the justification of our moral beliefs. Metaphysically, those counterparts need to partake in our robust moral reality: the moral facts that hold in their world must significantly overlap with those holding in the actual world. Or, more poetically: we both must be bound by most of the same moral laws (and seeking to uncover them). Why that? Because we couldn't have meaningful disagreement otherwise, let alone disagreement between epistemic peers. Rather, we would be talking past each other – just like two people 'disagreeing' over which ice cream flavor tastes best.⁴⁷ So, the argument must assume that our evolutionary counterparts are peers, both epistemically and metaphysically.

But, on reflection, those two assumptions are inconsistent – or, at least, stand in serious tension. Suppose that the first assumption is true: our counterparts radically disagree with us about moral matters because of their radically different evolutionary history. That seems to cast doubt on their putative status as both epistemic and metaphysical peers. To begin with, if their evolved moral beliefs are radically different to ours, yet still rationally formed, they probably rest on very different bodies of moral evidence, including moral intuitions and morally relevant non-moral facts.⁴⁸ But if their evidence so significantly differs from ours that it becomes unintelligible to us and their moral outlook strikes us as completely alien, it seems difficult to count them as our epistemic peers.⁴⁹ So, our evolutionary counterparts don't seem equally well placed to us, epistemically speaking.

Further, the first assumption also undermines their putative status as metaphysical peers. Suppose that our counterparts hold radically different moral beliefs due to alternative evolutionary pressures. If their moral beliefs are radically different, however, it seems plausible that other, morally relevant, non-doxastic features of their psychology would differ from ours as well. For instance, they might not be able to experience pain (or only certain kinds of pain) or they might lack the emotion of romantic love. But if our evolutionary counterparts experience and navigate the (moral) world so differently, why think that the same robust moral facts hold for them as for us? After all, many moral facts depend on the morally relevant, non-doxastic features of our psychologies (as even robust moral realists would admit).⁵⁰ For instance, take the fact that needlessly inflicting pain on others is morally wrong. That fact only holds if there are subjects capable of experiencing pain. If our evolutionary counterparts couldn't experience pain, they would not be bound by that moral fact. Similarly, if they were incapable of experiencing romantic love, many moral laws that specifically govern romantic interpersonal relationships wouldn't apply to them. For instance, it might not be wrong for our evolutionary counterparts to cheat on each other. So, it seems plausible that, if our evolutionary counterparts had very different moral beliefs

and thus different non-doxastic moral psychologies, the robustly moral facts that hold for them would differ as well. That, though, means that they cannot be our peers, metaphysically speaking. After all, those counterparts do not partake in our robust moral reality (or they aren't bound by most of the same moral laws).

In sum, it appears that the truth of the first assumption, namely that our counterparts have evolved to radically morally disagree with us, undermines the second assumption, namely that our counterparts are our peers, both epistemically and metaphysically. As a result, those two assumptions are inconsistent – or, at least, stand in serious tension. But, importantly, both assumptions are indispensable to any evolutionary debunking argument based on moral peer disagreement. We cannot just relax them.⁵¹ So, evolutionary debunkers such as Bogardus (2016) and Mogensen (ms, 2017) face a serious worry: it does not seem plausible that evolutionary considerations amount to evidence of moral peer disagreement. That, in turn, means that their argument doesn't get off the ground.

Where does that leave us? In this sub-section, I sketched how evolutionary debunkers such as Bogardus (2016) and Mogensen (ms, 2017) can avail themselves of the resources of the Total Evidence View to get around the self-defeat objection. Their key move consists in establishing that moral disagreements with evolutionary counterparts are structurally similar to **Offside Call** – and then adopt Kelly's (2010) diagnosis of such standard cases. However, that only works if we can plausibly suppose that evolutionary consideration supply evidence of moral peer disagreement. That supposition, though, strikes me as implausible. On reflection, we cannot simultaneously assume that our counterparts have evolved to radically morally disagree with us – and count them as our peers, both epistemically and metaphysically. So, their key move falters and, by implication, they cannot make good on the second part or (B) of my strategy. But if they cannot do that, the evolutionary debunking argument from moral peer disagreement cannot avoid self-defeat.⁵²

7. Conclusion

In this paper, I closely examined an evolutionary debunking argument based on the epistemic principle that higher-order evidence of error defeats justification. More precisely, I argued that any such argument ultimately cannot avert self-defeat. After presenting both the argument and the worry in more depth, my argument proceeded in two steps. First, I sketched out an initially promising strategy against self-defeat: evolutionary debunkers should reject Conciliationism and instead explore Kelly's (2010) Total Evidence View as their background view on the epistemic significance of higher-order evidence of error. However, as I argued in a second step, both versions of the evolutionary debunking argument from higher-order evidence of error fail to take advantage of that switch. More specifically, both versions fall short of establishing that the total evidence available to the robust moral realist defeats the justification of their moral beliefs. Debunkers such as Street (2006), who construe evolutionary considerations as evidence of moral unreliability, lack a plausible theoretical notion of evidential weight. In contrast, the evolutionary disagreement argument developed by Bogardus (2016) and Mogensen (ms, 2017) rests on two inconsistent assumptions about the epistemic (and metaphysical) credentials of our evolutionary counterparts.

My overall argument has two interesting implications. First, it implies that the prospects of developing an evolutionary debunking argument by appeal to higher-order evidence of error look dim. After all, such an argument runs into trouble, irrespective of which more general (background) view about the epistemic significance of higher-order evidence we endorse. If we accept Conciliationism (and thus *Independence*), the evolutionary debunking argument from higher-order evidence of error defeats itself, as Vavova (2015) persuasively claims. In contrast, if we accept the Total Evidence View as the correct account of how to rationally respond to higher-order evidence, the argument doesn't really get off the ground and thus remains toothless, as my line of thinking in §4-§6 suggests. So, evolutionary debunker should strongly examining alternative ways

of spelling out the epistemology of evolutionary debunking arguments. For instance, they might revisit Richard Joyce's (2006) thought that evolution, not mind-independent moral facts, best explains our moral beliefs.

Second, my discussion holds lessons for both proponents of *Conciliationism* and the *Total Evidence View*. After all, both views face issues when applied to evolutionary debunking. To start with and generally speaking, Conciliationism (and especially *Independence*) seems to lead to self-defeat whenever the higher-order evidence of error in question concerns an entire domain of inquiry, as is the case with the evolutionary debunking argument from higher-order evidence outlined in §2. This suggests that proponents of the view must look into ways to non-arbitrarily restrict the scope of *Independence*. If they cannot do that, their view won't be able to accommodate theoretically interesting instances of global defeat by higher-order evidence of error. In contrast, the Total Evidence View seems to hit an obstacle to generalization: without a plausible theoretical notion of evidential weight, the view won't be able to successfully theorize about the epistemic significance of higher-order evidence of error other than peer disagreement. So, my discussion points towards areas in which both views require refinement.

8. Bibliography

Ballantyne, Nathan (2013). The Problem of Historical Variability. In: *Disagreement and Skepticism* 46: 239.

Bogardus, Tomas (2009). A Vindication of the Equal-Weight View. *Episteme* 6 (3):324-335.

Bogardus, Tomas (2016). Only All Naturalists Should Worry About Only One Evolutionary Debunking Argument. *Ethics* 126 (3):636-661.

Brosnan, Kevin (2010). Do the evolutionary origins of our moral beliefs undermine moral knowledge? *Biology and Philosophy* 26 (1):51-64.

Christensen, David (2007). Epistemology of disagreement: The good news. *Philosophical Review* 116 (2):187-217.

Christensen, David (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass* 4 (5):756-767.

Christensen, David (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research* 81 (1):185-215.

Christensen, David (2011). Disagreement, Question-Begging, and Epistemic Self-Criticism. *Philosophers' Imprint* 11 (6).

Clarke-Doane, Justin (2012). Morality and Mathematics: The Evolutionary Challenge. *Ethics* 122 (2):313-340.

Clarke-Doane, Justin (2015). Justification and Explanation in Mathematics and Morality. Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol. 10.

Clarke-Doane, Justin (2016). Debunking and Dispensability. In: Uri D. Leibowitz & Neil Sinclair (eds.), *Explanation in Ethics and Mathematics: Debunking and Dispensability*. Oxford University Press.

Clarke-Doane, Justin (2017). What is the Benacerraf Problem? In: Fabrice Pataut (ed.), *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity*.

Clarke-Doane, Justin (forthcoming). Debunking Arguments: Mathematics, Logic, and Modal Security. In: Robert Richards and Michael Ruse (ed.), *Cambridge Handbook of Evolutionary Ethics*. Cambridge University Press.

Clarke-Doane, Justin & Baras, Dan (forthcoming). Modal Security. *Philosophy and Phenomenological Research*.

Cohen, G. A. (2000). If you're an egalitarian, how come you're so rich? *Journal of Ethics* 4 (1-2):1-26.

Das, Ramon (2016). Evolutionary debunking of morality: epistemological or metaphysical? *Philosophical Studies* 173 (2):417-435.

de Cruz, Helen ; Boudry, Maarten ; de Smedt, Johan & Blancke, Stefaan (2011). Evolutionary Approaches to Epistemic Justification. *Dialectica* 65 (4): 517-535.

Elga, Adam (2007). Reflection and disagreement. *Noûs* 41 (3):478–502.

Elga, Adam (unpublished manuscript). Lucky to be rational. URL = <
www.princeton.edu/~adame/papers/bellingham-lucky.pdf> (8.2.2018)

Enoch, David (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford University Press UK.

Feldman, Richard (2006). Epistemological puzzles about disagreement. In Stephen Hetherington (ed.), *Epistemology Futures*. Oxford University Press.

Fraser, Benjamin James (2014). Evolutionary debunking arguments and the reliability of moral cognition. *Philosophical Studies* 168 (2):457-473.

Gert, Bernard, 2005, *Morality: Its Nature and Justification*, Revised Edition, New York: Oxford University Press.

Gert, Bernard and Gert, Joshua (2017), "The Definition of Morality", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/morality-definition/> (8.2.2018).

Goldman, Alvin I. (2001). Experts: Which ones should you trust? *Philosophy and Phenomenological Research* 63 (1):85-110.

Huemer, Michael (2005). *Ethical Intuitionism*. Palgrave Macmillan.

Joyce, James M. (2005). How probabilities reflect evidence. *Philosophical Perspectives* 19 (1):153–178.

Joyce, Richard (2006). *The Evolution of Morality*. MIT Press.

Kahane, Guy (2011). Evolutionary Debunking Arguments. *Noûs* 45 (1):103-125.

Kelly, Thomas (2005). The epistemic significance of disagreement. In John Hawthorne & Tamar Gendler (eds.), *Oxford Studies in Epistemology*, Volume 1. Oxford University Press. pp. 167-196.

Kelly, Thomas (2010). Peer disagreement and higher order evidence. In Alvin I. Goldman & Dennis Whitcomb (eds.), *Social Epistemology: Essential Readings*. Oxford University Press. pp. 183--217.

King, Nathan L. (2012). Disagreement: What's the Problem? or A Good Peer is Hard to Find. *Philosophy and Phenomenological Research* 85 (2):249-272.

Klenk, Michael (2018). Evolution and Moral Disagreement. *Journal of Ethics and Social Philosophy* 14 (2).

Klenk, Michael (2019). Objectivist conditions for defeat and evolutionary debunking arguments. *Ratio*: 14.

Korman, Daniel Z. & Locke, Dustin (forthcoming). Against Minimalist Responses to Moral Debunking Arguments. *Oxford Studies in Metaethics*.

Kyriacou, Christos (forthcoming). Are Evolutionary Debunking Arguments Self-Debunking? *Philosophia*:1-16.

Lackey, Jennifer (2008). A justificationist view of disagreement's epistemic significance. In Alan Millar Adrian Haddock & Duncan Pritchard (eds.), *Proceedings of the Xxii World Congress of Philosophy*. Oxford University Press. pp. 145-154.

Lasonen-Aarnio, Maria (2014). Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research* 88 (2):314-345.

Lord, Errol (2013). From Independence to Conciliationism: An Obituary. *Australasian Journal of Philosophy* (2):1-13.

Lord, Errol & Maguire, Barry (eds.) (2016). *Weighing Reasons*. Oxford University Press USA.

Matheson, Jonathan (2015). Disagreement and Epistemic Peers. *Oxford Handbooks Online*.

May, Joshua (2018). *Regard for Reason in the Moral Mind*. Oxford University Press.

Mogensen, Andreas (ms). *Evolutionary Debunking Arguments*. DPhil Thesis. Oxford.

Mogensen, Andreas L. (2017). Contingency Anxiety and the Epistemology of Disagreement. *Pacific Philosophical Quarterly* 97 (2).

Schechter, Joshua (unpublished manuscript). Luck, Rationality, and Explanation. URL = <
[http://www.brown.edu/Departments/Philosophy/onlinepapers/schechter/LuckRationalityExp
lanation.pdf](http://www.brown.edu/Departments/Philosophy/onlinepapers/schechter/LuckRationalityExplanation.pdf) > (26.2.2017).

Schoenfield, Miriam (2013). Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief. *Noûs* 47 (1):193-218.

Schoenfield, Miriam (forthcoming). Meditations on Beliefs Formed Arbitrarily. *Oxford Studies in Epistemology* (forthcoming).

Setiya, Kieran (2013). *Knowing Right from Wrong*. Oxford University Press.

Shafer-Landau, Russ (2003). *Moral Realism: A Defence*. Oxford University Press.

Sher, George (2001). But I Could Be Wrong. *Social Philosophy and Policy* 18 (02): 64-78.

Sinclair, Neil (forthcoming). Belief Pills and the Possibility of Moral Epistemology. *Oxford Studies in Metaethics* 13 (forthcoming).

Sterpetti, Fabio (2015). Are Evolutionary Debunking Arguments Really Self-Defeating? *Philosophia* 43 (3):877-889.

Street, Sharon (2005). A Darwinian dilemma for realist theories of value. *Philosophical Studies* 127 (1):109-166.

Vavova, Katia (2015). Debunking Evolutionary Debunking. *Oxford Studies in Metaethics* 9:76-101.

Vavova, Katia (2016). Irrelevant Influences. *Philosophy and Phenomenological Research* 93 (3).

Wedgwood, Ralph (2007). *The Nature of Normativity*. Oxford University Press.

Wedgwood, Ralph (2010). The moral evil demons. In Richard Feldman & Ted A. Warfield (eds.), *Disagreement*. Oxford University Press.

White, Roger (2010). You just believe that because.... *Philosophical Perspectives* 24 (1): 573- 615.

Woods, Jack (2016). Mathematics, Morality, and Self-Effacement. *Noûs*.

¹ I am grateful to Dan Baras, Marius Baumann, Joshua Blanchard, James Brown, Matthew Chrisman, David Enoch, Giada Fratantonio, Camil Golub, David Plunkett, Duncan Pritchard, Geoff Sayre-McCord, Neil Sinclair, and especially Michael Klenk and Paul Silva for helpful feedback on previous drafts. I would also like to thank audiences at the Czech Academy of Sciences in Prague, the University of Edinburgh and the University of Nottingham for their input. This paper was written while holding a Swiss National Science Foundation Doc.Mobility Fellowship and a Janggen-Poehn Foundation Scholarship.

² See Enoch 2011, Huemer 2005, Shafer-Landau 2003 or Wedgwood 2007. Henceforth, I am only concerned with this view as a target of evolutionary debunking. All references to ‘moral realism’ should be understood accordingly.

³ See Kahane 2011. My discussion only concerns *epistemological*, not *metaphysical*, evolutionary debunking arguments. (For the latter, see Das 2016.) Further, I am only interested in *global*, not *local*, evolutionary debunking arguments: arguments that call into question all moral beliefs, robustly construed. Finally, for brevity, I shall omit the qualifier ‘realistically construed’ henceforth, trusting any instance of ‘moral belief’ is understood accordingly.

⁴ Vavova (2015, 2016) has contributed most to clarifying, reframing and criticizing evolutionary debunking arguments in terms of higher-order evidence of error. My discussion is heavily indebted to her framework and exclusively tackles this strand of evolutionary debunking. Other strands rely on the idea that our moral reliability requires explanation (e.g. Clarke-Doane 2012, Locke & Korman forthcoming) or the thought that evolution, not mind-independent moral facts, best explains our moral beliefs (e.g. Woods 2016). I will say a bit more about the problems the former faces in §3 below.

⁵ See Vavova’s (2015) compelling reconstruction. Whether there are other plausible interpretations is an exegetical matter beyond the scope of this paper.

⁶ See Christensen 2007, 2009, 2011 or Elga 2007.

⁷ See also May 2018.

⁸ See Christensen 2011, Lord 2013.

⁹ De Cruz et al. 2011, Kyriacou forthcoming and Sterpetti 2015 discuss a self-defeat worry for evolutionary debunking arguments. However, their discussion isn’t set within a higher-order evidence framework and fails to engage at all with Vavova’s (2015) specific objection.

¹⁰ Plausibly, the higher-order evidence of error would also need to be good, strong or weighty *enough*. Most common cases, including the ones below, intuitively meet that threshold. For more on the issue of evidential weight in more controversial cases such as evolutionary debunking, see §6.1 below.

¹¹ See Christensen 2010 and Lasonen-Aarnio 2014. Some higher-order evidence is evidence of epistemic success, not error (e.g. visiting your optometrist might confirm your visual reliability).

¹² See Elga 2007.

¹³ See Elga ms and Lasonen-Aarnio 2014.

¹⁴ Other kinds include following incorrect epistemic rules, following correct epistemic rules incorrectly, etc. For more, see Lasonen-Aarnio 2014: 315. Evolutionary debunkers focus on the two features in the main text, though.

¹⁵ Nothing below hinges on characterizing Conciliationism in terms of *justification* defeat. Instead, it could be defined in other epistemic terms (e.g. rationality, reasonability, confidence).

¹⁶ See Christensen 2007, Elga 2007, Feldman 2006, Bogardus 2009 and Matheson 2015.

¹⁷ See, for example, Lasonen-Aarnio 2014 or Vavova 2016.

¹⁸ For that very reason, the argument does not collapse into an argument for radical external world skepticism. To put it in Vavova's (2015: §3 & 4) terms, the argument is committed *Good*, not *No Good*.

¹⁹ See Matheson 2015.

²⁰ See Christensen 2011 in the context of peer disagreement. For criticism, see Lord 2013. Both evolutionary debunkers and their critics endorse *Independence*. See Bogardus 2016: 656 and Vavova 2015: 12.

²¹ For more on that implication, see Christensen 2009: 758f.

²² This is plausibly an instance of the *problem of irrelevant influences*: factors (e.g. upbringing, socio-economic background, gender – or evolution) that partially explain, yet don't give reasons for our beliefs, thus providing higher-order evidence of error. For more, see, for example, Ballantyne 2013, Cohen 2000, Mogensen 2017, Schoenfield 2013 & forthcoming, Sher 2001, Vavova 2016 and White 2010.

²³ See also Vavova 2015 who ultimately thinks that moral realists cannot *recognize* evolutionary evidence of moral unreliability as such. For more, see §3 below.

²⁴ See Fraser 2014 for empirical details.

²⁵ See Ballantyne 2013.

²⁶ In the text, Vavova frames her objection slightly differently: moral realists cannot *recognize* evolutionary evidence as good evidence of error – and that there are thus limits on our ability to get evidence of our own error, arising from the way such evidence works. However, her key move is denying that moral realists have a 'good *independent* reason' (92-96) to doubt their moral beliefs. So, it can be framed as an attack on *Independence* in the context of evolutionary debunking, as I do here. I am grateful to Neil Sinclair for helpful discussion of this point.

²⁷ Vavova also addresses some preliminary responses to her objection. For instance, what if the evolutionary debunker *extends* the *scope* of their attack to include *evaluative* in addition to moral beliefs? In that case, Vavova (2015: 87-89) argues, the evolutionary debunking argument defeats itself as well: now, to respect *Independence*, the moral realist cannot even rely on her beliefs about epistemic principles, including principles about how to evaluate evidence such as *Independence* or Conciliationism. What if the evolutionary debunker *restricts* the scope of their attack to deontological moral beliefs? In that case, Vavova (2015: 93-95) argues, their argument either collapses into a more ambitious form (such as the one against moral realism) – or the evolutionary story turns out to be idle, given other worries about deontology. Finally, what if the evolutionary debunker insists that her conclusion is merely *dialectical* (such that it establishes moral constructivism, say), not skeptical? According to Vavova (2015: 89), that wouldn't get them off the hook: even if the conclusion is dialectical, the inference of the argument must still go through. But it doesn't, if her objection is correct.

²⁸ The self-defeat objection might be even *more* general, arising for analogue arguments against realist construals of other putatively *apriori* beliefs about mathematics, logic, modality, epistemology or religion. Examining these analogue arguments, however, goes well beyond the scope of this paper.

²⁹ For a critical overview, see Bogardus 2016.

³⁰ Importantly, that assumption is innocuous in that context. For a different line of criticism, see also Klenk 2019, who argues that modal evolutionary debunking arguments depend on a problematic account of defeat.

³¹ For further discussion, see Woods 2016 and – in response – Baras & Clarke-Doane forthcoming.

³² I am grateful to Michael Klenk for pressing me on this point.

³³ Why not reformulate *Independence* instead of rejecting it? Because the most plausible ways of doing so strikes me as highly problematic. For instance, evolutionary debunkers might first distinguish between *substantive* and *formal* assumptions about moral truth (e.g. Gert 2005, Gert & Gert 2017, Sinclair forthcoming: 19-22) – and then argue that *Independence* tells us to set aside only substantive, yet not formal, moral assumptions when assessing higher-order evidence of error.

However, it is often difficult to draw the formal/substantive distinction, formal moral assumptions (such as *Motivational Internalism*) are often highly controversial even among moral realists, and it is unclear whether formal assumptions provide sufficient detail to assess whether moral realists are mistaken about morality.

³⁴ In what follows, I shall assume that epistemic akrasia is rationally impossible (and so-called ‘level-splitting’ therefore not plausible). In other words, it can never be rational to have high confidence in something like ‘P, but my evidence doesn’t support P’. Even though non-trivial, that assumption strikes me as very plausible. For a defense, see Horowitz 2014. For further critical discussion, see Lasonen-Aarnio 2014 and Silva 2016. Thanks to Paul Silva for alerting me to this.

³⁵ Another example would be Jennifer Lackey’s (2008) *Justificationism*. However, those two views ultimately converge, according to Matheson 2015.

³⁶ In football, delay of game is, at best, worthy of a yellow card. Players get awarded yellow cards for bad fouls that don’t warrant immediate ejection. Two yellow cards, though, equal a red card, which signifies immediate rejection. This case is modelled after so-called *Extreme Restaurant Cases*. See Christensen 2007: 199-203, Elga 2007: 490f, and Kelly 2010: 41f.

³⁷ See Kelly 2010: 37. To what extent the assumption of independent convergence holds is, of course, a psychological and sociological, not philosophical, matter. But for more on its epistemic importance, see Kelly 2010: 37-41 and Goldman 2001: 150-56.

³⁸ Following Kelly (2010), my presentation of the view focuses on peer disagreement here. But, at least initially, the view seems to plausibly generalize to other kinds of higher-order evidence of error. For more on whether that impression withstands scrutiny, see §6.1 below. For some common objections, see Kelly 2010: 42-64.

³⁹ J. Joyce (2005) distinguishes between the *balance* and *weight* of evidence. While the balance of evidence concerns whether a given body of evidence ‘points’ towards one set of propositions over another, its weight corresponds to the size of a body of evidence. Given his usage, an evolutionary debunker therefore requires a theoretical notion of evidential balance, not weight. However, since the distinction isn’t relevant in our discussion, I shall continue using them interchangeably.

⁴⁰ It might be unsurprising that Kelly’s (2010) view requires supplementation when generalizing it to higher-order evidence of error other than peer disagreement (for which it was conceived and developed). In fact, it might even be unsurprising that *any decent* epistemic theory requires a notion of weight. (For an argument that any decent *moral* theory requires weighted notions such as normative reasons (and thus: an account of their weights), see Lord & Maguire 2016: 1-8.) What *is* surprising, however, is that supplementing Kelly’s view proves *that* hard.

⁴¹ Of course, the total evidence also defeats the justification of Julian’s belief that the scorer was offside, as we shall see shortly.

⁴² For a similar line of thought, see Setiya 2013: Ch. 1.

⁴³ Some might worry that hypothetical disagreement can never be epistemically significant. But that strikes me as a red herring. For more on that, see Mogensen 2017.

⁴⁴ What follows is a somewhat condensed presentation of the worry I develop in Wittwer 2018: Ch. 3, 55-59. For a more sophisticated, paper-length defense of a similar line of criticism, see also Klenk 2018.

⁴⁵ We can imagine and spell out the empirical details in various ways: perhaps those evolutionary counterparts reproduce very differently, or incest strengthens group cohesion which, in turn, increases survival prospects, etc. For the purpose of our worry, those details don’t matter.

⁴⁶ See Kelly 2005: 10. For criticism, see King 2012.

⁴⁷ Of course, this assumes – plausibly in my mind – a roughly subjectivist or response-dependent account of taste judgments. Further, note that most paradigmatic cases of peer disagreement discussed in the literature trivially meet the criterion of metaphysical peerhood – and therefore don’t spell it out. Take **Offside Call** above: it is uncontroversial that there is a mind-independent (yet perhaps not practice/institution/independent) fact of the matter about whether the striker was offside or not.

⁴⁸ See Wedgwood 2010: 7.

⁴⁹ This point is exacerbated within the framework of the Total Evidence View: we are allowed to rely on our first-order moral evidence when assessing evidence of peer disagreement. But once we do, it becomes hard to regard our radically disagreeing evolutionary counterparts as epistemic peers – whether or not they share our evidence. Thanks to Camil Golub for this point.

⁵⁰ Not all, of course. For instance, it might be a fact that eco-systems have moral value. But their value wouldn't directly depend on any feature of our non-doxastic moral psychologies.

⁵¹ As rehearsed above, if we relax the first one, the argument ceases to be evolutionary – and if we relax the second one, it ceases to be about disagreement between peers.

⁵² Of course, the worry developed in this section might actually be so powerful that it would undermine any evolutionary debunking argument from moral peer disagreement irrespective of self-defeat. (Klenk 2018 draws that conclusion.) After all, it targets the assumption at its very core: that evolutionary considerations amount to such evidence in the first place.