**Higher-Order Defeat in Realist Moral Epistemology[[1]](#footnote-1)**

Brian C. Barnett (State University of New York at Geneseo & St. John Fisher College)

In *Higher Order Evidence and Moral Epistemology*

Ed. Michael Klenk, Routledge Studies in Epistemology

PENULTIMATE DRAFT—PLEASE CITE PUBLISHED VERSION

**Author Bio:** Brian C. Barnett is lecturer in philosophy at the State University of New York at Geneseo and St. John Fisher College in Rochester, New York. He received his PhD in Philosophy from the University of Rochester for a thesis on the nature and significance of higher-order evidence. His primary areas of research are epistemology, philosophy of religion, and the philosophy of nonviolence. Additional interests include logic, abstract mathematics, and formal methods in philosophy. Brian is currently at work in his role as editor of the *Introduction to Epistemology* volume in the *Introduction to Philosophy* open textbook series (ed. Christina Hendricks).

**Abstract:** On an optimistic version of realist moral epistemology, a significant range of ordinary moral beliefs, construed in realist terms, constitute knowledge—or at least some weaker positive epistemic status, such as epistemic justification. The “debunking challenge” to this view grants *prima facie* justification but claims that it is “debunked” (i.e., defeated), yielding the final verdict that moral beliefs are *ultima facie* unjustified. Notable candidate “debunkers” (i.e., defeaters) include the so-called “evolutionary debunking arguments,” the “Benacerraf-Field Challenge,” and persistent moral disagreement among epistemic peers. Such defeaters are best treated as higher-order evidence—viz., evidence contesting the merits of the first-order evidence on which moral beliefs are based. This chapter first develops a theory of higher-order defeat in general, which it then applies to debunking in particular. The result: the challenge fails entirely on epistemic grounds—regardless of whether or not its empirical and metaphysical presuppositions are correct. An advantage of this purely epistemic defense over alternative strategies is that the former extends even to laypersons who themselves lack the expertise necessary to formulate an adequate response. However, this leaves open the prospects for non-epistemological interpretations of debunking (e.g., moral or ontological). The chapter therefore concludes with brief suggestions in that direction.

**Introduction**

According to a *realist moral ontology*, moral properties are real features of the objective, mind-independent world. On a *robust* version of this thesis, these features occupy an ontological realm of their own: they are not part of the natural world, and for this reason are often taken to be causally inefficacious.[[2]](#footnote-2) If *cognitivism* is correct, these non-natural properties manage to figure in propositions that serve as the contents of moral beliefs, which are true or false in the full-fledged sense of bearing a correspondence to reality (the *correspondence theory of truth*). Assuming there’s no systematic reason for all such beliefs to be false (contra some error theories), at least some of them may (are likely to) be true whether or not we cannot tell which. To rule out sheer luck in obtaining moral truth, a *maximally* *robust moral realist package* (hereafter “realism”) combines the aforementioned views with an *optimistic moral epistemology*: many of our ordinary moral beliefs constitute knowledge, or at least some weaker positive epistemic status, such as epistemic justification.[[3]](#footnote-3)

Each component of this package—robust moral realism, cognitivism, the correspondence theory of truth, the denial of error theory, optimistic moral epistemology—has been the locus of attack.[[4]](#footnote-4) But how the package hangs together raises difficulties in its own right, since some of its components seem unlikely bedfellows. My focus here is whether optimistic moral epistemology is plausible given the rest of the package. Specifically, I will address a generalized epistemological version of the *debunking challenge*. This challenge grants the non-epistemic components of realism, along with some modest epistemological footing: many of our ordinary moral beliefs are *prima facie* justified. However, the debunking challenge purports to “debunk” this justification by establishing that realist assumptions yield defeaters (“debunkers”), so that moral beliefs are not *ultima facie* justified.

Among the most prominent candidate debunkers are the so-called *evolutionary debunking arguments*[[5]](#footnote-5) (*EDAs*), which begin with the premise that our moral beliefs are shaped by Darwinian processes (e.g., random gene introduction and mutation, genetic inheritance, and their survival value)—factors that have no connection to a non-causal, non-natural moral domain. As such, we should not expect moral beliefs to track the truth. They are unreliable, and therefore defeated (at least upon grasping the relevant facts). The problem generalizes to beliefs about any non-causal, non-natural ontological realm, such as Platonic abstracta, yielding the so-called *Benacerraf-Field Challenge* (*BFC*).[[6]](#footnote-6)

Another prominent candidate debunker arises in the disagreement literature.[[7]](#footnote-7) Given widespread and persistent moral disagreement, realism implies that we cannot consistently maintain our moral beliefs without dismissing those of others. Dismissal is epistemically unproblematic in cases in which we can rationally attribute a greater probability of error to interlocuters. However, it appears we are often confronted with *epistemic peers*—those who share the same information, are just as smart, equally evidentially responsive, and the like. In short, they are just as likely as we are to be right.[[8]](#footnote-8) According to *conciliationism* (in its strongest form[[9]](#footnote-9)), the rational response is to suspend judgment—i.e., to give up the contested beliefs.[[10]](#footnote-10) If correct, the ramifications are potentially quite skeptical.[[11]](#footnote-11) After all, you can find a smart philosopher who will disagree about nearly anything. Perhaps we can salvage at least a small core of moral beliefs that are sufficiently agreed upon that they are not subject to the problem. But we must consider whether even those beliefs are indirectly called into question by a conciliationist treatment of wider-ranging metaethical disagreement.

Many responses to debunking arguments have questioned their empirical details and metaphysical presuppositions.[[12]](#footnote-12) Unfortunately, these are shaky grounds on which to rest, even for those who are scientifically informed and well trained in metaphysical theorizing. So much the worse for the moderately informed layperson: an agent who possess the minimal information about evolution and disagreement sufficient to give the debunking threats some foothold, but who nevertheless have no recourse to the complex empirical and metaphysical challenges needed to dispel such threats. Realists presumably prefer a response that protects a wide range of ordinary moral beliefs of average folk. And in order to achieve this, we need our response to go epistemic.[[13]](#footnote-13)

In my preferred epistemological framework—internalist evidentialism—the debunking challenge amounts to the claim that there is available evidence such that those who are sufficiently aware of the relevant facts (about evolution, disagreement, etc.) are no longer justified in their moral beliefs (given the remainder of the realist package). More specifically, candidate debunkers fall under a class of evidence of recent interest to epistemologists—*higher-order evidence*, or evidence about evidence.[[14]](#footnote-14) Peer disagreement provides evidence about how others assess the evidence, which in turn yields comparative evidence about how well we have assessed ours. EDAs and the BFC provide evidence that casts doubt on the reliability of the evidence for our moral beliefs (moral intuition, reasoning, testimony, etc.). This higher-order evidence purports to defeat the first-order evidence for our moral beliefs. Whether the debunking challenge succeeds, then, is not determined solely by their empirical and metaphysical claims, but also in part by the nature of higher-order defeat.

My aim in this paper is twofold: one primary and one secondary. My primary aim is to defend an optimistic moral epistemology given the rest of the realist package by appeal to a theory of higher-order defeat in such a way that the defense extends even to laypersons who are not themselves in a position to formulate a response. My secondary aim is to introduce a theory of higher-order defeat and demonstrate its utility in moral epistemology. I first outline the theory in the next section, which I then apply to candidate debunkers in subsequent sections. I will show that they fail on purely epistemic grounds, even granting their empirical and metaphysical premises. Despite this failure, I conclude in the final section by suggesting alternative roles that debunking may continue to play in moral epistemology.

**Conditions on Higher-Order Defeat[[15]](#footnote-15)**

The proper starting point in a theory of higher-order defeat is the concept of higher-order evidence. Roughly, higher-order evidence is evidence about evidence. The evidence it is about is *lower-order evidence*, which may in turn be about further evidence. At some point, we reach *first-order evidence*: evidence directly concerning the proposition at the *object level* (one that is not evidence for some further proposition). Suppose, for example, that E2 is evidence about evidence E1, which is evidence about proposition P, which is at the bottom of the evidential chain. Then E2 is higher order (specifically second order), E1 is lower order (specifically first order), and P is at the object level.

If E2’s being about E1 suffices to render it higher order, then when some X is added to E2, E2+X continues to be about E1 and is likewise higher order, even if X is first order. Also notice that E2’s being about E1 in one way does not rule out E2 being about E1 in additional ways, nor does it rule out E2 being about some additional evidence E1\*. I call this *supportive complexity*. We will soon see that such seemingly trivial details matter.

Higher-order evidence raises two primary questions about evidential support. First, what bearing, if any, does evidence at a higher level have on the object level? Second, the question of *levels interaction*: how do different evidential levels interact when combined? What, in other words, does the total evidence support?

One promising answer to the first question is that evidential support can “filter” down from higher levels, through lower levels, and to the object level. As Feldman (2006) puts it, “evidence of evidence is evidence.” Here’s a more careful formulation:

*Filtration Principle:* If E2 is evidence that there is evidence E1 for P, then E2 is evidence for P.

For example, let E2 be evidence that a reliable mathematician sincerely tells you that there is a sound proof (E1) for theorem P. If all you know is E2, you have enough evidence to believe P. So, filtration is intuitively plausible.[[16]](#footnote-16) If true, this has immense epistemic importance, especially when one possesses higher-order evidence but lacks access to the first-order evidence. It potentially explains the evidential story in many cases of testimony,[[17]](#footnote-17) and how we retain justification in cases in which we no longer remember our original evidence but nevertheless remember having once had evidence (a solution to the problem of “forgotten evidence”[[18]](#footnote-18)).

Despite its plausibility, filtration has at least three exceptions. First, Hud Hudson observes that one can have a defeater D that defeats one’s support for P without defeating E2’s support for the claim that E1 supports P.[[19]](#footnote-19) For example, Hud convincingly lies to Rich that it’s his birthday, giving Hud evidence E2 that Rich has evidence E1 that it’s Hud’s birthday (P), even though Hud knows it’s not his birthday (~P). This is because he possesses prior evidence D against P, which defeats E2 but has no bearing on E1’s support for P. Feldman responds that this doesn’t refute filtration, since E2 continues to support P. What’s true is that E2&D does not support P. I agree. However, just relocate the defeater and voila: if E2\* = E2&D, then E2\* supports that E1 supports P, even though E2\* does not support P, contrary to filtration. So, filtration occurs only when the higher-order evidence does not contain defeaters.

A second exception arises when one doesn’t understand much about what one has evidence for. So, suppose that E2 supports that E1 is a sound argument for P. If E2 is testimonial evidence given by me in my logic class to a student who knows the standard definition of deductive validity but has a confused understanding of it, then perhaps E2 does not yield evidence for P. E2 needs to contain adequate conceptual information about the propositions it supports in order for filtration to kick in.

For the third exception, suppose we do a little reflection in epistemology class and conjure up some intuitive evidence E2, which supports that certain experiences (or propositions) E1 are evidence of trees. Contrary to filtration, if no one is having those experiences (or reason to think the propositions are true), then E2 doesn’t give us evidence of trees. In order for E2 to support P, it needs to *ground* E1 by supporting that E1 is an occurrent experience or contains true propositions.

Only in some cases does filtration require E2 to ground E1 or contain adequate conceptual information about E1’s relation to P. E2 needs to meet these conditions in order to support P *in virtue of its support for the claim that E1 supports P*. But due to the possibility of supportive complexity, E2 can support P in other ways. For example, E2 might also support that E1\* supports P. If E2 grounds E1\*, then it can support P via this route without grounding E1.

So, there are various restrictions and some of them only apply some of the time. But suppose that we have a case in which filtration occurs. Then it is crucial to note what I call the *dampening effect*. Suppose that E1 supports P to degree N1. And suppose that E2 supports to degree N2 that E1 supports P. I claim that in general, even when E2 meets the conditions for filtration, E2 supports P to some degree M < N1, N2. This is because the further the evidential distance from a proposition, the greater risk, the less likely it is true, the weaker the support—i.e., the more it has dampened. To illustrate the simplest case, where N1=N2=N: you are confident to degree N that you have a proof of theorem T vs. you are confident to degree N that you have a proof that there is a proof of theorem T. It seems to me the first justifies more confidence in T. However, dampening does not entail that higher-order evidence always provides weaker support at the object level than does first-order evidence, since for example higher-order evidence can accumulate in the absence of first-order accumulation.

We now turn to cases in which one has both higher-order and lower-order evidence, which raises the question: what does the total evidence E1+E2 support? In cases in which E2 is *friendly* to E1 (i.e., E2 supports E1, or agrees with E1 by supporting that it bears the relation to P that E1 actually does bear to P), there is little difficulty deciding how the two levels interact. Although there may be disputes about whether E1+E2 supports P more than E1 alone, there seems to be widespread presupposition that E1+E2 supports P. In *hostile cases*, E2 challenges E1 in some way. There are also *inert* cases—cases in which E2 is neither friendly nor hostile. Although this last case is not obvious and it will play a crucial role in my treatment of debunking, we will come back to it later. For now, focus on hostility.

The challenge to E1 in hostile cases can be *tangential* to E1: E2 supports that some further evidence E1\* is evidence against P. Then via filtration E2 is evidence against P, thus serving as a *rebutting defeater* for E1. The challenge to E1 can also be *direct*: E2 might challenge E1’s intrinsic merits (e.g., that it contains false propositions, or that it is a blurry perception) or challenge E1’s bearing on P, thereby serving as a potential *undercutting defeater*.

Now comes what I call the “latching” problem. Suppose E2 is evidence that some evidence supports ~P, but you don’t know what that evidence is. Unbeknownst to you, it is the evidence E1 that you are already relying on to support P. It cannot *latch onto* E1, and so cannot serve as an undercutter.[[20]](#footnote-20) From your perspective, E2 might as well be about some other evidence of which you know not. Even though it cannot undercut your evidence, E2 still has the potential to rebut E1 as evidence against P. But in order to do so, E2 will have to be evidence against P by meeting the conditions necessary for filtration. In that case, its negative impact is lessened due to dampening, and can more easily be withstood by the first-order evidence.

A few last notes about levels interaction. First, as with filtration, the above claims won’t apply in special cases due to conceptual inadequacy and supportive complexity. Second, we need a no-defeater clause, since even in a case of higher-order defeat, the defeater can itself be defeated. Third, defeat can be partial. When I don’t make that qualification, I mean full defeat, since that is the primary aim of candidate debunkers. So, when hostility leads to at least partial defeat, we need a condition that guarantees it meets some degree threshold for full defeat. One plausible view is *calibrationism*, which is roughly the view that the degree to which P is supported should match the degree to which E2 says E1 supports it.[[21]](#footnote-21) For now, we’ll just say that the degree has to be “sufficient” as a placeholder, and leave details to be worked out as the need arises.

So, the theory is this: Except in special cases (due to conceptual inadequacy and supportive complexity), higher-order evidence E2 (fully) defeats an agent’s lower-order evidence E1 concerning P (i.e., E1 supports P but E1+E2 does not support P) iff all of the following obtain:

1. The agent possesses E2;
2. The agent does not possess an undefeated defeater for E2;
3. E2 is directly or tangentially hostile to E1;
4. If E2 is tangentially hostile and has an object-level bearing in virtue of meeting the conditions necessary for filtration, it has at least some rebutting power, albeit dampened;
5. If E2 is directly hostility and latches onto E1, it has at least some undercutting power; and
6. E2’s hostility is sufficiently strong: its rebutting or undercutting power (the greater of the two if E2 has the potential for both) surpasses the minimum threshold.

**Debunkers as Higher-Order Defeaters**

Having outlined the theory of higher-order defeat, we now apply it. In this section, we see how candidate debunkers can be thought of as potential higher-order defeaters, make an initial assessment of whether they meet the conditions of the theory, and in the process see why existing responses to debunking fall short, thereby motivating the need for a new solution.

We’re granting for the sake of argument that we have evidence for our first-order moral beliefs. By definition, candidate debunkers aim to challenge this evidence, and so appear to be hostile. Indeed, by definition they purport to be defeaters. However, they cannot consistently purport to offer direct evidence against first-order moral beliefs, since that would be to put forward exactly what they contest.[[22]](#footnote-22) Candidate debunkers must pose higher-order challenges, and the theory of higher-order defeat therefore applies.

As a precondition of the theory, candidate debunkers must be possessed by the agents in question. Those who’ve never been exposed to the relevant considerations are not subject to epistemic threat. This is not a trivial point. Determining the relevant group of agents is more difficult than it might otherwise seem. One extreme is to think that only a handful of specialists who have read the literature, or discussed it in a seminar, are sufficiently aware of the relevant considerations. However, one need not know the details provided by such contexts for the challenge to kick in. It doesn’t take much to realize that there’s a lot of moral disagreement among smart, informed people. And most adults nowadays know about evolution, and at least some of the more reflective of them might have realized that this affects our moral judgments. It is plausible that these rudimentary considerations are enough to get the threat going. So, I assume that a significant percentage of average folk are in possession of at least one candidate debunker. Besides, protection from the threat by appeal to ignorance is unsatisfactory. After all, those of us attempting to respond to the debunking challenge are well informed, and we too want our moral beliefs to be safe! So, minimizing the relevant group doesn’t provide a satisfactory response.

The other precondition to be satisfied is exclusion from special cases, the first of which is easily dismissed: the group we have settled on (ranging from most average folk to experts) presumably has a sufficient handle on the basic concepts, such as evidence and support, and can carry out simple valid deductions. The other special case is supportive complexity. There is a simple trick for obtaining supportive complexity: plug into the theory one’s total body of higher-order evidence, which contains debunking evidence as a mere proper part. I will follow a more useful approach: isolate the debunking evidence, plug this alone into the theory, see what results, then separately consider after the fact whether the remaining evidence defeats those results. Since we are limiting consideration to consistent candidate debunkers (ones that do not presuppose first-order moral claims), there is no reason to think the isolated debunking evidence will enter into any evidential relations other than those which it wears on its sleeve: its negative relations to first-order moral evidence. So, with the proper construal of the higher-order evidence under consideration, we can safely set aside supportive complexity.

Having identified a relevant group of agents and ruled out special cases, the theory of higher-order defeat kicks in and its conditions must be applied. First, we should decide whether direct or tangential hostility is at issue. Some treatments of disagreement rely on the filtration principle: finding out that a peer disagrees yields evidence that there is evidence against one’s view, which via filtration is evidence against the view, thereby yielding tangential hostility.[[23]](#footnote-23) This is a mistake, even if the conditions necessary for filtration are satisfied. The dampening effect weakens the bearing of the evidence from disagreement on our moral beliefs, increasing the ease with which it can be outweighed by the first-order evidence. So, if treated as tangential hostility, the support for moral beliefs can remain intact in the face of disagreement albeit to a weakened degree. Of course, conciliationists might observe that we typically know numerous peers who disagree on any given topic, and so might make the argument that dampened support accumulates. However, as frequently observed, it is dubious that our peers’ beliefs are evidentially independent of one another, and to treat them as such would be illegitimate double-counting of the available support.[[24]](#footnote-24) So, it is doubtful that repeated instances of dampened support accumulate. If it does, then presumably not according to a linear measure, thereby limiting the prospects for any cumulative effect to surpass the threshold necessary to fully override the competing first-order evidential support.

There is a more promising treatment of disagreement. Discovering that a peer’s evidential assessment conflicts with one’s original assessment yields reason to doubt whether the evidence really supports what one initially thought—i.e., it’s directly hostile. The case for direct rather than tangential hostility is even clearer in the case of EDAs and the BFC, since they explicitly attempt to make trouble for our evidential basis.

Given that candidate debunkers are best treated as cases of direct hostility, the next condition to examine is the latching requirement. But we can easily see that the candidate debunkers are about the very evidence on which we rely.[[25]](#footnote-25) There’s only a little room for disconnect. Mental compartmentalization might do the trick, or Hume’s (in)famous treatment of skepticism: philosophical reflection giving rise to defeaters will inevitably be overturned by habit, at which point we’ll resume normal life and forget about the defeaters. Perhaps something like this can stop us from connecting the dots in everyday moral belief formation, preventing latching, saving us from defeat. But this will depend on how conscious evidence and belief need to be, which is a contentious matter.[[26]](#footnote-26) Moreover, most of us find such solutions unsatisfactory, since we prefer to retain justification even in circumstances of full and integrated awareness. Grant, then, that latching occurs, at least in most circumstances. As a result, there’s no need to resort to the weaker tangential treatment with the dampening effect of filtration.

Given the above, there’s at least some degree of defeat unless the defeaters are defeated. Now we must come to questions of degree. On a weak interpretation, candidate debunkers aim merely to show that our evidence is less reliable than initially thought. Although I will later return to this possibility, I set it aside for now, since something stronger is normally intended: candidate debunkers aim to show that our moral evidence should not be relied upon at all. On this stronger interpretation, since all the previous conditions are met, it follows from calibrationism that there is full defeat—unless the defeater is defeated.

This leads us to the dominant strategy in the current literature: defeat the defeater by providing direct evidence against the debunking premises. Rather than review the vast literature in the limited space here, it suffices to note that there are many versions of this strategy in the literature, each with its counterarguments, both sides appealing to subtle and complex empirical and metaphysical claims. This is not to deny that there is a more plausible side. But the level of dialectic and requisite background knowledge are largely inaccessible to non-experts. The weakness in this approach is that laypersons lack the defeater-defeaters.

It seems, then, that candidate debunkers satisfy all of the conditions on higher-order defeat, for at least a relatively large group of average people—despite the usual strategies for rebuttal. If so, debunking succeeds for the relevant group. Is there a way out?

**The Inertness Thesis**

There is one detail we have missed. I have glossed over whether candidate debunkers are hostile. They purport to be, and upon first inspection, it seems obvious that they are. Certainly, they aren’t friendly. I claim that they are inert (neither friendly nor hostile), and that this is the key to answering the debunking challenge. Seeing this requires a careful examination of the conditions under which non-friendly cases are hostile (hence have defeating power) or inert (thereby lack such power).

First, distinguish *evidential relations* from what I call *external measures of evidence*. The former is any relation in virtue of which something qualifies as evidence concerning a proposition. The latter is any other evaluation of the evidence. External measures capture how well evidence objectively matches reality. For example, consider a visual impression of a tree, which (along with conceptual information about treehood) is evidence that there is a tree. The relationship between the visual impression and the proposition is an evidential relation. Whether there really is a tree (whether the evidence is veridical or misleading) is an external measure. Likewise, the objective probability that there is a tree (given the visual experience) is an external measure. If you prefer, feel free to think of external measures in terms of reliability, safety, tracking, etc., though I will use these terms interchangeability, since the details won’t matter.[[27]](#footnote-27)

Now suppose your tree impression is hallucinatory (or that your environment makes this objectively probable), but you have no indication that this is so. Your tree impression continues to be evidence for the tree proposition despite its objective unreliability. The latter does not negate the former. In general, external measures on evidence do not alter the evidential relations themselves.[[28]](#footnote-28) Of course, if you find out about an external measure (regardless of whether your information about it is correct or misleading), this awareness is new evidence. If hostile, this new evidence undercuts the initial evidence. Still, the initial evidential relation holds. What fails to hold is a positive evidential relation between the new total body of evidence (the perception plus awareness of unreliability) and the proposition that there is a tree. In other words, defeat works not by nullifying an evidential relation that is defeated but by expanding the body of evidence to one that does not bear that evidential relation.

So, while external measures on one’s evidence are inert, awareness of positive external measures (more objectively probable than not) are friendly, and the awareness of negative external measures (less objectively probable than not) are hostile. This does not depend on whether the awareness is veridical. It can be misleading about the external measures on one’s first-order evidence. Whether misleading or veridical, it makes the same contribution to evidential support. Moreover, awareness of external measures is not required for evidential support. In other words, the absence of such higher-order evidence is inert. Think, for example, of a small child, who can have justified beliefs based on perceptual evidence without yet having a clue about external measures. In fact, it would lead to a vicious infinite regress to require that evidential support always requires additional evidence about external measures to confirm it.

Not only is the absence of any information about external measures inert. But awareness of one’s ignorance of such information is inert. Having just observed in the last paragraph that necessarily we have good evidence without other evidence about any external measures to support it, you and I are aware in this moment that we are ignorant in this respect, and surely this does not defeat our justification. Another example: consider adults unindoctrinated into the world of epistemology. When asked why they think what they think, they may be able to identify their evidence and cite it (or maybe not), but when challenged with questions about the objective likelihood that their evidence actually fits the external world (especially in light of skeptical scenarios) and asked how to noncircularly confirm reliability, they probably cannot give answers no matter how hard they try. After all, it’s controversial whether any expert epistemologists are up to the task.[[29]](#footnote-29) But this failure should not mean that our evidence is defeated, on pain of an extreme justificatory skepticism.[[30]](#footnote-30)

On the other hand, suppose you learn that you’ve been drugged. There’s a 50/50 chance of hallucination, which does seem to defeat. We now have a hostile situation. In contrast, the preceding observations make clear that being aware that one is ignorant of objective probability is inert. What’s the difference? In the hostile case, you have reason to think that the objective relation between your evidence and the proposition confers a 50/50 probability (i.e., the evidence confers what I call *neutral support*). In the inert case, there’s an absence of support altogether (neither positive, negative, nor neutral). The evidence simply fails to yield a verdict about what the probability is (not even that it lies within a relevant ballpark). In other words, it is entirely *inscrutable*.

Neutral probability is crucially different from inscrutability.[[31]](#footnote-31) To further illustrate the difference, suppose a quantum mechanical calculation yields a 50/50 probability that an electron will be spin up. Consider a layperson who is given the mathematical apparatus and told what it allows one to calculate but unable to perform the calculation. When asked what the probability is, her answer should be that she doesn’t know, not that the probability is 50/50. The probability is inscrutable to her, not neutral. This difference between a 50/50 probability and inscrutable probability corresponds to a distinction in the philosophy of probability. According to the infamous *indifference principle* (or *principle of insufficient reason*), when there’s no reason to prefer one proposition over another, they automatically receive equal probability. This is now in wide disrepute. In addition to not making sense of the quantum mechanics example, it is a well-known result that the principle yields incoherent probability assignments.[[32]](#footnote-32)

Let’s sum up. First, we are concerned here only with unfriendly cases that are directly about one’s first-order evidence, since candidate debunkers are not best treated as tangential. We need to divide the direct unfriendly cases into the hostile and inert categories. The direct unfriendly cases we have determined to be inert are (i) external measures on our evidence, (ii) lack of awareness of such external measures, (iii) awareness of our ignorance of external measures, and (iv) evidence of their inscrutability. The direct unfriendly cases that are hostile are (v) evidence of negative/neutral evidential relations and (vi) evidence of negative/neutral external measures.

To determine whether candidate debunkers are hostile or inert, we need to determine which of these categories they fall under. First consider peer disagreement. By definition, if a person S1 has good reason to believe that another S2 is an epistemic peer, this is evidence for S1 that S2 is as likely to be right. When S1 also finds out that S2 disagrees, S1’s total evidence yields a 50/50 shot that S1 is right. So, peer disagreement yields neutral, hostile support. After all, it is isomorphic to the above drug case, which is well accepted as a case of defeat.[[33]](#footnote-33) As this result shows, the theory of higher-order defeat entails conciliationism about peer disagreement (at least given peerhood as here conceived).

But how often is there good reason to accept peerhood? Although it is often irrational to deny peerhood, let’s be careful what we infer from this. We often think new thoughts that we think haven’t yet occurred to others. Moreover, people frequently change their minds when given new thoughts. So, we should recognize that we sometimes have evidence others lack which might change their minds—not because we are smarter but for practical reasons. Of course, we should also recognize that *others* sometimes have evidence *we* lack which might change *our* minds. And since thinking happens continually, our evidential situations are ever in flux. As a result, in many cases of disagreement with at least near-peers (that much is easily discerned), it’s an open question whether I’m currently in a better position to judge, they’re in a better position to judge, or we’re equally well positioned. So, I might be more likely to be right, they might be more likely, or we might be equally likely.[[34]](#footnote-34) The best description of this situation is that the probability is inscrutable, not 50/50. But peerhood is the minimal comparative evidential status which when combined with disagreement yields (full) defeat. Given this, along with peerhood inscrutability, the inertness thesis yields the conclusion that the total evidence in many ordinary cases of peer disagreement is inert rather than hostile despite being unfriendly.

Two qualifications are in order. First, it’s not always like this. Sometimes we so thoroughly discuss an issue with another to the point that it seems nothing further could be added to tip the scales, in which case it seems clear that we’re on equal footing, yielding a 50/50 chance of being on the right side of the dispute. Such cases yield neutral support, which is hostile and therefore defeats. Plausibly, however, these are uncommon in moral disputes. Even for moral beliefs nearing universal acceptance, there are dissenters motivated by metaethical concerns, the disputes over which reintroduce peerhood inscrutability. If so, many moral disagreements are only of the inert sort.

Second, up to this point we’ve considered only full defeat. Partial defeat I happily concede in a wide range of cases. Even if peerhood is inscrutable, near-peerhood (or some weaker positive status) is more easily discernible. Suppose you receive testimony against your belief, and your evidence affords the testifier some minimal positive evidential status. That there is some contest to your belief is not inscrutable, yet it is insufficiently hostile for full defeat. Its hostility is sufficient for mere partial defeat. Given partial defeat, you should maintain your doxastic stance with decreased confidence. Inscrutability doesn’t save you from this result; it at best offers protection from full defeat.

Turn now to EDAs, which are spelled out in various ways, sometimes without much to do with evolution. Some are purely probabilistic: one might suggest that any given moral belief, being only one of many competing beliefs, has a low probability due to this proportion.[[35]](#footnote-35) But this cannot be right, since it would yield reason to think the belief is probably false, from which one should infer its negation, yielding justification for a different moral proposition instead. This shifts justification rather than removes it. Better would be an argument establishing neutral probability. Such an argument cannot be made purely by appeal to the fact that my belief in P is only one option of two, the other being ~P. This would presuppose equal probabilities, which I have no reason to suspect given our assumption that I already have first-order evidence that seems to point in one direction over the other. No appeal to a lack of evidence about my evidence will yield a neutral probability, for that is to assume the indifference principle, ignoring the possibility of inscrutability. To establish neutral probability, I must be given some evidence that my evidence is as likely reliable as not. Perhaps evolution gives me this evidence by explaining moral beliefs in a way that does not link them up to the moral properties they are about. But the adequacy of such an explanation does not rule out there *also* being a “bridge” between those beliefs and the moral properties they are about, a bridge which grounds reliability. One option is to appeal to the fact that moral properties are non-causal and mind-independent, making it difficult to imagine such a bridge.

This takes us to the BFC, which denies the existence of a bridge on those grounds. But if we are to avoid lapsing back into any of the above strategies, we need to be given evidence that renders the probability of the bridge less than or equal to 50%. An appeal to naturalism is question-begging in this context. Nor should it be argued that a bridge is impossible on other grounds, since we can conceive of options (a special faculty of rational intuition, a constitutive relation between moral beliefs and properties, or divine revelation).[[36]](#footnote-36) The best bet is Ockham’s Razor: we already have an evolutionary explanation for moral beliefs, so no need to posit anything further, and the fewer entities, the simpler the theory, the more probable.

But it’s far from clear that this argument is legitimate in this context. First, we are already assuming there are non-natural moral properties. So, we should grant that they cannot be ruled out on explanatory grounds. And if such grounds don’t rule out such things, explanatory arguments against non-natural entities don’t generally work. It’s difficult to see why there would be a special explanatory reason that would rule out bridges in particular. Actually, on explanatory grounds, we might have reason to posit bridges. Perhaps cognitivism, which we are granting, is best explained by positing a connection between beliefs and the corresponding non-natural moral properties, a connection in virtue of which beliefs can be about them. Such a connection need not track the truth, of course, so this leaves open the epistemological question. But once we have some connection, it isn’t clear that explanatory considerations disfavor truth tracking. Perhaps granting cognitivism was unwise. If debunkers wish to drop it, fine. But then it’s no longer a question about moral epistemology but about the nature of mental content and language. Here I am only concerned with epistemology. Assuming cognitivism is necessary to isolate this concern. So, it seems, there’s no basis on which even experts can rule out a bridge in the current context. Inscrutability is the closest they can get—*a fortiori* for laypersons. Given inscrutability, EDAs and the BFC are inert, hence incapable of full defeat. But, as with disagreement, it is not inscrutable that there’s at least some chance that EDAs and the BFC are correct. As such, partial defeat remains feasible.

**Debunking Reconstrued**

I’ve argued in this chapter that the debunking challenge is best understood in terms of higher-order defeat. I also outlined a theory of such defeat, which allowed us to systematically eliminate subpar responses to the debunking challenge, ultimately revealing a small window through which to escape.

The result is that a significant range of ordinary first-order moral beliefs—of laypersons and experts alike—are safe from full defeat, even on the assumption of a maximally robust moral realist package. I have conceded that many such beliefs are nevertheless partially defeated. Moreover, I conceded this “happily.” The reason for this attitude is that it seems to promote the virtues of epistemic humility and open-mindedness without relinquishing our moral positions altogether. Insofar as these intellectual virtues are conducive to moral virtue, debunking may have the ironic result of moral improvement, especially in our approach to moral disputes.

On a final note, suppose we grant that our moral beliefs are fully defeated given realism. Still, it doesn’t obviously follow that our moral beliefs are defeated—if we abandon realism. Perhaps there’s a third way: keep the moral beliefs but abandon realism.[[37]](#footnote-37) Realism might still happen to be true, and our moral beliefs will be safe from defeat. In this way too, candidate debunkers can be inert at the object level, depending on what other views one wishes to retain alongside ordinary moral judgments. If debunkers are directed against realist metaethics rather than normative beliefs, they may circumvent the objection just suggested as well as my contentions in this chapter. My defense of optimistic moral epistemology from debunking thus has the potential advantage of accommodating other roles for debunkers.

**References**

Alexander, D.J. (2013). “The Problem of Respecting Higher-Order Doubt.” *Philosophers’ Imprint* 13(18).

Barnett, B.C. (2016). *Higher-Order Evidence: Its Nature and Epistemic Significance*. Ph.D. dissertation, University of Rochester. *ProQuest*.

Benacerraf, P. (1973). “Mathematical Truth.” *Journal of Philosophy* 70(19): 661.

Bergmann, M. (2005). “Defeaters and Higher-Level Requirements.” *The Philosophical Quarterly* 55(220): 419-436.

--- (2006). *Justification without Awareness: A Defense of Epistemic Externalism*. Oxford: Clarendon Press.

Campbell, R. (2015). “Moral Epistemology.” *Stanford Encyclopedia of Philosophy*.

Carter, J.A. (2018). “Meta-Epistemic Defeat.” *Synthese* 195(7): 2877-2896.

Christensen, D. (2007). “Epistemology of Disagreement: The Good News.” *Philosophical Review* 116(2): 187–217.

--- (2009). “Disagreement as Evidence.” *Philosophy Compass* 4(5): 756–67.

--- (2010). "Higher-Order Evidence." *Philosophy and Phenomenological Research* 81(1): 185–215.

Clarke-Doane, J. (2017). “What Is the Benacerraf Problem? In *Truth, Objects, Infinity: New Perspectives on the Philosophy of Paul Benacerraf*, ed. F. Patuat. Verlag: Springer, 17-43.

Conee, E., and R. Feldman. (2001). “Internalism Defended.” In *Epistemology: Internalism and Externalism*, ed. H. Kornblith. Oxford: Blackwell, 231-260.

--- (2008). “Evidence.” In *Epistemology: New Essays*, ed. Q. Smith. New York: Oxford UP: 83-104.

Copp, D. (2008). “Darwinian Skepticism about Moral Realism.” *Philosophical Issues* 18: 186-206.

Elga, A. (2010). "How to Disagree about How to Disagree." In *Disagreement*, eds. R. Feldman and T.A. Warfield. New York: Oxford UP, 175–186.

Enoch, D. (2010). “The Epistemological Challenge to Metanormative Realism: How Best to Understand it, and How to Cope with It.” *Philosophical studies* 148(3): 413-438.

--- (2011). *Taking Morality Seriously: A Defense of Robust Realism.* Oxford: Oxford UP.

Feldman, R. (1988). “Having Evidence.” In *Philosophical Analysis*, ed. D. Austin. Dordrecht: Kluwer, 83-104.

---(2005). "Respecting the Evidence." *Philosophical Perspectives* 19(1): 95–119.

--- (2006). "Epistemological Puzzles about Disagreement." In *Epistemology Futures*, ed. S. Hetherington. New York: Oxford UP, 216–236.

--- (2007). "Reasonable Religious Disagreements." In *Philosophers Without Gods: Meditations on Atheism and the Secular Life*, ed. L.M. Antony. New York: Oxford UP, 194–214.

--- (2009). "Evidentialism, Higher-Order Evidence, and Disagreement." *Episteme* 6: 294–312.

Field, H. (1989). *Realism, Mathematics, and Modality*. New York: Blackwell.

Fitelson, B. (2012). "Evidence of Evidence Is Not (Necessarily) Evidence." *Analysis* 72(1): 85–88.

FitzPatrick, W. (2015). “Debunking Evolutionary Debunking of Ethical Realism.” *Philosophical Studies* 172(4): 883-904.

Goldman, A.I. (2001). “Experts: Which Ones Should You Trust?” *Philosophy and Phenomenological Research* 63(1): 85-110.

Gutting, G. (1982). *Religious Belief and Religious Skepticism*. Notre Dame, IN: University of Notre Dame Press.

Huemer, M. (2016). “A Liberal Realist Answer to Debunking Skeptics: The Empirical Case for Realism.” *Philosophical Studies* 173(7): 1983-2010.

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

Kelly, T. 2005. "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology*, vol. 1, eds. J. Hawthorne and T.S. Gendler. New York: Oxford UP, 167–196.

--- 2010. "Peer Disagreement and Higher-Order Evidence." In *Disagreement*, eds. R. Feldman and T.A. Warfield. New York: Oxford UP, 111–174.

King, N.L. (2012). “Disagreement: What’s the Problem? or A Good Peer Is Hard to Find.” *Philosophy and Phenomenological Research* 85(2): 249-272.

Klenk, M. (2017a). “Old Wine in New Bottles: Evolutionary Debunking Arguments and the Benacerraf-Field Challenge.” *Ethical Theory and Moral Practice* 20(4): 781-795.

--- (2017b). “Third Factor Explanations and Disagreement in Metaethics.” *Synthese*: 1-20.

--- (2019a). “Objectivist Conditions for Defeat and Evolutionary Debunking Arguments.” *Ratio* <<https://doi.org/10.1111/rati.12230>>.

--- (2019b). “Hanno, Sauer, *Debunking Arguments in Ethics*.” *Utilitas* 8: 1-5.

Machuca, D. (2013). *Disagreement and Skepticism*. New York: Routledge.

Matheson, J. (2009). "Conciliatory Views of Disagreement and Higher-Order Evidence." *Episteme: A Journal of Social Epistemology* 6(3): 269–79.

Mellor, D.H. 2005. *Probability: A Philosophical Introduction*. New York: Routledge.

Plantinga, A. 1993. *Warrant and Proper Function*. New York: Oxford UP.

Rini, R.A. (2016). “Debunking Debunking: A Regress Challenge for Psychological Threats to Moral Judgment.” *Philosophical Studies* 173(3): 675-607.

Sauer, H. (2018). *Debunking Arguments in Ethics*. New York: Cambridge UP.

Shafer-Landau, R. (2003). *Moral Realism: A Defence*. New York: Oxford UP.

--- (2012). “Evolutionary Debunking, Moral Realism and Moral Knowledge.” *Journal of Ethics and Social Philosophy* 7(1): 1-37.

Sliwa, P., and S. Horowitz (2015). “Respecting *All* the Evidence.” *Philosophical Studies* 172: 2835-2858.

Street, S. (2006). “A Darwinian Dilemma for Realist Theories of Value.” *Philosophical Studies* 127(1): 109-166.

Sturgeon, S. (2014). “Pollock on Defeasible Reasons” *Philosophical Studies* 169: 105-118.

Tersman, F. (2017). “Debunking and Disagreement.” *Nous* 51(4): 754-774.

Vavoka, K. (2015). “Evolutionary Debunking of Moral Realism.” *Philosophy Compass* 10(2): 104-116.

Wedgwood, R. (2010). “The Moral Evil Demons.” In *Disagreement*, eds. R. Feldman and T.A. Warfield. New York: Oxford UP, 216-246.

1. I am grateful to Michael Klenk, Margaret Greta Turnbull, and Eric Sampson for helpful feedback on an earlier draft. [↑](#footnote-ref-1)
2. For an exception to the causal claim, see Oddie (2009). [↑](#footnote-ref-2)
3. Enoch (2011) and Shafer-Landau (2003, 2012) defend this sort of package. [↑](#footnote-ref-3)
4. A helpful survey is given by Campbell (2016). [↑](#footnote-ref-4)
5. Influential proponents include Joyce (2006) and Street (2006). [↑](#footnote-ref-5)
6. The BFC originates in the work of Benacerraf (1973) and Field (1989). Klenk (2017a) mounts a strong argument that EDAs depend on the BFC. [↑](#footnote-ref-6)
7. E.g., Wedgwood (2010). [↑](#footnote-ref-7)
8. The notion of epistemic peerhood derives from Gutting (1982). Cf. Turnbull and Sampson’s contribution in this volume for a weaker conception. [↑](#footnote-ref-8)
9. Sometimes called the “equal weight view” (Kelly 2010), though there are weaker versions that only require decreased confidence in the face of peer disagreement. [↑](#footnote-ref-9)
10. For defenses of conciliationism, see Christensen (2007, 2009, 2010); Elga (2010); Feldman (2006, 2007, 2009); and Matheson (2009). For non-conciliatory views, see Kelly (2005, 2010) and Wedgwood (2010). See also Turnbull and Sampson’s contribution to this volume, which develops an account of non-conciliationism in terms of rational “level-splitting” beliefs. [↑](#footnote-ref-10)
11. See Machuca (2013) on the prospective skeptical implications of disagreement. [↑](#footnote-ref-11)
12. E.g., Copp (2008); FitzPatrick (2015); Huemer (2016); Klenk (2017a, 2017b); and Tersman (2017). [↑](#footnote-ref-12)
13. For existing epistemic responses, see Clarke-Doane (2017), Vavova (2015), and the non-conciliationists referenced above. [↑](#footnote-ref-13)
14. For a sampling of this growing body of literature, see Christensen (2009, 2010); Feldman (2005, 2006, 2007, 2009); Kelly (2005, 2010); Matheson (2009); and Sliwa and Horowitz (2015). [↑](#footnote-ref-14)
15. Barnett (2016) presents a fuller account. [↑](#footnote-ref-15)
16. Fitelson (2012) and Kelly (2005, 2010) raise objections to filtration to which Barnett (2016), Feldman (2009), and Matheson (2009) respond. [↑](#footnote-ref-16)
17. See the contributions by Lee, Robson, and Sinclair in this volume for higher-order defenses of testimony. [↑](#footnote-ref-17)
18. For discussion of this problem, see Conee and Feldman (2001). [↑](#footnote-ref-18)
19. In p.c. to Feldman (2009). [↑](#footnote-ref-19)
20. See Sturgeon (2014) and Klenk (2019a) for the related debate over subjective vs. objective defeat. [↑](#footnote-ref-20)
21. Silwa and Horowitz (2015) defend a version of this. [↑](#footnote-ref-21)
22. Thus, candidate debunkers as construed here are not subject to the self-refutation raised by Rini (2016). [↑](#footnote-ref-22)
23. Feldman (2006, 2007, 2009); Matheson (2009) [↑](#footnote-ref-23)
24. Christensen (2007), Elga (2010), Goldman (2001), and Kelly (2010) propose independence conditions. [↑](#footnote-ref-24)
25. Turnbull and Sampson argue that certain level-splitting beliefs can be rational—e.g., “P, but the evidence as construed by my peer doesn’t support P” (my wording). I can accommodate this with latching. If the evidence as construed by my peer doesn’t support P yet I’m justified in maintaining my stance, I must have reason to expect an evidential difference between us. This can only happen given the weaker notion of peerhood, which doesn’t require sharing the same evidence. As such, what I know about my peer’s evidence doesn’t quite map onto mine, hence doesn’t latch, therefore doesn’t undercut. I can still account for my peer’s evidence without latching via filtration, but then there’s dampening. Either way, the level-splitting belief remains rational. However, once faced with a peer in my stronger sense, this no longer applies. [↑](#footnote-ref-25)
26. See Feldman (1988) for discussion. [↑](#footnote-ref-26)
27. See also Conee and Feldman (2008) for this distinction. [↑](#footnote-ref-27)
28. This is part of the reason why, as Feldman (2009: 309-310) puts it, evidential relations are “timeless and eternal and necessary.” For accounts of such relations, see Conee and Feldman (2008) and Barnett (2016). [↑](#footnote-ref-28)
29. Bergmann (2006) provides a survey and worthwhile attempt of his own. [↑](#footnote-ref-29)
30. The usual responses to skepticism cannot bypass this point. Contextualism *concedes* skepticism in contexts in which skeptical questions are posed. Standard anti-skeptical theories (e.g., Mooreanism and explanationism) contend that there exist positive external measures on our evidence (even if we cannot say much about them). However, ordinary folk won’t know these arguments, and often admit they don’t have knowledge of external measures—yet retain a strong intuition they do know the first-order claims. This natural response seems rational. The only way to accommodate it is to accept that awareness of ignorance of external measures is inert. Hence, any adequate response to skepticism must supplement this point, not replace it. [↑](#footnote-ref-30)
31. Alexander (2013), Bergmann (2005), and Plantinga (1993) make the related distinction between suspended judgment (corresponding to neutral support) and the absence of a doxastic attitude (corresponding to inscrutable support). [↑](#footnote-ref-31)
32. Mellor (2005) [↑](#footnote-ref-32)
33. Given this isomorphism, and the implausibility of escaping defeat in the drug case by endorsing “epistemic optionalism” (the denial of the Uniqueness Thesis), we should say the same about peer disagreement. [↑](#footnote-ref-33)
34. In King’s (2012) phrase, “A good peer is hard to find.” [↑](#footnote-ref-34)
35. Consider Street (2006) and Shafer-Landau’s (2012) interpretation. [↑](#footnote-ref-35)
36. Klenk (2018a) [↑](#footnote-ref-36)
37. See Sauer’s (2018) argument and Klenk’s (2019b) response, along with Carter (2018). [↑](#footnote-ref-37)