1. Introduction

Debunking arguments—also known as etiological arguments, genealogical arguments, access problems, isolation objections, and reliability challenges—arise in a wide variety of domains, including causation, chance, color, consciousness, epistemic reasons, free will, grounding, laws of nature, logic, mathematics, modality, morality, natural kinds, ordinary objects, religion, and time. What unifies the arguments is the transition from a premise about what does or doesn’t explain why we have certain mental states to a negative assessment of their epistemic status. My aim in this paper is not to assess the debunking arguments, but rather to take a bird’s eye view on the debates arising in these different domains, to show what they have in common.

2. The Varieties of Debunking Arguments

Let’s begin by illustrating how debunking arguments arise in different domains, beginning with morality. One common strategy for debunking moral beliefs and intuitions is to point to their evolutionary roots. We have the moral attitudes that we do, the debunker might say, because those attitudes motivated our ancestors to perform actions that tended to enhance reproductive success. Moral facts have no role to play in such explanations; it is not as if we need to suppose that we really are obligated to feed our children in order to explain why it’s adaptive to think that we are. Accordingly, it could only be a coincidence if evolutionary forces led us to moral beliefs that line up with the moral facts. Since we have no rational grounds for believing that we got lucky, we shouldn’t think that we did, and so we should suspend all of our moral beliefs.

Debunking arguments also arise in the philosophy of mathematics. Suppose that, as Platonists think, mathematical beliefs are about mind-independent abstract objects (numbers, sets, etc.). Since such objects would be causally inert and have no way of influencing our beliefs, there would evidently be no conceivable explanation of our accuracy about them. Upon recognizing this, the Platonist must either abandon her mathematical beliefs or else reject the Platonist conception of mathematics that generates
the problem in the first place. It is easy to see how this sort of argument will generalize to other abstract domains, for instance logic and modality.

Beliefs about the colors of objects are meant to be debunked by the observation that colors themselves (if there are any) have no role to play in explaining our color experiences. Religious beliefs are meant to be debunked by explanations from evolutionary psychology, for instance that belief in God is a natural byproduct of an adaptive tendency to prefer agential explanations of events whose causes have not been observed. Experiences of time “flowing”, invoked in support of A-theory, have been held to be debunked by explanations from cognitive science, as have intuitions about grounding, the essences of natural kinds, and free will. Not even our perceptual beliefs about ordinary objects are safe from debunking: after all, given our hard wiring, atoms arranged treewise are bound to cause us to have an experience as of a tree, regardless of whether atoms so arranged in fact do compose a single object.

Debunking arguments can also take the form of global challenges to the reliability of all belief-forming methods. Given that our belief-forming methods are the products of natural selection, the idea goes, we ought to think it is highly unlikely that any of those methods reliably produce accurate beliefs. Why? Because the evolutionary explanation for why some method yields the beliefs it does will have little if anything to do with the accuracy of those beliefs and everything to do with which sorts of beliefs enhance reproductive success.

3. The Structure of Debunking Arguments

Debunking arguments in any given domain D can be divided into two categories. First, there are skeptical debunking arguments, which aim to establish that a given range of beliefs or other propositional attitudes have some negative epistemic status. Schematically:

(SD1) Our D-attitudes and the D-facts do not stand in explanatory relation E
(SD2) If so, then our D-attitudes have negative epistemic status S
(SD3) So, our D-attitudes have negative epistemic status S
Second, there are targeted debunking arguments, which aim to establish only that a certain sort of theorist is committed to the attitudes’ having a negative epistemic status:

(TD1) Theorist T is rationally committed to believing that her D-attitudes and the D-facts do not stand in explanatory relation E
(TD2) If so, then T is rationally committed to taking her D-attitudes to have negative epistemic status S
(TD3) So, T is rationally committed to taking her D-attitudes to have negative epistemic status S

Instances of either schema will specify a domain D (e.g. morality), the explanatory relation E that is said to be absent (e.g., that D-facts don’t explain D-attitudes), the type of attitude in question (e.g. belief or intuition), and the status S that the attitude is said to have (e.g. being unjustified) A targeted debunking argument must additionally specify the range of theorists whose attitudes are being said to be debunked (e.g. realists, Platonists, or naturalists).

Each debunking argument has an explanatory premise (SD1 or TD1) and an epistemic premise (SD2 or TD2). Approaches to defending the explanatory premise, and in particular the claim that D-facts don’t explain D-attitudes, may be categorized as either negative or positive.

The negative approach to defending the explanatory premise is to argue that D-facts (as conceived by theorist T) aren’t the sorts of things that could enter into an explanation of D-attitudes. For instance, debunkers may insist that the D-facts (so conceived) can’t explain our D-attitudes insofar as they purport to be facts about abstract and causally inert objects or properties.

The positive approach to defending the explanatory premise is to provide a “sparse” explanation of our D-attitudes, one that makes no reference to D-facts. Debunkers may, for instance, point to an evolutionary explanation of our D-attitudes, entirely in terms of the adaptive advantage of having such attitudes, or they may point to more proximate explanations, in terms of facts about worldly objects or features that are not among the D-facts. The idea would then be that, given the adequacy of the sparse explanation, alternative
explanations of D-attitudes that do make reference to D-facts would be unparsimonious or otherwise objectionable. viii

The idea behind the epistemic premise is that the indicated explanatory revelation operates as a defeater. Defenses of this premise can be separated into two categories, according to how the explanatory revelations are meant to undermine the D-attitudes. Indirect defenses take the revelations to undermine D-attitudes indirectly, by way of revealing them to have some other, more fundamental deficiency, for instance that they are unsafe or insensitive or at best coincidentally accurate. Direct defenses take the revelations to undermine D-attitudes directly: it is directly in virtue of the recognition or concession that your attitudes aren’t appropriately connected to associated facts that those attitudes are undermined. ix

4. Explanationist Responses

Let’s turn now to strategies for resisting debunking arguments. These divide into two categories. First, there are explanationist strategies, which challenge the explanatory premise by insisting that the specified explanatory relation does obtain between D-facts and D-attitudes. Second, there are minimalist strategies, which challenge the epistemic premise, maintaining that the recognized absence of the specified explanatory relation is not sufficient to undermine our D-attitudes.

Because the exact explanatory relation at issue may vary from one formulation to the next—depending on what the debunker fills in for E in the argument schema—a strategy will count as explanationist only relative to a formulation. For concreteness, we’ll focus on responses to the following partial completion of the schematic premise SD1:

(SD1*) Our D-attitudes neither explain nor are explained by the D-facts

There will then be two sorts of explanationist strategies. First, there are alethic explanationist strategies, on which the D-facts do explain our D-attitudes. Second, there are doxastic explanationist strategies, on which it is the D-attitudes that explain the D-facts. We’ll consider the possibility of a more attenuated explanatory connection at the end of this section.

Alethic strategies come in many different varieties. There are reductionist strategies, on which one identifies the D-facts with some range of facts that
straightforwardly figure in the explanation of D-attitudes. For instance, one might insist that colors just are the very spectral reflectance properties that the debunker takes to explain our color experiences, or that the facts about rightness or wrongness just are facts about what does and doesn’t promote the flourishing of our species.\footnote{x} Other alethic strategies involve working (unreduced) D-facts into the causal, constitutive, semantic, theological, or evolutionary explanations of D-beliefs.\footnote{xi}

Doxastic strategies are available to certain sorts of anti-realists in the relevant domains. One might hold, for instance, that the moral facts are as they are because we have the moral attitudes that we do, or that things have the colors that they do in virtue of our being disposed to have certain sorts of color experiences in their presence.\footnote{xii}

Even if D-facts neither explain nor are explained by D-attitudes, there could still be a more attenuated explanatory connection between them. According to third-factor strategies, there is some further fact that explains both the D-attitudes and the D-facts. The fact that feeding one’s children promotes their survival, for instance, may be cited as the third factor that figures both in the (evolutionary) explanation of why we believe it’s good to feed them and in the (moral) explanation of why it is good to feed them. Whether this qualifies as an explanationist or a minimalist strategy depends on the exact formulation of the argument. If the explanatory premise affirms the absence of a third-factor explanation, then this is an explanationist strategy; if it doesn’t (as in SD1*), it’s a minimalist strategy.

5. Minimalist Responses

Minimalist responses are those that challenge the epistemic premise of the debunking arguments. As we saw above (§3), defenses of this premise may be either direct or indirect. Since the different styles of defense invite different strategies for resistance, we’ll consider them separately.

5.1 Indirect Defenses of the Epistemic Premise

According to the indirect debunker, explanatory revelations debunk by revealing D-attitudes to be in one way or another “precarious”. Such a debunker might contend that, if indeed the factors responsible for our D-beliefs were in no way influenced by the D-facts, then our D-attitudes are unsafe: we could easily have ended up with inaccurate D-
attitudes. Or that our D-attitudes are insensitive: we would have had the same D-attitudes even had the D-facts been different. Or that the D-attitudes are unreliable, or that it could only be a coincidence or an accident or a stroke of luck if we ended up with accurate D-attitudes. And it’s the recognition that the attitudes, even if accurate, are precarious in some such way that ultimately does the defeating.xiii

Minimalist responses to indirect defenses divide into two categories: stable and unstable. Unstable minimalists grant that the D-attitudes are precarious, but deny that this concession undermines their D-attitudes.xiv After all, they will say, one can sometimes have good evidence that something unlikely or massively coincidental has occurred. I can check my roster and find that every one of the thirty students enrolled in my class next term is named ‘Sam’. Realizing that this could only be a massive coincidence doesn’t undermine my belief that they all have the same name, since I have good evidence (the roster) that the coincidence did occur.

Suppose, then, that the debunker insists that we would have to have gotten massively lucky to end up with accurate moral beliefs. The unstable moral minimalist will grant the point. But she will insist that we did get massively lucky and that there is good evidence that we did. After all, we can check whether those actions that we believe to be wrong really are wrong, for instance by consulting our intuitions about whether the actions are right or wrong, or by checking whether the actions have the natural features that (intuitively) make right actions right and wrong actions wrong. And when we do, we find that, as luck would have it, those actions that we believe to be right are indeed right, and that those we believe to be wrong are wrong.

Stable minimalists grant the absence of the relevant explanatory connection but, deny that our D-attitudes are precarious, pointing to some features of D-beliefs or D-facts that secures their accuracy—no luck or coincidence required.xv As an illustration, take the question of whether your belief that 1+1=2 is safe. Are you mistaken in any nearby worlds about whether 1+1=2? No, says the stable minimalist. After all, it’s not just true but necessarily true that 1+1=2, so (a fortiori) it’s true in all nearby worlds. Moreover, you believe that 1+1=2 in all the nearby worlds; perhaps the evolution of mathematical cognition could have taken a different course but, given the evident adaptive value of our mathematical beliefs, this isn’t something that could easily have happened. Putting the
pieces together, you have a true belief about whether 1+1=2 in all nearby worlds. In other words, you couldn’t easily have been mistaken; the belief is safe.

One might naturally complain that these minimalist lines of reasoning are question-begging or circular. After all, the reasoning involves relying on the very D-attitudes that the debunker means to be calling into question (e.g., moral intuitions or mathematical beliefs), in an attempt to vindicate those very attitudes. Minimalists will likely respond that there is nothing illicit about relying on these attitudes. They will likely insist that some amount of epistemic circularity and question-begging is inescapable, on pain of global skepticism. And they will likely insist that their D-attitudes have not yet been debunked simply by the revelation that they bear no appropriate explanatory connection to the D-facts. By the indirect debunker’s own lights, explanatory revelations defeat only by way of convincing one of the precariousness of the attitudes, in which case there should be nothing wrong with relying on the not-yet-impugned attitudes to check whether the explanatory disconnect renders the attitudes precarious.xvi

[[In a longer version of the paper, I also consider “plenitudinous” strategies.]]

5.2. Direct Defenses of the Epistemic Premise

According to direct debunkers, the epistemic premise is underwritten by a more general explanatory constraint. Here are some candidate formulations of such a constraint:

(C₁) If S believes that the fact that p does not explain her belief that p, then S is thereby rationally committed to withholding belief in p

(C₂) If S believes that the fact that p neither explains nor is explained by her belief that p, then S is thereby rationally committed to withholding belief in p

The conclusion that D-beliefs have the relevant negative epistemic status would then be an immediate consequence of the concession that those beliefs lack the appropriate explanatory connection to the D-facts.xvii

Direct defenses circumvent the minimalist responses considered in the previous subsection (with the possible exception of plenitudinists). As we saw, the minimalist’s vindication of her D-attitudes relies on those very attitudes, which (she will insist) is unobjectionable so long as the attitudes have not yet been defeated. But if the direct debunker is right, then the minimalists’ D-attitudes have already been defeated by their
explanatory concession, and it is neither here nor there that they can reason from the impugned attitudes to the conclusion that the attitudes aren’t precarious or that they are luckily accurate. The minimalist must therefore reject the direct debunker’s explanatory constraint.

One common objection is that such constraints are *self-refuting*. To see the problem, take C1. Does C1 itself—that is, the putative fact expressed by the sentence labeled ‘C1’—explain one’s belief in C1? The debunker will likely concede that it doesn’t; after all, C1 looks to be precisely the sort of abstract, normative fact that according to debunkers doesn’t or can’t explain our beliefs. But then, given C1, the debunker’s belief in C1 is undermined, thereby undermining her reasons for accepting the premises of her own debunking argument.xviii

What this reasoning shows is that a certain sort of skeptical debunker cannot make use of C1 or C2—specifically, a debunker whose reasons for rejecting explanationist accounts of D-attitudes also preclude explanationist accounts of beliefs about C1 and C2. The reasoning does not show that principles like C1 and C2 are false, nor does it show that *no one* can rationally accept them. Anti-realists who embrace doxastic explanations of D-facts, for instance, are well-positioned to embrace C2 and launch *targeted* debunking arguments against realism that are driven by C2, so long as they extend their anti-realism to epistemic facts like C2.xix

Another objection to explanatory constraints is that they overgeneralize and end up undermining our beliefs about the future and other inductive beliefs. After all, the fact that the sun will rise tomorrow surely is no part of the explanation of my belief that it will rise tomorrow, on pain of backwards causation. Nor does my believing it to be so make it so. By C1 and C2, then, I ought to withhold belief about whether the sun will rise tomorrow—which is plausibly regarded as a reductio of those constraints.xx

Direct debunkers and other friends of explanatory constraints will likely concede that C1 and C2 are too strong, and attempt to weaken the explanatory constraint to make room for justified beliefs about the future.xxi Here is one illustration:

(C3) If S believes that the fact that p neither explains nor is explained by her belief that p and that there is no further fact that explains both the fact that p and the belief that p, then S is thereby rationally committed to withholding belief in p
C₃ arguably does not prohibit beliefs about future sunrises, since the laws of nature plausibly figure in the explanation of both beliefs and facts about future sunrises. A debunker who weakens the constraint in some such way will of course have to make sure that she has not weakened it so much that the D-attitudes she is trying to debunk end up satisfying the constraint. (C₃, for instance, leaves the door wide open for a third-factor response.)

References
Baras, Dan (2017b), ‘Our Reliability is in Principle Explainable’, *Episteme* 14: 197-211.
Baron, Sam (2017), ‘Feel the Flow’, *Synthese* 194: 609-630.


Copp, David (forthcoming), ‘How to Avoid Begging the Question Against Evolutionary Debunking Arguments’, *Ratio*.


Isserow, Jessica (forthcoming), ‘Evolutionary Hypotheses and Moral Skepticism’, *Erkenntnis*.

Jaquet, François (forthcoming), ‘A Debunking Argument Against Speciesism’, *Synthese*.


Klenk, Michael (forthcoming), ‘Third Factor Explanations and Disagreement in Metaethics’, *Synthese*.


*Philosophia Mathematica* 27: 88-104.

Kovacs, David M. (forthcoming), ‘How to be an Uncompromising Revisionary 
Ontologist’, *Synthese*.

Kyriacou, Christos (2016), ‘Are Evolutionary Debunking Arguments Self-Debunking’, 
*Philosophia* 44: 1351-1366.

Ladyman, James, and Ross, Don (2007), *Every Thing Must Go* (Oxford: Oxford 
University Press).


*Philosophy Compass* 6: 552-563.

Leng, Mary (forthcoming), ‘Debunking, Supervenience, and Hume’s Principle’, 
*Canadian Journal of Philosophy*.


Liggins, David (2010), ‘Epistemological Objections to Platonism’, *Philosophy of 
Compass* 5: 67-77.

Lillehammer, Hallvard (2003), ‘Debunking Morality: Evolutionary Naturalism and Moral 
Error Theory’, *Biology and Philosophy* 18: 567-581.

Locke, Dustin (2014), ‘Darwinian Normative Skepticism’, in Michael Bergmann and 
220-236.

Lott, Micah (2018), ‘Must Realists Be Skeptics? An Aristotelian Reply to a Darwinian 
Dilemma’, *Philosophical Studies* 175: 71-96.


Maddy, Penelope (1980), ‘Perception and Mathematical Intuition’, *Philosophical Review* 
89: 163-196.


Mizrahi, Vivian (2006), ‘Color Objectivism and Color Pluralism’, *Dialectica* 60: 283


Rose, David and Jonathan Schaffer (forthcoming), ‘Folk Mereology is Teleological’, *Noûs*.


Setiya, Kieran (2012), Knowing Right from Wrong (Oxford: Oxford University Press).


Thomasson, Amie L. (forthcoming), ‘How Can We Come to Know Metaphysical Modal Truths?’, *Synthese*.


---


ii On color, see, e.g., Goldman (1992: chs. 2.3 and 3.1), Chalmers (2006: §6), Pautz (2011: §3), Schaffer (2016: §2), and Cutter (2018: §3). In the color literature, debunking arguments are far less common than arguments from variation in color experience, but interestingly the range of responses to variation arguments (dispositionalism, reductionism, externalism, pluralism) double as responses to debunking arguments.


ix On framing the debate in terms of defeaters, see Plantinga (1993: ch. 12), Merricks (2003), and Thurow (2013).


For a direct defense, see Lutz (2017).


See Fumerton (1995: 50-51) for more on self-refuting arguments.

On the future, see White (2010: 582-583), Dogramaci (2017), and Fairchild and Hawthorne (2018: 52).