Etiological Debunking Beyond Belief[*]

Joshua Schechter

Abstract. Learning information about the etiology of one's beliefs can reduce the justification a thinker has for those beliefs. Learning information about the etiology of one's desires, emotions, or concepts can similarly have a debunking effect. In this chapter, I develop a unified account of etiological debunking that applies across these different kinds of cases. According to this account, etiological debunking arguments work by providing reason to think that there is no satisfying explanation of how it is that some part of our mental life is fitting. The role that etiological information plays is to help to rule out (or render less plausible) potential explanations, given our background views about the world. This account suggests several potential morals for epistemology, the philosophy of mind, and metaethics.

1. INTRODUCTION

Sometimes when we learn information about the etiology of some of our beliefs, this can have a debunking effect.[1] Here are a few examples:

- I learn a broadly evolutionary story about the etiology of my moral (or normative) beliefs according to which the relevant evolutionary factors were not tied to the truth of the relevant moral (or normative) claims. This would seem to debunk those beliefs or their status as objective.[2]
- I learn that had I been brought up in a different neighborhood, I would have been similarly intelligent and well-informed, but I likely would have had very different political (or moral or religious) beliefs.[3]
- I learn that some of my philosophical beliefs are due to the influence of my training at university X rather than at (the equally good) university Y.[4] (Perhaps I couldn't decide where to go and flipped a fair coin to choose between the two graduate programs.)

These examples are all at least somewhat controversial. But there are also more quotidian (and less controversial) cases of the etiological debunking of beliefs. Examples include cases in which

[*] This is the penultimate draft of a paper forthcoming in *Oxford Studies in Metaethics* 19.
[1] Etiological information can also have a vindicating (or more complex) effect.
[2] See Ruse and Wilson (1986), Joyce (2006), and Street (2006).
[3] See Cohen (2000, chapter 1), Mill (1991 [1859], 229-30), and Sher (2001).
[4] See Cohen (2000, chapter 1).

a thinker learns information that suggests that one of their beliefs was formed on the basis of wishful thinking or is due to some kind of bias.

Etiological debunking does not only arise for beliefs. Learning information about the etiology of other parts of our mental life can also have a debunking effect. Here are a few cases:

- I learn that my general philosophical methodology is due to the influence of my training at university X rather than at (the equally good) university Y.
- I learn that some of my general methods of inquiry are due to the unconscious influence of third-rate detective novels.
- I learn a broadly evolutionary story about the etiology of some of my basic motivational states according to which the relevant evolutionary factors were not tied to the aptness of those motivational states.
- I learn that my current desire to dance like a chicken is due to my having been hypnotized by an entertainer at a party.[5]
- I learn that my preference for Brand X personal care product is due to Brand X's clever marketing campaign.[6]
- I learn that my preference to listen to jazz rather than pop is due to my teenage self's desire to "be different".
- I learn that my intense anger at a friend's behavior is due to its resonance with something from my early childhood.
- I learn that my fear of lizards is the result of a false memory that researchers cleverly implanted in me (or is the result of having been surreptitiously fed a fear-of-lizards pill, or is due to my distant ancestors having been terrorized by now-extinct giant lizards, or…).
- I learn that my moral (or political or religious) concepts were deliberately spread by members of some group to maintain their privileged social status.[7]
- I learn that a particular scientific concept came to be accepted because of the beauty of the illustrations in the text introducing it.

To make these cases precise would require filling in various details, which I don't have the space to do here. But even gesturing at the cases, it should be clear that learning one of these pieces of information would be arresting. It would (at least apparently) have a debunking effect. The thought that my preferences about personal care products are due to the manipulations of

[5] See Elga (ms b).
[6] See Elga (ms b). Elga's specific example was his realizing that his preference for a certain brand of shaving cream was likely due to the fact that a can was included in his undergraduate university's "welcome package". As it happens, I attended the same university as Elga. Upon reading his paper, I had the uncomfortable realization that the very same thing had happened to me.
[7] This example is due to Matti Eklund.

clever marketeers is a somewhat disturbing one. It seems to cast doubt on my preferences. The thought that my intense anger at my friend's behavior is due to its resonance with something in my early childhood seems to cast doubt on my anger. And so on for each of the examples.

It is worth being explicit about how I'm using the term "debunk" here. According to my usage in this chapter, information debunks a subject's mental state when the subject's possession of the information renders the mental state less rational for the subject to have. According to this definition, then, a belief can be debunked even if it is true. The examples listed above are examples in which learning information about the etiology of some mental state (plausibly) makes it less rational for the subject to retain that mental state.

For my purposes here, one doesn't need to accept all of these examples of etiological debunking.[8] What is important is not that each of the cases is fully persuasive, but that there is a wide range of cases of (apparent) etiological debunking involving a wide variety of mental states. Examples of etiological debunking do not solely concern beliefs and other cognitive states. They also concern methods of inquiry, conative states, affective states, and even parts of one's conceptual apparatus.[9]

That is not to say that every kind of mental state can be debunked by etiological information. If I find out that my headache has a problematic etiology, that does nothing to debunk it.[10] Headaches don't count as rational or irrational, and so cannot be debunked. More generally, non-debunkable mental states plausibly include sensations (e.g., itches and pains), basic appetites (e.g., hunger and thirst), perceptual experiences (both veridical and

[8] In conversation, I've found the greatest resistance has been to the case of musical preferences.
[9] See Milligram (1997, chapter 2) for an argument that desires can be undermined by etiological information. See Joyce (2013, pp. 366-70) for defense of the etiological debunking of emotions and of "likings" and "dislikings".
[10] See Joyce (2013, p. 367).

hallucinatory),[11,12] as well as mere propositional entertainings. But very many kinds of mental states can (at least apparently) be debunked.[13]

Indeed, it's easy to generate examples of (apparent) etiology debunking for:

- Cognitive states including beliefs, disbeliefs, credences, hunches, best guesses, assumptions, and expectations
- Conative states including desires, preferences, wants, wishes, hopes, plans, and intentions
- Affective states including fear, anger, joy, sadness, guilt, blame, resentment, and happiness
- Deductive rules, inductive practices, deliberative and decision-making procedures, and other parts of our theoretical and practical methodologies
- Concepts and other parts of our conceptual apparatus

This list focuses on what's internal to an individual's mind, but the same phenomenon extends to external behavior, including methods of gathering evidence and other patterns of action. It also extends beyond the individual to the social, including practices of group inquiry and group deliberation as well as other social practices. So the phenomenon of etiological information yielding (at least apparent) debunking is very broad.

In this chapter, I'd like to discuss the following question: How does etiological debunking work? That is, how is it that learning information about the etiology of a mental state can render that mental state less rational? And what is the normative principle (or principles) involved in such etiological debunking?

[11] But see Siegel (2017) for a view on which perceptual experiences are rationally assessable, and so presumably can be debunked.

[12] There is a subtlety here. If I learn that one of my perceptual experiences has a problematic etiology – for instance, it was generated by my accidental ingestion of a hallucinogen – this does not make my having the experience any less rational. It does, however, make it less rational for me to believe the content of the experience. The etiological information thus has downstream effects. One could define a second sense of "debunk" according to which a mental state is debunked by etiological information when the information makes it less rational to deploy the state in practical and theoretical reasoning in the canonical way. On this alternative definition, additional mental states would count as debunkable. Thanks to Suzy Killmister for discussion.

[13] A natural hypothesis is that the mental states that can be debunked by etiological information are all and only the mental states that are rationally assessable. This raises the questions: (i) Which mental states are rationally assessable? (ii) Why are those states the ones that are rationally assessable? These are interesting and important questions, but not ones I can do justice to here.

It would be strange if the etiological debunking of beliefs worked in a different way than the etiological debunking of desires, which worked in a still different way than the etiological debunking of emotions. The cases of apparent etiological debunking all seem to be similar. There seems to be a unified phenomenon here. So it would be theoretically more elegant – and intellectually more satisfying – if there were a general account of etiological debunking that applied across the different kinds of debunkable mental states.

The aim of this chapter is to develop a unified account of etiological debunking that applies across these different kinds of mental states. According to the account developed here, etiological debunking arguments work by providing reason to think that there is no plausible explanation of how it is that some part of our mental life is fitting (roughly: correct) that is compatible with our background views of the world. The role that etiological information plays is to help to rule out (or render less plausible) potential explanations given our background views. This generates rational pressure on the combination of (i) the relevant parts of our mental life, (ii) our acceptance of the etiological information, and (iii) the relevant background views.

This chapter will proceed as follows. In the next section, I'll make a few brief clarifications to get the target phenomenon in better focus. Then in section 3, I will discuss the etiological debunking of beliefs. This will yield a model of etiological debunking. In section 4, I'll extend this model to other kinds of mental states. This general model requires that debunkable mental states have fittingness conditions. In section 5, I'll discuss the fittingness conditions of several different kinds of mental states. Section 6 is devoted to responding to potential objections to my account. Finally, in section 7, I'll briefly present several potential morals that can be drawn from my discussion.

2. CLARIFICATIONS

Before I start to develop the account of etiological debunking, let me first make a few necessary clarifications.

On any reasonable view, the etiology of our beliefs is relevant to their doxastic justification. Recall the distinction between propositional and doxastic justification. A thinker is propositionally justified in believing some proposition if (more or less) the thinker has good grounds for believing it. A thinker's belief is doxastically justified if (more or less) the belief is based on good grounds. The two can come apart. A thinker may have good grounds for believing some proposition, and may in fact believe it, but may believe it on bad grounds (for example, wishful thinking). Such a thinker is propositionally justified in believing the relevant proposition. But the thinker's belief does not count as doxastically justified.

On almost every view, for a belief to be based on good epistemic grounds, it must be formed and maintained in the right way. So the etiology of a belief is directly relevant to its doxastic justification. But that's not my topic here. Instead, my focus is on *information about etiology*, not etiology itself. The issue I'm concerned with is whether learning information about the etiology of a mental state can have a normative impact on the mental state. For instance, can learning information about the etiology of a belief lessen the epistemic justification that one has for holding the belief? Can learning information about the etiology of a desire make it less rational for one to have the desire? And so on. This is most naturally understood to be an issue concerning propositional justification (and its analogues for other kinds of mental states), not doxastic justification (and its analogues for other kinds of mental states).

Notice that information about the etiology of a mental state can be misleading. For instance, perhaps my desire to dance like a chicken is not in fact due to my having been

hypnotized by an entertainer. If I nonetheless acquire information that strongly suggests that being hypnotized is what caused my desire, this information would seem to have a debunking effect. Indeed, this information would seem to have the very same debunking effect that it would have had were it not misleading.

Some theorists provide stories that present an imaginary genealogy of beliefs and concepts.[14] My focus here is on information about the actual causal history of our mental states, not on "just so" or "state of nature" stories. Such stories raise other issues. I'm also not here interested in "merely psychological" effects of etiological information (for instance, making salient the paucity of one's evidence or the existence of alternative views). My interest is in the normative significance of etiological information, not in its ability to remind us of other issues.

In what follows, I will focus on rationality (broadly understood) rather than pragmatic, aesthetic, moral, or political normativity. I will try to stay as neutral as possible on the nature of beliefs, desires, emotions, and concepts. Finally, I will assume that there genuinely is a phenomenon of etiological debunking.[15]

3. BELIEF

Let me start by considering the etiological debunking of belief. It will turn out to be useful to examine an evolutionary debunking argument. In particular, I'll focus on the evolutionary debunking argument for normative beliefs due to Sharon Street.[16] I start with Street's version of

[14] For recent examples, see Craig (1991) on knowledge and Williams (2002) on truthfulness.
[15] See Elga (ms a) and White (2010) for concerns about whether there is a genuine phenomenon of etiological debunking that does not collapse into some other issue.
[16] See Street (2006; 2008).

the argument because it is among the most sophisticated versions in the literature, and because discussing it will elicit some useful morals.

Here is a sketch of Street's argument:

> *Darwinian Dilemma.* Our normative capacities are the products of evolution by natural selection. Indeed, the forces of natural selection have had a tremendous impact on the contents of our normative beliefs. According to normative realism, the fundamental normative truths are mind independent. However, there are good scientific reasons to think that our ancestors were not selected for believing mind-independent normative truths (or for having the motivational tendencies that would lead to believing such truths). There are very many different possible collections of normative beliefs. It would therefore be an astonishing coincidence if our normative beliefs matched the independent normative truth. Thus, normative realism leads to the conclusion that "we are in all likelihood hopeless at recognizing the independent normative truth."[17] This is an intolerable skeptical conclusion. Therefore, normative realism is false.

This is not a direct quote, but it is a close rendering of what I take to be the central line of thought in Street's paper. For simplicity, it helps to understand "normative realism" as referring to "robust" non-naturalist versions of normative realism.[18]

Notice that the target of the argument is not the truth of our normative beliefs, but rather normative realism.[19] On many accounts of etiological debunking, it is our first-order beliefs about some domain that lose justification. What is particularly interesting about Street's version of the argument is that what ultimately get debunked are not our first-order normative beliefs but our view about their status (i.e., normative realism).

There are many interesting questions that can be raised about Street's argument. For instance, does it have force only against non-naturalistic versions of normative realism, or does it also have force against naturalistic versions? Do non-realist views, such as (perhaps) normative expressivism, end up facing a similar problem?[20] And so on.

---

[17] Street (2011, p. 15).
[18] See, for example, Enoch (2011), Moore (1912), and Shafer-Landau (2003).
[19] This is an important contrast with the etiological debunking argument in Joyce (2006).
[20] See Street (2011).

My interest here, however, is not to delve into the details of Street's precise formulation of the argument, or to consider exactly how far the argument extends. Rather, my interest is in a more basic issue. How does the argument work? In particular, what is the central normative principle featuring in the argument?

A natural way to try to understand the argument is as an "irrelevant influences" argument.[21] According to this line of thought, the crux of the argument should be understood as follows:

> *Irrelevant Influences*. Assuming normative realism, we have good reason to believe that our normative beliefs were formed in a way that reflects the significant impact of irrelevant factors – causal factors disconnected from the truth of the beliefs in question. (The irrelevant influences at issue are ones that were important for the evolution of our motivational tendencies – for instance, conduciveness to survival and reproduction.) According to a general epistemological principle, if under certain background assumptions we have good reason to believe that our beliefs about a domain were formed in a way that reflects the significant impact of irrelevant factors, this defeats the initial justification we may have had for the combination of our beliefs about the domain with the background assumptions. Thus, there is a tension in the normative realist's package of views. This puts pressure on the combination of normative realism, our normative beliefs, and our general background views about the world. The best way to resolve this tension is to reject normative realism.

There are several problems facing this explication of Street's argument. A first issue is that there are numerous counterexamples to the "general epistemological principle". Most if not all our beliefs were formed in ways that involve the significant impact of irrelevant factors. For instance, the fact that I was earlier a bit peckish caused me to look in my fridge and notice that the light bulb in my fridge had burned out. But the fact that I was peckish is not connected to facts about the state of the bulb. Similarly, there are many things I believe about Providence, Rhode Island because I was interested in philosophy, decided to pursue it as a career, and ended

[21] See Vavova (2016).

up being hired by an institution in that city. But my interest in philosophy is not connected to the truth of the claims about Providence. And so forth.

Perhaps we can avoid these counterexamples by suitably restricting the "general epistemological principle". I don't see how to do so.[22] And even if there is a suitable way to restrict the principle, there are other concerns with the proposed explication of the argument. One concern is that the explication does not obviously capture the intuitive force of there being an astonishing coincidence. The claim that it is an astonishing coincidence that our normative beliefs are true plays an important role in Street's argument, or the intuitive argument underlying it. But it does not play an obvious role in the irrelevant influence argument, at least as stated. This suggests that the irrelevant influence argument is missing something important.[23]

A different issue is that the "general epistemological principle", even if true, is not a good candidate for being a basic principle in epistemology. (This may be especially so if the principle is restricted in some way.) Why does having reason to think that our beliefs were impacted by the presence of irrelevant factors put pressure on the combination of those beliefs with our background views of the world? Something needs to be said here, presumably concerning how irrelevant factors tend not to yield true beliefs, and how this defeats justification for the relevant beliefs. So the argument depends on more basic epistemic principles. It would be good to have a sense of what those more basic principles are supposed to be. And it would be good to formulate

[22] See Schechter (2018; 2023) for discussion of possible approaches.

[23] One could try to fill out the irrelevant influence argument by claiming that good reason to think that one's beliefs about a domain were formed in a way that reflects the significant impact of irrelevant factors provides good reason to believe that it would be an astonishing coincidence if one's beliefs about the domain are true, and this is what puts pressure on the combination of our beliefs about the domain with the relevant background assumptions. This would be an improved version of the argument. But it would still face the other difficulties. And my sense is that proponents of irrelevant influence arguments intend the presence of irrelevant causal factors to be the relevant defeater, not the presence of an astonishing coincidence. Thanks to Jamie Dreier for suggesting this response.

the argument in terms of those more basic principles, and not in terms of a principle concerning irrelevant influences.

Even putting these issues aside, there is a bigger problem facing the irrelevant influences explication of the argument. Namely, it is unclear why an evolutionary story is needed to pose a problem for normative realism. Street grants than any plausible causal history of our normative beliefs would suffice for her argument – for example, cultural evolution. But it is not the *presence* of a causal history of our normative beliefs that generates the problem. Rather, it is the *absence* of an explanation of our reliability. Or, more accurately, the problem is raised by there being *good reason to think that there is no satisfying explanation* of our reliability. If we were to possess a satisfying explanation of the reliability of our normative beliefs that is compatible with our background views, the fact that there were irrelevant casual factors would be unproblematic.

What this suggests is that we should understand the argument in a different way. We should understand the argument not in terms of irrelevant causal factors but in terms of explanation. Simplifying slightly, what lessens the justification of our normative beliefs is that we have reason to think there is no satisfying explanation of how we get it right about the normative.

Here, then, is a more promising explication of the argument:

> *Reliability Challenge*. We seem to be reliable about the normative in the following sense: By and large, the contents of our normative beliefs are true and the contents of our normative disbeliefs are false, at least given sufficient reflection and discussion. This is a striking fact, one that "cries out" for explanation. In particular, explanation is needed of how it is that we have a cognitive mechanism (or mechanisms) for making normative judgments that is reliable. Given normative realism, there is no satisfying explanation of our reliability that can be provided. According to a general epistemological principle, it is a cost of a theory if it treats a striking phenomenon within the scope of the theory as accidental or otherwise inexplicable. Thus, there is a tension in the normative realist's overall package of views. This puts pressure on the combination of normative realism,

> the claim that we are reliable about normativity, and our general background views of the world. The best way to resolve this tension is to reject normative realism.[24]

Notice that the normative principle featuring in this argument – it is a cost of a theory if it treats a striking phenomenon as inexplicable – is a general epistemic principle not specifically tied to etiological debunking or to irrelevant influences. It is a more plausible candidate for being a basic normative principle governing belief.

I think this is a better explication of the argument. But it raises an important question: Where exactly does evolution fit in? My claim is that it doesn't.

Pointing to the evolution of our normative capacities is part of what makes the Darwinian dilemma so arresting. But this is a "merely psychological" effect rather than an epistemological matter. The evolutionary story makes salient a worry we should already have had – namely, that there is no explanation of how it is that we are reliable about the normative, at least assuming non-naturalist normative realism (and various other background claims).[25] One way to see this point is to note that the evolutionary story doesn't rule out any potential explanation of our reliability that we didn't antecedently have reason to disbelieve, at least given the acceptance of a broadly scientific picture of the world.

Generalizing to other etiological debunking arguments for beliefs, this suggests the following conjecture:

> *Initial Conjecture*. Etiological debunking arguments (concerning beliefs) work by vividly highlighting the fact that there is reason to think that there is no satisfying explanation of how it is that we have true beliefs about a topic that is compatible with our background views of the world. The role of etiological information is "merely psychological" rather than genuinely epistemologically significant. The causal story makes salient a worry we should already have had.

[24] This argument is analogous to the Benacerraf-Field challenge to mathematical Platonism. See the introduction and title essay in Field (1989). See Benacerraf (1973) for an important precursor. See Enoch (2011) for discussion of the analogous argument for the case of normativity and Schechter (2010) for the case of logic.
[25] The dialectical situation is more complicated when the argument is targeted against naturalistic views.

Indeed, this conjecture seems to fit many familiar examples of etiological debunking arguments for beliefs. When I learn that had I been brought up in a different neighborhood, I would have had different political (or moral or religious) beliefs, the specific etiology of these beliefs seems irrelevant. Instead, the crux of the issue seems to be that I have reason to think that there is no satisfying explanation of how it is that I got it right about these matters.

However, this conjecture is false. It is based on an impoverished budget of cases.

Consider the following counterexample: Suppose that I have many beliefs about the moon Europa – what its atmosphere is like, whether there might be life there, and so on – on the basis of talking to an eminent astrobiologist. I later learn that the astrobiologist is an inveterate prankster and enjoys misleading non-scientists.

In this case, there is etiological debunking of my beliefs about Europa. This is so even though there is a perfectly good explanation of how it is my beliefs about Europa *could* be reliable – the astrobiologist knows the facts about Europa and accurately communicated them to me. So in this case at least, the etiological information is doing more than highlighting a pre-existing worry.

Indeed, even in the Darwinian dilemma, there is epistemologically important etiological information – generated not by the evolutionary story but by a broadly scientific picture of the world. This picture rules out various potential explanations of our reliability – e.g., causal interactions with elementary normative particles or with abstract normative entities.

This suggests a different conjecture about etiological debunking:

> *Improved Conjecture*. Etiological debunking arguments (concerning beliefs) work by providing reason to think that there is no satisfying explanation of how it is that we have true beliefs about a topic that is compatible with our background views of the world. The epistemic role that etiological information plays is to help to rule out (or render less plausible) potential explanations, given our background views of the world. This

> generates rational pressure on the combination of (i) the beliefs, (ii) our acceptance of the etiological information, and (iii) the relevant background views.

I think that this conjecture is correct. But there is one complication that should be noted. Recall that my explication of Street's argument involves the claim that our reliability about normative matters is a striking fact in need of explanation. Presumably, it is also a striking fact that someone's moral, political, or religious beliefs are reliable, or that they are reliable in their beliefs about a given subject matter. But the fact that one has a true one-off belief about some topic does not seem nearly as striking. The complication is that there are etiological debunking arguments that apply to a single isolated belief. For instance, I might learn that one of my philosophical beliefs is due to the influence of my graduate institution. Or perhaps I believe a single claim about Europa based on conversing with an eminent astrobiologist, who I later find out is a prankster. Learning this information would still seem to have a debunking effect. What is going on in such a case?

There are several possible views one could take here. For instance, one might claim that having a true belief is always a striking fact that cries out for explanation. Alternatively, one might claim that etiological information plays an additional role in such cases – it can be striking that a single belief is true given its etiology. Or, one might claim that there is general rational pressure against simultaneously having a belief and believing there is no satisfying explanation of how it is that one has a true belief. (On this view, the normative principle in play doesn't depend on a more fundamental normative principle involving strikingness.) It is not clear to me which of these three views is best. But I do think that any of them could play a role in a plausible model of the etiological debunking of a single belief.

4. BEYOND BELIEF

Let's move from beliefs to other kinds of mental states. Does the model of the etiological debunking of beliefs extend to conative states, affective states, and concepts? I claim that the answer is yes. The model can be smoothly extended to other parts of our mental life.

Consider the etiological debunking of our normative beliefs. As Street points out, it is implausible that natural selection selected our ancestors to have certain normative beliefs. Rather, a better picture is that natural selection selected our ancestors to have certain motivational tendencies (desires, dispositions to act, and so forth).[26] And it seems that we could run an etiological debunking argument directly against these motivational states: Our basic motivational tendencies seem by and large apt for the situations we find ourselves in. This is a striking fact, in need of explanation. But it is difficult to see what the explanation of this fact is, given our background views. This provides rational pressure against the combination of the relevant motivational tendencies with the relevant background views.

I don't think that this argument ultimately works. I suspect that there is a good explanation of how it is that we by and large have apt motivational tendencies. But the point is that if we do in fact have good reason to think there is no satisfying explanation, this generates pressure to adjust either our motivational tendencies or our background views.

Now consider the case of acquiring information that suggests that my current desire to dance like a chicken is due to the machinations of a party entertainer. Again, we can extend the model of etiological debunking to this case. When I learn the relevant etiological information, this puts pressure on the claim that the desire is a fitting one for me to have – that it coheres with my other beliefs and desires, that it is authentically "due to me", that it is conducive to my well-

[26] See Street (2006, p. 118).

being, and so forth. It does so by ruling out or rendering less plausible potential explanations of how I came to have a fitting desire.

More generally, the examples at the start of this chapter suggest that there is a unified account of etiological debunking that applies across mental states. The examples all fit the kind of story I have been telling, where the debunking at base depends on there being reason to think that there isn't a certain kind of explanation – an explanation of how it is that we have *fitting* (roughly: correct) mental states.[27] For the case of belief, fittingness should be identified with truth. For the cases of conative and affective states (and concept-possession), fittingness should be identified with something else.

What this suggests is that we should endorse the following conjecture about the etiological debunking of mental states, which generalizes the "improved conjecture" above:

> *General Conjecture*. Etiological debunking arguments work by providing reason to think that there is no satisfying explanation of how it is that we have fitting mental states that is compatible with our background views of the world. The normative role that etiological information plays is to help to rule out (or render less plausible) potential explanations, given our background views of the world. This generates rational pressure on the combination of (i) the relevant parts of our mental life, (ii) our acceptance of the etiological information, and (iii) the relevant background views.

Two clarifications may be helpful here. First, we can distinguish between two different explanatory questions – "how is it that some specific mental states we have are fitting?" and "how is it that we came to have fitting mental states?". The latter question is the relevant one. That is where etiological information is most directly relevant. Second, for etiological information to debunk a subject's mental states, the subject need not be able to reason about whether the relevant mental states are fitting. The subject need not even possess the concept *fitting mental state*. Rather, what is important is that the etiological information rules out (or

[27] For discussion of the fittingness of emotions, see D'Arms and Jacobson (2000). For a survey of fittingness in general, see Howard (2018).

renders less plausible) the claim that the relevant mental state satisfies its fittingness condition, where this is understood de re and not de dicto. For example, it is plausible that a fear is fitting only if the target of the fear is genuinely dangerous. For etiological information to debunk a fear, it suffices that the information rule out (or render less plausible) potential explanations of how it is the subject came to be afraid of something that is genuinely dangerous.

Is the general conjecture true? I think that it provides a correct account of etiological debunking or is in the ballpark of a correct account. To fully evaluate the conjecture would require going through the various cases I began with, one by one, to see how well they fit the model I've been proposing. I don't have the space to do that here. Instead, I'll consider a more general issue. The account of etiological debunking I've been suggesting requires that there be fittingness conditions for any debunkable mental state. So for the view to be plausible, it's important that there be good candidates for such fittingness conditions. The next section will focus on that issue. In the section after that, I'll respond to a few objections to the account. Discussing these objections will enable me to flesh out the account in a bit more detail. After that, I'll conclude by eliciting a few potential morals from my discussion.

5. FITTINGNESS CONDITIONS

The account of etiological debunking I've been suggesting predicts that the mental states that can be debunked by etiological information are precisely the ones that have (non-trivial) fittingness conditions. The account is compatible with any account of the fittingness conditions of mental states. But since it is plausible that a wide range of mental states can be debunked by etiological information, for the account to be plausible, there need to be attractive candidates for the fittingness conditions of a wide range of mental states. In this section, I'll briefly present some

candidates. I do not wish to commit myself to the claim that these candidates are correct. Rather, they are only intended to provide a proof of concept to show that there are plausible ways to develop the account of etiological debunking.

As I'm using the term, "fitting" is a near synonym for "correct". The fittingness condition of a mental state is something like its correctness condition.[28] For instance, it is plausible that a belief is fitting just in case it's true.[29] Similarly, it is plausible that the acceptance of a method of reasoning, such as a rule of inference, is fitting just in case the method of reasoning is truth conducive.[30]

For a case of an emotion, consider fear. It is plausible that a fear of (say) bears is fitting only if bears are in fact dangerous. Or, more precisely, a fear that one will be harmed by bears is fitting only if there is a danger that one will be harmed by bears. A fear of being lonely is fitting only if there is a danger of being lonely. In general, a fear that p is fitting only if there is a danger that p – that is, there is a prospect that p and the outcome that p would be a bad one.

This is a simplification of the full story. Fears come in degrees – one may be mildly afraid all the way through completely terrified. The fittingness condition of a fear should take its degree into account. Presumably, to be fitting, the degree of the fear should correspond to – or at least be no greater than – the level of danger, which itself depends on the likelihood or easiness of the relevant scenario coming to pass and the degree to which that scenario would be bad.

[28] I use "fittingness" rather than "correctness" for two reasons. First, "correctness" misleadingly suggests that the relevant conditions only concern matters external to the thinker. Second, perceptual experiences can be correct or incorrect in the sense of being veridical or non-veridical. However, perceptual experiences plausibly lack fittingness conditions.

[29] There is an alternative proposal that might be preferred by knowledge-first epistemologists: A belief is fitting just in case it is an instance of knowledge.

[30] The case of credences is a bit trickier. A natural proposal is that the fittingness of credences comes in degrees, and a credence is fitting to the degree that it is accurate according to some strictly proper scoring rule.

There are plausible candidates for the fittingness conditions of other emotions, too. Here are a few: Anger is fitting only if the target of the anger is responsible for inappropriately harming something that one cares about. Sadness is fitting only if one has suffered a significant loss. Shame is fitting only if one has failed to live up to some ideal. Admiration is fitting only if the admired has a positive feature or has accomplished something good. Curiosity about a question is fitting only if one doesn't know (or, perhaps, doesn't have a justified belief in) the answer to the question. Perhaps it should also be required that the question not be insignificant. Surprise is fitting only if what one is surprised about is not to be expected given one's background views (or, perhaps, given what one's background views rationally should be). And so forth. Just as for the case of fear, presumably these proposals should be complicated to take into account the degree of the relevant emotion. But there does not seem to be any in principle difficulty in doing so.

Now let's consider the case of desire. There are several kinds of desires – including instrumental desires, realizer desires, and ultimate desires. An instrumental desire is a desire for something as a means of satisfying some other, more fundamental, desire. A realizer desire is a desire for something as a way of satisfying some other, more fundamental, desire. An ultimate desire is a desire for something "for its own sake", and not as a means or ways of satisfying some other desire.

There is a natural proposal to make for the fittingness conditions of instrumental and realizer desires: An instrumental desire is fitting only if satisfying the desire would in fact be a good means of satisfying the more fundamental desire that the instrumental desire aims at satisfying. A realizer desire is fitting only if satisfying the desire would in fact be a good way of

satisfying the more fundamental desire that the realizer desire aims at satisfying. (Perhaps it should also be added that the more fundamental desire is itself required to be fitting.)

The case of ultimate desires is more vexed. Some philosophers – with broadly Humean views of the nature of desire[31] – will claim that ultimate desires cannot be evaluated as rational or irrational, and so cannot be debunked. Presumably, on such a view, ultimate desires either lack fittingness conditions or have trivial fittingness conditions – that is, every ultimate desire one possesses counts as fitting. It is worth emphasizing that this view is compatible with the account of etiological debunking developed here. Given any view of the fittingness conditions that some kind of mental states possess, the account will specify whether and how that kind of mental state can be debunked by etiological information. If it turns out that certain mental states either lack fittingness conditions or have trivial fittingness conditions, the account will entail that those mental states cannot be debunked. So the account of etiological debunking is not committed to any particular view of the fittingness conditions of ultimate desires.

As it happens, though, I am sympathetic to the view that ultimate desires can count as rational or irrational. There are intuitively compelling cases where an ultimate desire strikes us as irrational (e.g., an ultimate desire to devote considerable time and effort to constructing the fourth largest ball of twine on the East Coast and to maintaining its status as fourth largest). More interestingly for present purposes, there are intuitively compelling cases where an ultimate desire can be debunked by etiological information. For instance, consider the case of someone who desires to dance like a chicken. In some versions of the case, the desire to dance like a chicken is an instrumental or realizer desire – the subject may have a desire to enjoy themself and desire to dance like a chicken as a means toward or a way of achieving that end. In a

[31] This kind of view goes back to Hume (1975 [1739-40], 2.3.3.6).

different version of the case, the desire is an ultimate desire – the subject may desire to dance like a chicken "just because". It's intuitive that even if the desire to dance like a chicken is an ultimate desire, if the subject learns information that strongly suggests that this desire was inculcated by a party entertainer, this information would debunk the desire. If that's right, this suggests that ultimate desires have nontrivial fittingness conditions. A plausible candidate for such a fittingness condition is that the ultimate desire be authentically due to (or ratified by) the subject.[32] Perhaps, too, for an ultimate desire to be fitting, satisfying the desire cannot strongly negatively contribute to the subject's wellbeing, have a large amount of disvalue, or something of the sort.[33]

What is the fittingness condition for possessing a concept? If a conceptual role semantics (or metasemantics) is correct for some concepts, possessing any of those concepts involves having certain beliefs or following certain inferential rules. This is most plausible for logical concepts – on a natural view, possessing the concept *if … then* involves inferring in accordance with Modus Ponens and also, perhaps, Conditional Proof. If a view of this sort is correct, presumably such a concept is fitting only if its constitutive beliefs are true and its constitutive rules are truth-conducive. Even if conceptual role semantics is not correct for a given concept, there are natural fittingness conditions for possessing it. For instance, to be fitting, presumably a concept should have a semantic value. Perhaps, too, the concept should be suitable for representing the world or for guiding inquiry (or for some other important cognitive practice).[34]

[32] Beliefs and emotions can also be debunked by the information that the relevant mental state was inculcated by (e.g.) an entertainer in a problematic way. This suggests that it may be part of the fittingness conditions of any debunkable mental state that the state is authentically due to (or ratified by) the subject. If so, it is a delicate matter how best to articulate this condition since it's important to avoid classifying mental states acquired via testimony or training as unfitting.

[33] On an alternative proposal, the fittingness conditions for desires are uniform across kinds of desires. For instance: A desire is fitting just in case what is desired is good (or, perhaps, is desirable).

[34] See Plunkett (2016).

There is plenty more that can be said about the fittingness conditions of various mental states. But the discussion here is hopefully enough to provide a sense that there are attractive candidates for the fittingness conditions of a wide range of mental states. In the next section, I'll turn to discussing and responding to several objections to the account of etiological debunking.

6. Objections and Replies

Objection 1: Epistemic normativity only applies to beliefs and other cognitive states. So there cannot be epistemic debunking of other kinds of mental states – including conative states, affective states, and the possession of concepts.

Reply: It is plausible that epistemic normativity (narrowly construed) only applies to cognitive states such as beliefs, credences, and the employment of methods of reasoning. But the examples at the start of the chapter suggest that non-cognitive mental states can nevertheless be debunked. That is, etiological information can render them less rational. We have a practice of assessing many kinds of mental states for rationality. For instance, we might judge that someone's fear of garter snakes is irrational, or that someone's instrumental desire to commute 50 miles by car rather than by foot is rational. When a non-cognitive mental state is debunked, it loses rationality in this sense. The relevant kind of normativity here is not narrowly epistemic. But it is an important kind of normativity, nonetheless.

Objection 2: When rational pressure is brought to bear against some of our beliefs, we can give them up or reduce our confidence in them. What can we do for desires, emotions, or concepts?

The worry is that if there is nothing we can do, there can't be any rational pressure brought to bear against them. So there can't be etiological debunking of non-cognitive states.

Reply: Like beliefs, desires and emotions can be given up or reduced in intensity. We can also "distance" ourselves from them. We can fail to deliberate with them or act on them. For concepts, we can stop deploying them in thought, employ replacement concepts, or be cautious about the products of inquiry involving them.

Objection 3: We don't have significant control over our desires, emotions, or concepts. Without such control, there can't be any rational pressure brought to bear against them. So there can't be etiological debunking of non-cognitive states.

Reply: It is true that we don't have significant control of our desires, emotions, or conceptual apparatus. But the same is true for our beliefs. I cannot simply decide to believe that there is an elephant in the corner or that there is no desk in front of me. So this is not a relevant disanalogy between beliefs and the other kinds of mental states that can apparently be debunked. If a lack of control is not a problem for belief, it is not a problem for the other kinds of mental states.

Objection 4: Some updates to one's beliefs count as rational revisions. Others do not. By contrast, revisions to one's desires, emotions, and conceptual apparatus do not count as rational or irrational. If a mental state is not apt for rational revision, there can't be any rational pressure brought to bear against it. So there can't be etiological debunking of non-cognitive states.[35]

[35] Thanks to Jamie Dreier for pressing me on this issue.

Reply: I'm not sure whether we should accept the claim that only rationally revisable states count as rational or irrational. That is connected to the question of what the main purpose of rational appraisal is. If the main purpose of rational appraisal is in first-personally guiding one's mental life, then it is plausible that only revisable states can count as rational or irrational. If, by contrast, the main purpose of rational appraisal is in third-personal evaluations of others, then it is not clear that only revisable states can count as rational or irrational.

In any event, we should not accept the claim that desires, emotions, and concepts are not rationally revisable. One can come to realize that one's desire is compulsive or arbitrary and rationally come to lose the desire. One can come to realize that something that one fears is not in fact dangerous and rationally come to lose the fear.

Objection 5: The etiological debunking of mental states other than belief is parasitic on the etiological debunking of belief. So there is no need for a general model of etiological debunking that applies to different kinds of mental states.

Reply: I can think of three ways to develop a view according to which the etiological debunking of mental states other than belief is connected to the etiological debunking of belief. According to the first approach, debunkable mental states are partly constituted by beliefs. Such a mental state is debunked just in case one of its constitutive beliefs is debunked.

The main problem with this view is that it is implausible that debunkable mental states always have constitutive beliefs. Consider the case of fear. It is plausible that fear does have an associated content – for instance, that there is a danger that such-and-such. But this content plays a different role in our cognitive economy than does the content of a corresponding belief. Unlike

an ordinary belief, the content of a fear is closely tied to motivation. Unlike an ordinary belief, it cannot easily be deployed in complex inferences. It is also more insulated from the impact of evidence than an ordinary belief is. Thus, even if fear involves a propositional attitude, it is a mistake to think that fear is partly constituted by a belief.

This naturally suggests a second approach. According to this approach, debunkable mental states have propositional contents. Such a mental state is debunked just in case a belief in the associated content is (or would be) debunked.[36]

This is a more plausible view. However, it faces difficulties, too. One issue is that there are non-debunkable mental states with propositional contents – for example, perceptual experiences and propositional entertainings. So it is not a sufficient condition on being debunkable that a mental state has a propositional content. There may also be debunkable mental states that lack propositional contents – for instance, moods. It's plausible that moods such as boredom, ennui, excitement, generalized anxiety, and the like can be debunked by etiological information. But it's also plausible that they don't have contents, or at least no very specific contents. So there is reason to think that it is not a necessary condition on being debunkable that a mental state has a propositional content.

That leaves a third approach. According to this approach, debunkable mental states have fittingness conditions. A mental state is debunked just in case a belief that the relevant fittingness condition is satisfied is (or would be) debunked. (Here, the claim that the fittingness condition is satisfied should be understood de re and not de dicto.)

[36] Due to the conditional fallacy, it would be better to avoid using a counterfactual here and instead talk about the debunking of the subject's propositional justification to believe the associated content. (And similarly for the third approach.)

Perhaps something like this view is correct – I have some sympathy for it.[37] Perhaps, then, there is a connection between the debunking of beliefs and the debunking of other mental states. But this does not show that the debunking of mental states other than beliefs is *parasitic* on the debunking of beliefs. Rather, a more natural picture is that the mental states are debunked in parallel to the debunking of the beliefs. Both the mental state and the belief lose rationality in the same way and for the same reason. There is no obvious reason to think that the beliefs come first.

I'm sure there are additional objections that ought to be considered but let me move on and conclude by considering some morals suggested by my discussion.

7. CONCLUSION

If the account of etiological debunking presented here is on the right track, it suggests several potential morals.

A first potential moral is that epistemic normativity is closely connected to a more general kind of rationality. Etiological debunking arguments concerning beliefs reduce the epistemic justification of the target beliefs (or other beliefs, such as a belief in realism about the relevant domain). If what I've said is on the right track, such arguments should be assimilated to a more general kind of etiological debunking argument that does not involve epistemic justification but a more general kind of rational normativity. This suggests that epistemic justification is closely tied to – and plausibly a species of – a more general kind of rationality. Epistemic justification is the kind of rationality appropriate to beliefs. But it is not the only kind

[37] But see Schechter (ms) for reasons to worry.

of rationality. There are other kinds of rationality appropriate to other kinds of mental states. And such kinds of rationality are closely related.

A second potential moral is that the problem posed by etiological debunking arguments does not reduce to the problem of peer disagreement. White (2010) and Mogensen (2015) have argued that the normative significance of etiological debunking arguments traces back to peer disagreement – in particular, to disagreement with actual and possible thinkers.[38] But if my discussion is on the right track, and there is a general normative principle governing debunking across a variety of mental states, this isn't correct. We don't have disagreements with other people's desires, feelings, concepts, etc. We may find them irrational, off-putting, or inappropriate. But we don't disagree with them in anything like the sense in which we disagree with other people's beliefs. A more plausible view is that peer disagreement and etiological debunking arguments are species of a common genus. For instance, perhaps learning of the disagreement of a peer can defeat one of my beliefs because it raises the need to explain why I'm right and my peer is wrong, rather than vice versa.

A third potential moral is that etiological debunking arguments can be used to shed light on the nature of various kinds of mental states. Investigating the effectiveness of debunking arguments may help us to determine the ways in which affective states, conative states, and concepts can be defective. If we can identify which kinds of debunking arguments have force and which do not, this can potentially help us to determine the fittingness conditions of various kinds of mental states. Thus, perhaps debunking arguments can be used as a philosophical tool to investigate questions on the philosophy of mind/epistemology border.

[38] See Vavova (2016) for arguments against the view that the force of etiological debunking traces back to peer disagreement.

A fourth potential moral is that etiological debunking arguments can be used to provide a challenge to broadly Humean views of desire. Since there are plausible examples of the etiological debunking of ultimate desires, it is plausible that ultimate desires can be assessed for rationality, contrary to Humean views.

All of this is a bit speculative. But these are interesting and important issues that deserve to be further explored.[39]

References

Benacerraf, Paul. (1973) "Mathematical Truth," *The Journal of Philosophy* 70: 661–79.

Cohen, G.A. (2000) *If You're an Egalitarian, How Come You're So Rich?* Cambridge, Mass.: Harvard University Press.

Craig, Edward. (1991) *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford: Clarendon Press.

D'Arms, Justin and Jacobson, Daniel. (2000) "The Moralistic Fallacy: On the 'Appropriateness' of Emotions," *Philosophy and Phenomenological Research* 61: 65–90.

Elga, Adam. (ms a) "Lucky to be Rational," presented at the 2008 Bellingham Summer Philosophy Conference. Available at https://www.princeton.edu/~adame/papers/bellingham-lucky.pdf

Elga, Adam. (ms b) "Suspiciously-Formed Preferences." Talk delivered in 2013.

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

Field, Hartry. (1989) *Realism, Mathematics, and Modality*. Oxford: Basil Blackwell.

Howard, Christopher. (2018) "Fittingness", *Philosophy Compass* 13: e12542.

Hume, David. (1975) [1739-40] *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge, 2nd ed. revised by P. H. Nidditch. Oxford: Clarendon Press.

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

Joyce, Richard. (2013) "Irrealism and the Genealogy of Morals," *Ratio* 26: 351–372.

[39] Earlier versions of this chapter were presented at a Workshop on Conceptual History at the Arché Centre at the University of St. Andrews, an APA Eastern Division Meeting (where my commentator was Roger White), a University of Connecticut Philosophy Department brown bag, the Vancouver Summer Philosophy Conference (where my commentators were Magdalena Balcerak Jackson and Ned Markosian), a graduate seminar at Virginia Tech, and the Madison Metaethics Workshop. I'd like to thank the participants at these events for their questions and comments. Thanks especially to Nomy Arpaly, Magdalena Balcerak Jackson, Donald Baxter, Selim Berker, Paul Bloomfield, David Christensen, Jamie Dreier, Adam Elga, Suzy Killmister, Michael Klenk, Stephan Krämer, Stephan Leuenberger, Ned Markosian, David Plunkett, Bernard Reginster, Geoff Sayre-McCord, Miriam Schoenfield, Lionel Shapiro, Katia Vavova, Roger White, Bruno Whittle, and Masahiro Yamada for helpful conversations.

Mill, John Stuart. (1991) [1859] "On Liberty," in *The Collected Works of John Stuart Mill*, edited by John M. Robson. Toronto: Toronto University Press.
Milligram, Elijah. (1997) *Practical Induction.* Cambridge, Mass.: Harvard University Press.
Mogensen, Andreas. (2015) "Contingency Anxiety and the Epistemology of Disagreement," *Pacific Philosophical Quarterly*.
Moore, G. E. (1912) *Ethics*, London: Williams and Norgate.
Plunkett, David. (2016) "Conceptual History, Conceptual Ethics, and the Aims of Inquiry: A Framework for Thinking about the Relevance of the History/Genealogy of Concepts to Normative Inquiry," *Ergo* 3: 27–64.
Ruse, Michael and Wilson, Edward O. (1986) "Moral Philosophy as Applied Science," *Philosophy* 61: 173–192.
Schechter, Joshua. (2010) "The Reliability Challenge and the Epistemology of Logic", *Philosophical Perspectives* 24, *Epistemology*, 437–64.
Schechter, Joshua. (2018) "Explanatory Challenges in Metaethics," in T. McPherson and D. Plunkett (eds.), *The Routledge Handbook of* Metaethics. Abingdon: Routledge, pp. 443–58.
Schechter, Joshua. (2023) "Five Kinds of Epistemic Arguments Against Robust Moral Realism," in P. Bloomfield and D. Copp (eds.), *The Oxford Handbook of Moral Realism*, Oxford.
Schechter, Joshua. (ms) "Rational Defeat Beyond Belief"
Shafer-Landau, Russ. (2003) *Moral Realism: A Defense.* Oxford: Oxford University Press.
Sher, Georges. (2001) "But I Could Be Wrong," *Social Philosophy and Policy* 18: 64–78.
Siegel, Susanna. (2017) *The Rationality of Perception*. Oxford: Oxford University Press.
Street, Sharon. (2006) "A Darwinian Dilemma for Realist Theories of Value" *Philosophical Studies* 127:109–166.
Street, Sharon. (2008) "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About," *Philosophical Issues* 18: 207–28.
Street, Sharon. (2011) "Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways," *Oxford Studies in Metaethics* 6: 1–32.
Vavova, Katia. (2016) "Irrelevant Influences," *Philosophy and Phenomenological Research*.
White, Roger. (2010) "You Just Believe That Because…," *Philosophical Perspectives* 24: 573-615.
Williams, Bernard. (2002) *Truth and Truthfulness: An Essay in Genealogy*. Princeton: Princeton University Press.