“Third-Factor Defenses of Moral Knowledge.”

Abstract:

I argue that “third-factor” defenses against evolutionary debunking arguments fail because they predict a lack of epistemic safety for the beliefs that they defend.

Third-Factor Responses to Evolutionary Debunking Arguments.

Evolutionary Debunking Arguments (EDAs) attack moral knowledge or realist moral knowledge by alleging that the forces that created human moral belief-forming dispositions were disconnected in epistemically relevant ways from whatever moral truths there may be (for some EDAs, only if those truths are construed in a realist way).

For example, Street’s (2006) much-discussed EDA against moral knowledge, realistically construed, has been reconstructed by Vavova (2014):

(1) Evolutionary forces select for creatures with characteristics that increase fitness.
(2) The true evaluative beliefs and the adaptive evaluative beliefs come apart.
(3) So, evolutionary forces are off-track: they do not track the evaluative truth. [1, 2]
(4) Evolutionary forces have influenced our evaluative beliefs.
(5) So, off-track forces have influenced our evaluative beliefs. [3, 4]
(6) So, we have good reason to think that our evaluative beliefs are mistaken. [5]
(7) If you have good reason to think that your belief is mistaken, then you cannot rationally maintain it.
(8) So, we cannot rationally maintain our evaluative beliefs. [6, 7]

Here premise (2) claims a disconnect between the true evaluative beliefs and the adaptive evaluative beliefs, such that, given that human evaluative beliefs are adaptive, we end up lacking reason to think that they are true.

Brosanan (2011) reconstructs Street’s EDA along with EDAs from Joyce (2006) and Ruse and Wilson (1986) as having a similar claim embedded:

(9) If our moral beliefs are products of a process that fails to track moral facts, then we do not possess moral knowledge.
(10) Our moral beliefs are products of evolution by natural selection.
(11) Evolution by natural selection is a process that fails to track moral facts.
(12) Our moral beliefs are products of a process that fails to track moral facts.
(13) Therefore, we do not possess moral knowledge.

Here premise (11) claims that evolution is a process that “fails to track moral facts.” The “failure to track” gets interpreted to mean either that basic evaluative attitudes formed by evolutionarily-influenced cognition could well be adaptive but not true; or else that as an explanation of those attitudes’ content, adaptiveness is a superior explanation to truth. Truth then gets jettisoned by parsimony:

[H]ow does the tracking account [the claim that our basic evaluative attitudes evolved because of a relation to the truth] explain the remarkable coincidence that so many of the truths it posits turn out to
be exactly the same judgements that forge adaptive links between circumstance and response—the very same judgements we would expect to see if our judgements had been selected on those grounds alone, regardless of their truth? The tracking account has no answer to this question that does not run right back into the parsimony and clarity problems just discussed. (Street 2006, p. 132)

A popular response to such moral-knowledge EDAs has been to argue that some third factor is probabilistically connected (whether by explanation, causation, partial constitution, or otherwise) with both (at least some important) moral truths T and with the evolution of a human disposition to believe T, and that therefore those two things—the moral truths T and human disposition to believe T—are correlated. This allows the defender of moral knowledge to grant an evolutionary influence on the relevant belief-forming dispositions but still argue that there is a correlation between human moral belief-dispositions and moral truths.

For example, Kevin Brosnan grants that the adaptiveness of human moral beliefs “screens off” those beliefs from the moral truth. That is, the probability of humans believing the moral propositions P that they do is no greater given both adaptiveness and truth than it is given adaptiveness alone. But Brosnan points out that on such a picture, human moral beliefs are still correlated with the truth, because some of the reasons why believing P is adaptive also raise the probability of the truth of P (the arrows represent a relation of probability raising):

\[
\text{We believe that cooperation is morally good} \quad \text{Cooperation is morally good}
\]

\[
\text{It helps our group for us to believe that cooperation is morally good}
\]

(Brosnan 2011)

Here, “It helps our group for us to believe that cooperation is morally good” is a third factor raising the probability both of humans believing that cooperation is morally good and of it’s being true that cooperation is morally good. The existence of such a third factor raises the probability that humans will believe the truth, even though human belief is screened-off from true belief. Many such third-factors have been suggested in response to moral-knowledge EDAs.

For example, here are some third-factors that the following authors argue suffice to salvage moral knowledge from EDAs like Street’s:

- Enoch (2010): ‘Our survival is good.’
- Wielenberg (2010): ‘If one has the capacity to form beliefs, then one has rights.’
- Skarsaune (2011): ‘Pleasure is usually good and pain is usually bad.’
- Brosnan (2011): ‘It helps out group to believe that cooperation is good.’
- Behrends (2013): ‘If we have reason for doing something or other, then we have reason to pursue our continued existence.’
- Sterelny & Fraser (2016): ‘Cooperation is good.’
• Lott (2017): ‘An animal’s characteristic ways of living and acting determine it’s good.’

Various responses from EDA-advocates have been offered, including the claim that some of these third-factor responses beg the question against the debunker. Third-factor responses like Enoch’s, for example, clearly appeal to an evaluative claim of the very sort that the EDA purports to call into question. But I doubt that question-begging is a problem here because it is the task of the debunker to formulate an argument based on plausible premises. It is unreasonable for the debunker to require her interlocutors to set aside all moral beliefs before engaging with the EDA.

But in any case, further discussion of the question-beggingness debate will take me too far afield. I will now argue that even granting that such third-factor responses establish a correlation with truth for our basic evaluative attitudes (or moral beliefs, etc.), they still fail to salvage moral knowledge. This is because they imply a lack of warrant for the beliefs they defend.

A Safe Connection to the Truth?

Many epistemic externalists defend the idea that epistemic safety is implied by knowledge. Even those who do not include safety as a constitutive factor in knowledge often suppose that it is entailed by their favored constitutive factors (for instance, Sosa 2017). Epistemic safety is essentially the idea that if S believes truly, then not easily would she believe falsely. While internalists about knowledge will generally not suppose that knowledge implies safety, they should be uncomfortable with total evidence against epistemic safety, since that would be extremely strong evidence against the relevant belief (most internalist accounts allow one’s evidence to play a large role in one’s knowledge). Additionally, most moderate externalist accounts should not allow total evidence against safety, since only radical externalisms posit that one’s evidence doesn’t matter at all for one’s knowledge. So I take total evidence against epistemic safety to be generally considered a defeater for knowledge. If “warrant” is that element that turns merely true belief into knowledge, then total evidence against safety is more specifically a defeater for warrant. In any case, I’ll restrict myself to arguing that third-factor predict a lack of safety, and take this to be a significant result even without rigorous argument that it entails a lack of knowledge.

Safety is commonly articulated via a possible-worlds heuristic, where for one’s belief to be “easily false” means false in nearby possible worlds. For some time it has been recognized that beliefs in nearby possible worlds can reflect upon safety even if they have different content from the belief under evaluation. This is because a safety condition that only looked at beliefs with the same content would have the implausible implication that belief in necessary truths cannot be unsafe. A more plausible safety condition goes something like:

SAFETY: S’s belief that p via method M is safe iff in nearby worlds where S believes some similar p* via some similar M*, P* is true.

If p is a substitution-instance of p*, then beliefs both with the same content and relevantly similar content reflect upon p’s safety.

1 Williamson (2000), for instance. Some of these (like Williamson) take an externalist view of evidence, and so could claim that evidence does matter for knowledge, but this externalist view of evidence is even more radical. I am assuming an internalist account of evidence.

2 Dunaway (2017).
Now consider our third-factor defenses. In order to even allow for epistemic safety, it will need to be the case that the third factor doesn’t raise the probability of relevantly similar *false* beliefs in epistemically relevant possible worlds. But, for example, in Brosnan’s case, the third factor:

“It helps our group for us to believe that cooperation is morally good”

*does* raise the probability (in epistemically relevant possible worlds) of humans forming dispositions to have relevantly similar but false beliefs. This is because it raises the probability of humans forming any and every belief-forming disposition that would be adaptive given that “It helps our group for us to believe that cooperation is morally good.”

And (given the particular hominin ancestral environment) this includes dispositions to believe plausibly false things like:

- “If I am powerful, then cooperation is good only when I arbitrarily desire it”
- “Cooperation exclusively within our group is good”
- “Cooperation with a small undiscovered group of freeriders is good”
- “Cooperation only when pleasurable is good” [given that it is also morally good to experience pleasure when doing something morally good]
- “Cooperation to an increasing degree as someone shares my genetics is good”

In general, if it contributes to adaptiveness to have a belief-disposition that cooperation is good, and if cooperation is good, then it will also contribute to adaptiveness to have a host of other belief-dispositions, many of them false. And so on for any third factor: if X raises the probability of the adaptiveness of some accurate moral-belief disposition, X will also raise the probability of the adaptiveness of a host of other moral belief-dispositions, many of them systematically inaccurate.

For example, take Lott’s (2017) third-factor:

“An animal’s characteristic ways of living and acting determine it’s good.”

To adapt Brosnan’s handy visual aid, Lott’s third-factor is supposed to work like this (the arrow again represent relations of probability-raising):

```
We believe in the goodness of our characteristic ways of living and acting

Our characteristic ways of living and acting are good

Our characteristic ways of living and acting determine our good
```

The left-hand arrow is supposed to be supported by the observation that it will generally be adaptive for cognizing beings to have normative attitudes that sanction the behaviors that are adaptive
for them – in this case, human “characteristic ways of living and acting.” But we need only look back at
the list of false moral belief-dispositions that Brosnan’s third factor gives rise to, and adapt them to
Lott’s account. One feature of human evolution is that many contradictory behaviors in the moral realm
have been adaptive in our evolutionary history. It has been adaptive for powerful individuals and groups
to oppress the weak, as well as for the powerful and the weak to cooperate. But this means that in
relevantly similar possible worlds, humans have the disposition to believe that oppression of the weak is
good, even though in those worlds our characteristic ways of living and acting are not much different
form the actual world. So if oppression of the weak is made wrong in the actual world (by our
characteristic ways of living and acting), it is also made wrong in those possible worlds (since our
characteristic ways of living and acting are not much different there). But then humans have a
disposition to believe falsely in those relevantly similar possible worlds. And so on for the other
(plausibly) false moral belief-dispositions in that list above.

The problem for the third-factor accounts is that adaptiveness ends up needing to play a key
role in all of them, and adaptiveness is plausibly too disconnected from moral truth to safely influence
true moral belief. It is indeed connected enough to underwrite some probability-raising relations, as in
Brosnan’s account. And it’s connected enough to moral truths to underwrite some explanations and
partial constitution relations, as in Lott’s account and others’ accounts. But those relations admit of
correlations that are far to weak to underwrite epistemic safety. Put another way, in the diagram of
Brosnan’s and of Lott’s account, the left-side arrow is far too weak to support epistemic safety. And we
should be able to see that this will be the case for these third-factor accounts generally. They give
evidence of a correlation with truth in the actual world, given what we think the moral truths are. But
that is different from evidence of safety. If the correlation is not modally robust, then this is
simultaneous evidence for truth and against safety.

I argued earlier that if our total evidence is against the safety of some belief, this should have
epistemically significant implications. Now, of course evidence from a third-factor account in the EDA
literature is not ipso facto enough to swing our total evidence toward a lack of safety. If we have
independent evidence that our moral beliefs (certain important ones anyway) are epistemically safe,
then my argument merely shows that third-factor explanations are a poor explanation of the safety of
our moral beliefs. But if they are a poor explanation of the safety of our moral beliefs, then they are a
poor explanation of our moral knowledge, and that was the whole point of their existence! Taken on
their own, these third-factor explanations actually predict a lack of epistemic safety for the beliefs they
explain. And so if someone is in need of some defense of their moral knowledge, perhaps in the face of
an EDA, a third-factor defense is inadequate to the task.

References.

Behrends, Jeff (2013). “Meta-Normative Realism, Evolution, and Our Reasons to Survive,” Pacific

Brosnan, Kevin (2011). “Do the evolutionary origins of our moral beliefs undermine moral knowledge?”
Biology and Philosophy 26: 51-64.


it, and how to hope with it,” Philosophical Studies, 148: 413-438.


