Luck: Evolutionary and Epistemic∗

Billy Dunaway
University of Oxford

Abstract
This paper advances two theses about evolutionary “debunking” arguments. The first is that, while such arguments are often put forward using the rhetoric of “luck”, proponents of these arguments have not distinguished between the kinds of luck that might lead to the formation of a true belief. Once we make the needed distinctions, the relevance of the kind of luck which can be derived from broadly evolutionary explanations to the epistemological conclusions the debunkers draw is suspect. The second thesis is that successful versions of these arguments might still be in the offing, but they will need to include specific premises about the kind of evolutionary mechanism that explains the beliefs that the debunker targets. Proponents of debunking arguments, if they continue to insist on using luck-based considerations, are then hostage to empirical fortune in a way not previously recognized.

1 Preliminaries
Recently a quite large literature has emerged on the topic of evolutionary “debunking” arguments in ethics.1 Broadly, these arguments aim to show that, owing to some general features of the evolutionary origins of certain kinds of belief, those beliefs are epistemically suspect. This is just a rough schema for an argument, and within the burgeoning literature there are a variety of ways of filling out the schema. (In particular, which kinds of belief are problematic, how evolutionary origins bear on formation of the relevant beliefs, and what “epistemically suspect” amounts to are all questions any respectable version will need to answer.) One tempting way to fill out the argument, which debunkers consistently find attractive, is via the notion of luck. It is a common theme that the

∗Special thanks to Matthew Benton, Julien Dutant, Daniel Fogal, John Hawthorne, Brian Hedden, Maria Lasonen-Aarnio, David Manley, Katia Vavova, Dani Rabinowitz, and an audience at University College, London for many valuable comments and/or discussion of this paper.

effects of evolutionary forces that shape our ethical beliefs show that the beliefs in question require a kind of luck to be true. And many epistemic notions—knowledge, defeat, and the like—plausibly require the absence of a certain kind of luck. The question this paper addresses is whether the luck-based motivations for worrying about the epistemic effects of evolution are apt.

Progress on this question is important since, in addition to its intrinsic interest, its answer may have important ramifications for questions of first-order ethics. To give just one example: Sharon Street (2008) has claimed that evolutionary debunking arguments motivate adoption of a “Constructivist” account of ethical truth. Roughly, the idea is that my belief that it is wrong for me to tell a lie is, given its alleged evolutionary origins, epistemically unproblematic only if the fact that my lie-telling is wrong is (as the Constructivist claims) grounded in the fact that I would believe on reflection that lie-telling is wrong. But what others would believe about ethics on reflection is a contingent matter—and as many Constructivists allow, some people will, even after engaging in the appropriate kind of reflection, hold some very odd (or worse) ethical beliefs. For instance there are others who will reflect on the wrongness of lie-telling and conclude that it is permissible for them to lie. According to the Constructivist view Street favors, it is not wrong for such people to lie. Thus if Street is correct, evolutionary debunking arguments in ethics motivate some surprising first-order normative views. It is important to see whether luck-based considerations can support these conclusions, or if we need to look elsewhere for the needed support.

It is admittedly very natural to speak in terms of luck when considering the question of whether evolutionary forces might have produced true ethical beliefs. Street, for instance, uses such language in the following passage:

[I]t turns out that we have been evolving towards affirming whatever evaluative content tends to promote reproductive success. We have thus been guided by the wrong sort of influence from the very outset of our evaluative history, and so, more likely than not, most of our evaluative judgements have nothing to do with the truth. Of course it’s possible that as a matter of sheer chance, some large portion of our evaluative judgements ended up true, due to a happy coincidence between the realist’s independent evaluative truths and the evaluative directions in which natural selection tended to push us, but this would require a fluke of luck that’s not only extremely unlikely, in view of the huge universe of logically possible evaluative judgements and truths, but also astoundingly convenient to the realist. Barring such a coincidence, the only conclusion remaining is that many or most of our evaluative judgements are off track.\(^3\)

\(^2\)The thought appears to be that, given the Constructivism metaphysic, our ethical beliefs are not at risk of error (since, schematically, my believing that \(p\) makes it the case that \(p\) is true, there is no risk that I form a false belief when I believe that \(p\)). This is a superficially attractive but underinvestigated claim. See Dunaway (MS) for some arguments that the principles needed to secure this advantage of Constructivism are not as plausible as initial appearances would suggest.

\(^3\)Street (2006, 122), emphasis added. See also Street (2008, 208-9).
And Richard Joyce, when picking an analogy, chooses a case where one is slipped a pill that causes belief in a certain truth—that Napoleon lost Waterloo. It is extremely natural, given the set-up, to view one’s Napoleon-beliefs as only true owing to an accident of luck. One could easily have been slipped a pill which caused one to believe that Napoleon won instead. Joyce draws the analogy in the following passage:

Suppose that there were a pill that makes you believe that Napoleon won Waterloo, and another one that makes you believe that he lost […] Now imagine that you are proceeding through life happily believing that Napoleon lost Waterloo (as, indeed, you are), and then you discover that at some point in your past someone slipped you a “Napoleon lost Waterloo” belief pill. It is not a matter of your learning of the existence such pills and having no way of knowing whether you have ever taken one; rather, we are imagining that you somehow discover beyond any shred of doubt that your belief is the product of such a pill. Should this undermine your faith in your belief that Napoleon lost Waterloo? Of course it should.4

If ethical beliefs are suspect for the same reason that pill-caused Napoleon beliefs are suspect, this is because they require an objectionable kind of luck.

This observation is distinct from the interpretive claim that the official versions of the debunking arguments of Street, Joyce, and others, are best read as using premises about the presence of “luck” in ethical belief. The points I have been making so far have no implications concerning what the official versions of these arguments are. Rather, I have only been indicating that some of the apparent force of evolutionary considerations is derived from the need for luck to arrive at a true belief. We can ask how much of this force is warranted.5

With this background in place, I will argue for two theses. The first is that the appeals to luck in evolutionary debunking arguments frequently ignore important distinctions between different kinds of luck that are broadly relevant to forming a true belief. Not all types of luck in processes leading up to formation of our beliefs are epistemically problematic—some are perfectly benign. Proponents of debunking arguments, in using the language of luck to motivate central premises,

4Joyce (2006, 179). Similar rhetoric is found in the presentation of evolutionary arguments in other domains in Bloom (2009, 129) and Wilkins and Griffiths (2013, 143).

5While I have not taken a stand on how to interpret the arguments offered by the debunkers, we can think of the relationship between the luck rhetoric, which I am interested in, and the official debunking arguments in the following way. When debunkers make claims about a certain process being “off-track” or relying on a “coincidence” (Street), or claim that a particular belief is “unjustified” (Joyce), we will face the question of how to understand these proprietary terms. One option, which is naturally suggested by what debunkers say, is to take the luck rhetoric at face-value and assume that, if a belief instantiates one of these properties, it is because it is subject to epistemically problematic luck. Or, we can ignore the informal connection with epistemic luck: but then, we will need to find another account of why beliefs that are off-track, unjustified, etc. are to be avoided epistemically speaking. Associations with luck will not do the job. I return to this second possibility briefly in the concluding section of this paper.
have systematically ignored these distinctions. Absent further premises, the evolution-induced presence of luck in ethical belief implies nothing about the epistemic status of these beliefs. It is irrelevant to the conclusion the debunkers wish to draw.

But the second thesis of this paper is that there may well be additional premises available to the debunker which do establish the needed connection between evolutionary luck and the epistemic status of ethical beliefs. In particular, if debunkers can provide additional evidence about the specific way in which evolutionary forces have shaped our ethical beliefs, they can rule out the possibility that the luck involved in such belief-formation is epistemically benign. I will outline the what the specific evolutionary processes would have to look like in order to yield this conclusion. While I will not attempt to settle whether these processes have in fact taken the required shape, one lesson from this second thesis is that the relevance of luck to evolutionary debunking is hostage to empirical fortune. Not every evolutionary explanation that proceeds broadly by a mechanism of natural selection will do the job. It is an open (and under-explored) question, to be settled by naturalistic sciences and not philosophical debate, whether luck is the right tool for casting doubt on our ethical beliefs.

A few preliminary methodological points are in order before proceeding. The primary focus of the following sections will be on whether the debunker’s focus on luck is warranted. I will understand this primarily to be a question about the relationship between the luck arising from the evolutionary origins of ethical belief and the status of those beliefs as knowledge. This is not entirely in keeping with the terms used in the debunking literature, where the focus is not on ethical knowledge but instead is on whether ethical beliefs are ‘off-track’, ‘distorted’, or are at best true owing to a ‘coincidence’. But plausibly the lessons for knowledge will carry over to these other notions. Just as there are kinds of luck which are incompatible with knowledge, and others that are perfectly benign, an epistemically significant notion of ‘off-trackness’ will tolerate some kinds of luck but not others. (And similarly for ‘distortion’, ‘coincidence’, etc.) As a related point, some will want to think of debunking arguments as showing that, once one knows of the evolutionary origins for one’s ethical beliefs, one acquires a ‘defeater’ for them. Again I will not focus on the notion of a defeater here, but the lessons below carry over: one way of acquiring a defeater is by learning that one’s belief, even if true, was formed in a way that prevents one from knowing. Thus proponents of defeat-based arguments might choose to focus on the evolution-induced need for luck to arrive at a true ethical belief, and the same distinctions between objectionable and benign luck will be relevant in the same ways.

2 Knowledge and risk of error

These preliminary remarks point to some important work for the notion of luck that is connected to what can be known. It is a commonplace claim in recent epistemology that certain kinds of freedom from luck, or the absence of a risk of error, are necessary for knowledge. One can think of this in the context of traditional cases from Gettier (1963): Jones knows that Smith has 10 coins in her pocket, falsely believes Smith will get the job, and deduces that the person who
will get the job has 10 coins in her pocket. This belief turns out to be true, since Jones herself has 10 coins in her pocket, and is the one who will get the job. The belief Jones forms as a result of the deduction is, while true, not knowledge; anti-luck epistemologists hold that this is because the truth of her belief is a matter of luck.

Not just any kind of luck in true belief is problematic in a way that makes the belief fail to be a candidate for knowledge—a point that will be crucial for disentangling luck that arises from evolutionary explanations of a belief and luck that is incompatible with knowledge in later sections. If I happen to turn my head just in time to see that John is smiling, I can know that John is smiling, even though I am in a sense lucky to believe this at all. The Gettier case and the head-turning case involve different kinds of luck, only one of which is relevant to knowledge. One might use a simple “safety” condition on knowledge to capture this distinction; this is Simple Safety:

**Simple Safety** If one knows \( p \) in a world \( w \), then for any world \( w^* \) close to \( w \) where one believes that \( p \) by a similar method, \( p \) is true in \( w^* \).

Although Simple Safety provides a natural diagnosis of why true belief in the above Gettier case isn’t knowledge (on the grounds that it could easily have been false that Jones has 10 coins in her pocket), it cannot deliver the same verdict in similar cases where the belief in question is a belief in a necessarily true proposition, since these propositions are guaranteed to be true in all close worlds. This problem is especially relevant in when evolutionary debunking of “basic” ethical beliefs is at issue, as the contents of these beliefs are plausible candidates for propositions that are necessarily true if true at all.

Here is a revision of Simple Safety that avoids these and other difficulties, which I will take for granted in subsequent discussion as the best extant attempt at capturing the connection between knowledge and the absence of risk or luck:

**Safety** If one knows \( p \) in a world \( w \), then for any world \( w^* \) close to \( w \) where in \( w^* \) one believes a proposition \( p^* \) similar to \( p \), and the token causal process which produces the belief in \( p^* \) in \( w^* \) is similar to the token causal process which produces the belief \( p \) in \( w \), the belief in \( p^* \) in \( w^* \) is true in \( w^* \).

According to Safety, true beliefs in necessary propositions can fail to be knowledge. This is because of the requirement that similar beliefs in close worlds be true. If I correctly guess that 154 is the sum of 81 and 73, I still won’t count as knowing

---

6See similar examples in Pritchard (2004, 133 ff.); see also Kahane (2011, 105-6) for a preliminary discussion of kinds of luck as they pertain to debunking arguments.

7Compare Sosa (1999, 142), Williamson (2000, 325), Pritchard (2004, 163), Hawthorne (2004, 55-6). One might wish to modify Safety to account for cases where one’s belief-forming process fails to produce any belief at all (perhaps owing to reference failure—see Manley (2007)). I will omit these complications from the main discussion.
that 154 is the sum: given that I was guessing, I could easily have guessed a
different number and believed (say) that 183 is the sum. Since the proposition
that 183 is the sum is (in this context) sufficiently similar to the proposition that 154 is
the sum, and the token causal processes producing each guess are quite similar,
my belief doesn’t satisfy SAFETY and hence isn’t knowledge.8

Moreover, and importantly for a discussion of debunking arguments, the
causal history of a belief can in principle show that the belief is lucky in a way
that is incompatible with knowledge. But not every instance of luck in a causal
history implies a failure of SAFETY; only similar beliefs in nearby worlds that are
formed by similar token processes need to be true. This makes the exact nature of
the causal history of a belief highly relevant to its epistemic status.

As a warm-up, consider two simple cases which the causal history of one’s
belief shows that luck of some kind or other is needed for arriving at a true
belief. The first is COIN IN THE HEAD, adapted from a case in White (2010, 599):

COIN IN THE HEAD: You are forming beliefs about your immediate
environment. Inside your head, the various neural pathways involved
in your belief-forming process all lead to a box. In the box is a coin that
gets flipped around. The outcome of the coin flipping is objectively
random, and as it flips your beliefs change in a systematic way: when
the coin lands ‘heads’, you form the beliefs that you are located on
Earth, and that you are a human with a body. These beliefs are, how-
ever, “hardwired” into your brain; all that is needed to produce them is
your specific cognitive architecture and the coin’s being heads side up.
When the coin lands ‘tails’, your brain is hardwired to form different
beliefs: you form the false beliefs that you are located on Mars, and
that you are merely a brain with quasi-perceptual experiences caused
by electrode stimulation. As the coin is flipped, you lose memory
of your previous beliefs and cannot notice any inconsistency between
your quasi-perceptual beliefs over time.

SAFETY provides a natural and plausible explanation for why your current
beliefs in COIN IN THE HEAD are not knowledge, even when the coin is on ‘heads’
and your beliefs are mostly true. Worlds where the coin is on ‘tails’ are very
close—if an objectively random coin flip had turned out differently, you would
have been in these worlds. In these worlds, one’s beliefs in similar propositions
about one’s immediate environment are false. And importantly, these false beliefs
are formed by very similar causal processes, as in both cases the process involves
a hardwired connection between coin flips and beliefs.

But by varying the causal history of your belief without eliminating the need
for luck in arriving at a true belief, we arrive at a very different verdict. Consider
a related case, which we can call GOD’S COIN:

8Terminological note: here and in what follows, I will speak of beliefs satisfying of failing to
satisfy SAFETY. By this, I mean that they do not satisfy the consequent of the conditional labeled
SAFETY, and hence on present assumptions do not constitute knowledge.
God’s COIN: God created you five minutes ago, and before doing so, God flipped a coin. If the coin landed ‘heads’, God would create you in the environment you are actually in, and moreover would give you a normal human visual system and cognitive architecture that produces the true beliefs that you are located on Earth and are a human with a body. If the coin landed ‘tails’, God would not have given you a perceptual system, but rather would have given you credible but misleading testimony, on the basis of which you would form the false beliefs that you are located on Mars and are merely a brain with quasi-perceptual experiences caused by electrode stimulation. The outcome of God’s coin flip is objectively random.

Suppose again the coin landed ‘heads’. Unlike COIN IN THE HEAD, it is quite plausible to say that one knows propositions about one’s condition and environment. But intuitively worlds where one has false beliefs about being an envatted brain on Mars are just as close in God’s COIN as they are in COIN IN THE HEAD: the outcome of the coin flips are objectively random in both cases, and hence the coin could easily have landed ‘tails’ in each. And in each ‘tails’ scenario, you have the same false beliefs. The difference is, these beliefs are plausibly produced by sufficiently dissimilar processes in God’s COIN. While the false beliefs are produced by a ‘tails’ outcome are the products of believing on the basis of testimony, the true beliefs in a ‘heads’ outcome are the products of a human visual system interacting with its environment. Since beliefs produced by similar perceptual processes are true in all nearby worlds, one’s true beliefs in the ‘heads’ outcome satisfy SAFETY.

Comparison of God’s COIN and COIN IN THE HEAD serves to illustrate the importance of similarity of belief-forming processes to SAFETY. But more generally, it points to the way in which SAFETY captures an important distinction between different kinds of luck in true belief, some but not all of which are incompatible with knowledge. Let us reserve the term epistemic luck for the kind of luck that begets a failure of SAFETY. This is the kind of luck that is present in COIN IN THE HEAD. There are other kinds of non-epistemic luck that might lead to formation of a true belief, and these kinds of luck are not incompatible with knowledge. For instance, whatever luck is required for a true belief in God’s COIN is not epistemic luck in the sense in which we are using the term here.

We need, then, to keep different kinds of luck separate when assessing whether a the need for luck of some kind in the causal history of a belief implies that it cannot be knowledge. In the next section, I make the case that appeals to luck in evolutionary debunking arguments have run afoul of these distinctions.

3 Evolutionary luck

3.1 The appeal to luck in general form

As outlined in §1, some proponents of evolutionary debunking arguments highlight the epistemological relevance of the evolutionary forces which produced our ethical beliefs, and they do this by emphasizing the need for luck if these forces are to produce true beliefs. It is easy to see why luck of some kind or other would be necessary: it is plausible (and I will assume here) that the evolutionary forces
that shape these beliefs are not themselves concerned with producing true beliefs. While evolution might select for creatures who are disposed to form certain ethical beliefs, it is no part of the selective process that these beliefs are true. If one forms the beliefs evolution selects for, and thereby forms true beliefs, one is, in a sense, lucky.

But since we concluded the previous section by noting that not every instance of luck in forming a true belief is an instance of epistemic luck, we must be careful about what kind of luck is present owing to the evolutionary origins of the relevant beliefs. One might be lucky in some sense but still have knowledge, as in God’s Coin. We need to look more closely at exactly what kind of luck is present in the evolutionary origins of ethical beliefs.

Debunkers do not typically claim that every belief that has an evolutionary explanation is epistemically suspect, but only beliefs with a certain kind of evolutionary explanation. To get an initial grasp of the kind of explanation that debunkers have in mind, a contrast case may be helpful. First a little more detail on ethical beliefs: they are paradigmatic instances of beliefs whose evolutionary explanation—whatever it is—does not imply that those beliefs are true. Whatever selection advantages are conferred on organisms who form ethical beliefs, these advantages aren’t enjoyed by virtue of these beliefs being true.

For one illustration of this idea we can take the outline of the explanation favoured in Joyce (2006): ethical belief provides a selection advantage because organisms who have the capacity for ethical belief will do better at resisting the temptation to cheat to obtain short-term goals when such cheating is disadvantageous in the long term. Let us suppose that there is, as Joyce suggests, a selection advantage being resistant to short-term temptations, and that one way of acquiring this advantage is to acquire the capacity to form ethical beliefs. The crucial feature of this story for the debunker’s purposes is that this selection advantage is available regardless of whether it is true that one ought to avoid succumbing to temptation.

In this respect there is an important contrast with an evolutionary explanation of perceptual beliefs. A system for forming beliefs about one’s immediate environment will plausibly confer a selection advantage on an organism only if it produces true beliefs about the organism’s environment. Quite plausibly, the

---

9 See Vavova (forthcoming) for more on the details of the debunker’s evolutionary premise.

10 Here is Joyce:

[A] person who in addition to being sympathetic judges that cheating is morally wrong will feel very differently if on occasion she succumbs to temptation. She can tell herself that she has done something wrong, that her action was unfair or unjust, that she must make amends, that she not only has risked punishment but deserves it. The emotion of guilt is available to her [...] The fact that these more robust forms of self-recrimination are available to the moralized thinker when she does cheat strongly suggests that when she is behaving herself her motivation not to cheat is more reliable and resolute than that of her non-moralized counterpart. (Joyce (2006, 112-3))

It should be noted that this idea is not wholly uncontroversial: one can read Enoch (2011, 168 ff.) as challenging the claim that every evolutionary explanation for ethical belief will not imply that, or even raise the probability that, those beliefs are true.
advantage conferred by perceptual beliefs is due to their facilitation of finding food, avoiding predators, and engaging in other behaviours that are favoured by natural selection. Only perceptual systems that generally produce true beliefs about these matters will play this role.

If our ethical beliefs are true, then, it is because of some kind of luck which is not needed for our perceptual beliefs to be true. The luck-based rhetoric which accompanies many debunking arguments is plausibly put forward with this contrast between the evolutionary explanations of ethical belief and perceptual belief in mind. While it is no accident that (most of) our perceptual beliefs are true, given their evolutionary explanation, the same can’t be said for ethical beliefs. Thus it is uncontroversial (or so I will assume) that there is a kind of luck needed for ethical beliefs to be true. Let us call this *evolutionary luck*. Ethical beliefs require evolutionary luck in order to be true, while perceptual beliefs do not.

By calling this a species of luck, we haven’t thereby shown anything about the epistemic status of these beliefs. Evolutionary luck isn’t necessarily a species of epistemic luck. Rather, accusations of a need for luck will be relevant to a successful debunking argument only if an additional premise connecting the presence of evolutionary luck with epistemic luck holds. This is LUCK CONNECTION:

\[
\text{LUCK CONNECTION} \quad \text{If a belief requires evolutionary luck, then it requires epistemic luck.}^{11}
\]

If LUCK CONNECTION is true, then the epistemic consequences that attend to epistemic luck, including its incompatibility with knowledge, are available to the proponent of the debunking argument. Allegations of an objectionable reliance on luck in ethical belief will be entirely apt. In particular, the debunker can claim that since it is a consequence of evolutionary theory that ethical beliefs are subject to evolutionary luck, it follows from LUCK CONNECTION that these beliefs are subject to epistemic luck. Then, since epistemic luck is incompatible with knowledge, it follows that these beliefs are not knowledge. They should be given up.

In the remainder of this section, I will consider two ways of arguing that LUCK CONNECTION is true.

3.2 Conceptual possibilities

One way of doing this is to appeal to a sufficiently broad space of possibilities, and then to point out that, even if evolutionary influenced beliefs happen to be true, those same beliefs are false in most of the worlds in the space of possibilities. Street suggests a picture like this when she says:

[A]s a matter of sheer luck, evolutionary pressures affected our evaluative attitudes in such a way that they just happened to land on or near the true normative views among all the conceptually possible ones.\(^{12}\)

---

\(^{11}\)Terminological note: I will use phrases like ‘belief \(b\) requires evolutionary luck’ as shorthand for ‘if \(b\) is true, then \(b\) ’s truth is a matter of evolutionary luck’. Similarly for the phrase ‘\(b\) requires epistemic luck’.

\(^{12}\)Street (2008, 208-9). See also Clarke-Doane (2012, 334 ff.) for more discussion of this aspect of the debunking argument.
If the evolutionary explanation for a belief doesn’t imply that the belief is true (that is, if the truth of the belief requires evolutionary luck), then it is conceptually possible that the evolutionary explanation is true but the belief false. This gives us a first premise:

(1) If a belief requires evolutionary luck, then most conceptually possible worlds are worlds where it is formed by similar processes but is false;

This observation can then be turned into an argument for Luck Connection with the following premise:

(2) If most conceptually possible worlds are worlds where a belief is formed by similar processes but is false, then that belief requires epistemic luck.

Given that (1) seems plausible, a successful debunking argument would be up and running if (2) were true. But it is not. The conceptually possible worlds in question need not be—and are very likely not to be—close worlds, as Safety requires.\(^\text{13}\)

There is more to be said about (2), although it is of questionable help to the debunker. In a Safety-centric framework, these distant possibilities are not irrelevant to what one can know. This is because of the following: while knowing \(p\) requires that the belief that \(p\) is safe; knowing that one knows \(p\) requires safely believing that one knows \(p\). Knowing that one knows \(p\), in other words, requires that the belief in \(p\) be safely safe: not only could the belief that \(p\) not easily have been false, but in addition the belief that belief in \(p\) is safe itself could not easily have been false.

These do not come to the same thing. To see this, consider the following case:

**Roller Coaster** Sally is a 8 year old child who wants to ride the roller coaster Vortex. There is a minimum height requirement of 50 inches for the ride; anyone who fails to meet this threshold cannot ride. As a matter of fact, Sally is 52 inches tall and can ride, but you cannot know her exact height: you are looking at her from a distance, have no independent information about her. Your powers of discrimination in environments like this are good: the margin for error for your estimates is ±2 in, so for any height \(h\) you can know that a person of height \(h\) in your environment is between \(h - 2\) and \(h + 2\) inches tall. Since Sally is as a matter of fact 52 inches tall, if you believe that she is between 50 and 54 inches tall, you know that she is between 50 and 54 inches tall, and hence know that she can ride.\(^\text{14}\)

\(^{13}\)Take a belief in the proposition \(p\) in the actual world, and suppose that \(p\) is true but there are conceptually possible worlds \(w^\dagger\) where in \(w^\dagger\) one believes a proposition \(p^\dagger\) similar to \(p\), and the token process which produces the belief in \(p^\dagger\) in \(w^\dagger\) is similar to the token process which produces belief in \(p\) in the actual world, but the belief in \(p^\dagger\) is not true in \(w^\dagger\). Since the conceptually possible world \(w^\dagger\) is plausibly not a world one could easily have been in, it is not close in the relevant sense. Hence we do not have a failure of Safety for the relevant beliefs.

See Clark and Rabinowitz (2011) for a related point in the case of religious belief.

\(^{14}\)See similar cases in Williamson (2000).
In Roller Coaster, your belief that Sally's height is between 50 and 54 inches tall is safe, given the margin for error of your perceptual faculties in your environment. Call the content of this belief $S_{50-54}$. And let there be a “world” for each height Sally might have: a world $w_{48}$ where she is 48 inches tall, a world $w_{49}$ where she is 49 inches tall, and so on. $S_{50-54}$ is false in $w_{49}$, but you couldn’t easily believe $S_{50-54}$ in $w_{49}$: by hypothesis, the perceptual faculties you use to form your belief are fairly accurate in your environment. So a true belief in $S_{50-54}$ safe in $w_{52}$; hence one is in a position to know $S_{50-54}$ in $w_{52}$.

But it isn’t safely safe, and hence you can’t know that you know $S_{50-54}$ in $w_{52}$. This is, you can’t know $KS_{50-54}$. This is because the belief in $KS_{50-54}$ would be unsafe in $w_{52}$. Suppose you believe $KS_{50-54}$ in $w_{52}$. This belief is safe only if the belief in $S_{50-54}$ is safe in close worlds. $w_{51}$ is, by hypothesis, a close world. But the belief in $S_{50-54}$ isn’t safe there: in $w_{51}$, $w_{49}$ is a close world given your perceptual faculties and environment, and in $w_{49}$ it is false that Sally is between 50 and 54 inches tall.

The important point here is just that $w_{49}$ is not close to $w_{52}$, but it is close to worlds $w_{52}$ is close to, and hence some proposition that you can know in $w_{52}$ but is false in $w_{49}$ cannot be known to be known. A similar pattern applies to further iterations of knowledge.

This suggests that the conceptual possibilities, where one’s ethical beliefs that are produced by the same evolutionary mechanisms but are false, are not irrelevant for what one actually knows. If these worlds are close to close worlds, or are close to worlds that are close to close worlds, etc., it follows that one cannot have some iteration of knowledge of ethical propositions. (One cannot know that one knows them, or one cannot know that one knows that one knows them, etc.) Thus the presence of evolutionary luck in ethical belief, on some additional assumptions, implies that at least some iteration of knowledge is unavailable.

But debunking arguments typically take the epistemic import of evolutionary luck to be stronger than this—they not only take the presence of evolutionary luck to show that some iteration of knowledge of ethical claims is unavailable; they take it to show that no knowledge is available. Thus the argument from conceptual possibilities needs to get from (1) the first iteration of knowledge of ethical claims is unavailable. We could get this result if we help ourselves to a number of instances a “collapse” premise which claims that, if one doesn’t know that one knows a claim, one doesn’t know that claim either. While such assumptions are sometimes tacitly appealed to and can seem appealing when not made entirely explicit, it would be unfortunate from the debunker’s perspective if her reliance on luck-based considerations required acceptance of such a controversial principle.

### 3.3 Safety and token causal processes

Coin in the Head illustrates another way in which ethical beliefs might require epistemic luck. In this case, one fails to satisfy Safety because in nearby worlds one has similar but false beliefs. Cases like this are importantly different

---

15 Here and in what follows I use the operator $K$ to stand for ‘one is in a position to know that’.

16 This is labelled the ‘KK’ premise in Williamson (2000).
from the (very distant) conceptual possibilities where evolution leads one form
the same ethical beliefs but the ethical facts are different; here it is variation in
one’s beliefs, rather than variation in the target facts, that potentially subject one’s
actual beliefs to a risk of error of the relevant kind. Thus the COIN IN THE HEAD
model thus suggests another strategy for establishing LUCK CONNECTION.

The salient feature of COIN IN THE HEAD that gives rise to epistemic luck is the
presence of a cognitive mechanism—in this case, a box containing an objectively
random coin flip that causally explains one’s beliefs—which produces similar but
false beliefs in nearby worlds. Evolution might be thought to be working like the
box in cases where beliefs are subject to evolutionary luck: if natural selection
doesn’t favour ethical beliefs because they are true, then it might have taken a
different course where it favours similar but false beliefs.

Just as the box in COIN IN THE HEAD implies that a problematic kind of
epistemic luck is present, one might think that for a similar reason evolutionary
luck implies that epistemic luck is present. This would establish LUCK CONNEC-
tion. Roughly: if a belief is subject to evolutionary luck, then there could easily
have been different evolutionary histories that produce different and incompatible
beliefs. Hence beliefs that require evolutionary luck fail SAFETY.

There are two premises in this argument:

(3) If a belief requires evolutionary luck, then evolution could have easily have
produced a similar but false belief;

(4) If a belief is such that evolution could have easily have produced a similar
but false belief, then that belief requires epistemic luck.

Suppose we grant (3). Still, (4) is false. The reason is that (4) doesn’t guarantee
that the false belief in the nearby world is formed by a sufficiently similar token
causal process.\footnote{Take a belief in the proposition \( p \) in the actual world, where the belief has an evolutionary
explanation which implies that it is subject to evolutionary luck, and suppose that \( p \) is true. There
are nearby worlds \( w^4 \) where in \( w^4 \) the evolutionary forces produces a belief in \( \neg p \), which is similar to
belief in \( p \) in the actual world. Moreover, we can suppose that the belief in \( \neg p \) is not true in \( w^4 \) since
nothing in the evolutionary explanation of the belief in \( \neg p \) implies that this belief is true. But since
the belief in \( \neg p \) in the world \( w^4 \) need not be formed by a sufficiently similar token process—and
plausibly will be a very different process, since by hypothesis the course of evolution that produces
the belief is very different in \( w^4 \)—we do not guarantee a failure of SAFETY for the belief in \( p \).}

This isn’t a mere technical quibble. As an illustration, consider a simple version
of a plausible evolutionary explanation for religious belief which plausibly implies
that religious beliefs are subject to evolutionary luck. On this hypothesis, belief in
non-embodied supernatural entities is a product of a hyper-active agency detection
system (a “HAAD” for short). A HAAD is a cognitive mechanism that is prone to
attribute changes in one’s environment to agents, rather than non-agential forces.
The selection advantage of a HAAD is of the “better safe than sorry” variety:
believing that the rustling in nearby bushes is caused by a predator or enemy
rather than the wind will cause you to take evasive measures, and such behaviour
promotes survival over the long run. But such a system will overgenerate, though
not for reasons that directly contribute to its survival value: creatures with a
HAAD will be disposed to find gods in cloud formations, spirits in unexplained occurrences in their surroundings, and so on.\textsuperscript{18} The HAAD hypothesis is one on which supernatural belief requires evolutionary luck, in our sense: a HAAD would have survival value whether or not the supernatural beliefs it produces are true.

But the this hypothesis does not imply that beliefs in the supernatural are epistemically lucky simply because evolution could have easily have produced a similar but false belief. One might imagine evolution taking a different path and thereby producing organisms who believe in the non-existence of supernatural entities of any kind. But plausibly these beliefs will be produced by a very dissimilar process—in worlds where evolutionary processes explain rejection of the supernatural, nothing like a HAAD will be involved in the token causal processes leading to the production of these beliefs. So (4) is false for the reasons sketched above in the HAAD case. Similar points apply to the epistemic import of evolutionary luck and errors in nearby worlds in ethical belief as well.

4 When do epistemic and evolutionary luck coincide?: the way forward

The previous section rejects two arguments for Luck Connection. This naturally suggests that the mere existence of evolutionary luck is not sufficient to establish a failure of Safety. And the bigger picture lesson here is that, absent further arguments for Luck Connection, any plausibility debunking arguments derive from their emphasis on luck (and related notions) is undeserved. There is a real risk that by emphasizing the need for luck of some kind on the basis of an evolutionary explanation of ethical belief, debunkers conflate two distinct notions.

While there is a risk that these categories are conflated, this doesn’t mean that the categories of evolutionary and epistemic luck never overlap. One might attempt a less ambitious project and try to show that, even if a failure of Safety does not follow from the presence of evolutionary luck alone, it does follow given other features of the evolutionary explanation for ethical belief. The prospects for luck to assist the debunker’s argument are then not closed, but will have to pay special attention to whether it is empirically plausible that the best evolutionary explanations for ethical belief have the needed features. I will close by outlining some constraints on a project of this kind.

4.1 Belief-forming processes

I have been granting for the sake of argument that if there is an evolutionary explanation for a belief, then individual tokens of the belief produced by the actual process of evolution are all the products of similar token causal processes. Without this assumption, the relationship between evolutionary luck and a failure of Safety would be in even worse shape than what I have claimed so far. To see why, suppose we deny this and hold that, for two tokens of an ethical belief, \(b_1\) and \(b_2\), it is possible that (i) ethical beliefs are explained by evolutionary processes as they actually occurred, but (ii) \(b_1\) and \(b_2\) are formed by very dissimilar token causal processes. In this case, the fact that ethical beliefs share an evolutionary

\textsuperscript{18}Bloom (2009)
explanation as in (i) would not even be a starting point for showing that anyone’s ethical beliefs fail to satisfy Safety. For if (ii) were true, then the fact that evolutionary processes could have easily produced false beliefs gives no reason to think that knowledge is not present; the nearby false beliefs might all be beliefs formed by sufficiently different processes. To avoid this result, and to preserve the possibility of a relationship between debunking arguments and epistemic luck, we need an evolutionary explanation \( E \) that satisfies the following claim:

**Process Similarity** For any two token ethical beliefs \( b_1 \) and \( b_2 \), if ethical beliefs have the evolutionary explanation \( E \), then \( b_1 \) and \( b_2 \) are formed by similar token causal processes.

It isn’t obvious that Process Similarity is true for any evolutionary explanations in other domains: presumably all human visual systems have the same evolutionary explanation, but it is false that any two token beliefs in the proposition that *War and Peace* is a long book formed by perception are produces of relevantly similar causal processes. You might believe this by looking at the table of contents and noticing the page count, and your friend might believe it by looking at the closed book from the side and noticing that it is thick. Suppose both processes are ones that you might have instantiated (you and your friend are in the same room) and you could easily have formed a false belief by looking at the table of contents (since the printer was highly unreliable). If you actually look at the book from the side, it doesn’t follow that you could easily have had a false belief by a similar process because you could easily have looked at the table of contents instead. And this is so even though you would be using your visual system in either case. The token causal visual processes producing the belief that *War and Peace* is long are not sufficiently similar.

There is however some cause for optimism for the debunker here. It is somewhat plausible that, given some psychological explanations for religious belief, the processes leading up to formation of such beliefs are sufficiently robust so as to imply analogues of Process Similarity. Suppose for example that religious belief is the product of wishful thinking. Leaving the empirical plausibility of the wishful thinking theory to the side, the theory plausibly which entails the failure of Safety for religious beliefs: even if wishful thinking produces true religious beliefs in a subject, similar but false beliefs could easily have been formed by a similar process. One could easily have wanted different things (one could want for instance that atheism rather than theism to be true), and in this case one would have formed similar by false beliefs by similar processes.\(^{19}\)

The upshot is that the wishful thinking theory plausibly entails that individual beliefs produced by wishful thinking will count as products of sufficiently similar

\(^{19}\)See for example Nagel (1997, 130):

I want atheism to be true and am made uneasy by the fact that some of the most intelligent and well-informed people I know are religious believers. It isn’t just that I don’t believe in God and, naturally, hope that I’m right in my belief. It’s that I hope there is no God! I don’t want there to be a God; I don’t want the universe to be like that.
token causal processes. An analogue of Process Similarity is then true of religious belief on the wishful thinking theory. But this isn’t guaranteed to hold once we substitute other theories and other target beliefs into the equation. Taking different evolutionary explanations as values of $E$ in Process Similarity is not guaranteed to preserve truth-value. Thus the prospects for reviving luck-based considerations in debunking arguments will depend on whether the robustness of the relevant mechanisms in the most empirically plausible evolutionary explanations for ethical belief.

4.2 Propensities and environmental luck

Another issue that has so far been left to the side, but will need to be raised for a full evaluation of the relationship between luck and debunking, is the following. It is unlikely that evolution properly explains ethical belief; what is more likely is that it explains a widespread propensity to form these beliefs. Propensities will, at best, constitute only a partial explanation for why a particular individual has a belief. To illustrate with a previous example in the case of religious belief: suppose a HAAD-like faculty is responsible for a propensity to form beliefs in the supernatural. One wouldn’t want to say that the presence of a HAAD guarantees that an agent will have such beliefs—plenty of atheists exist, and presumably their cognitive architectures include a HAAD which confers an unactualised propensity to form beliefs in the supernatural.

If mere propensities are the products of evolutionary explanation, this is doubly relevant to questions of Safety. First, it makes the claim that one could easily have formed similar but false beliefs easier to establish. But, second, it also makes the claim that the false beliefs would have been formed by a similar process more difficult to establish. I take each of these in turn.

First: as an initial approximation, take the relevant propensities to be dispositions that manifest themselves in some but not necessarily all environments. A propensity to form certain ethical beliefs, then, is a disposition to form such beliefs in certain environments. Moreover, this disposition need not manifest itself by producing the same ethical beliefs across a range of different environments. (Recall Joyce’s explanation of ethical belief as conferring a selection advantage by enabling organisms to resist the temptation to cheat for short-term gains. The disposition that confers the relevant selection advantage underdetermines which specific ethical beliefs an organism with the disposition will have: formation of sophisticated utilitarian ethical beliefs will not necessarily confer more or less of an advantage than formation of ethical beliefs in a rights-based morality.) Which specific set of beliefs a token individual ends up with is explained not just by this disposition, but by its interaction with the individual’s other cognitive faculties, cultural setting, and other environmental factors.

If the disposition to form ethical beliefs might have produced different beliefs in different environments, one might naturally reason as follows: one could easily have been in a different environment, and if one were in that environment, one’s disposition to form ethical beliefs would have manifested itself by leading to the formation of ethical beliefs that are distinct from the beliefs one actually has. Taking propensities seriously thus makes one component of a failure of Safety
easier to establish, since it is highly plausible that the same propensity could easily have been set in different environments, and hence could easily have produced different beliefs.

Now for the second point: if we focus on the fact that evolution confers only dispositions to form ethical beliefs, and that these dispositions output different beliefs when placed in different environments, it becomes less obvious that the output ethical beliefs are the products of sufficiently similar token causal processes. Let $D$ be the disposition to form ethical beliefs which evolution has selected for. And let $E_1$ and $E_2$ be different environments broadly construed in which $D$ might be placed, where $D$ produces the belief $b_1$ in $E_1$ and the belief $b_2$ in $E_2$. And finally, let us suppose that the beliefs $b_1$ and $b_2$ cannot both be true in their respective environments. The question pertinent to Safety is then whether the token causal process which involves $D$ in the environment $E_1$ and produces $b_1$ is sufficiently similar to the token process which involves $D$ in $E_2$ and produces $b_2$.

If evolution is going to have implications for the presence of epistemic luck, the answer to be ‘yes’. This is a question, however, which cannot be answered in the abstract. Merely from the fact that $D$ has an evolutionary explanation, nothing follows for questions such as whether token processes involving $D$-in-$E_1$ and $D$-in-$E_2$ are sufficiently similar. Different answers to this question may well be plausible depending on what exactly the disposition $D$ is. Some dispositions evolution selects for might be realised in very robust cognitive mechanisms where it is quite natural to think that any two beliefs that are products of such a mechanisms are formed by similar token processes, even if strictly speaking other inputs to the mechanism are needed for it to produce any beliefs. But there is no guarantee that evolution selects for a disposition that is realised in this way; other hypotheses may deliver different results. Crucial questions about the status of ethical belief vis-à-vis Safety are such that their answers depend on—and require attention to—the results of evolutionary biology and cognitive science.

5 Conclusions

With Safety as a plausible formulation of the kind of luck that is incompatible with knowledge, we can make substantial progress on the extent to which luck is relevant to evolutionary debunking arguments in ethics. Debunkers frequently allude to the need for luck to arrive at true ethical belief, given the broad contours of their evolutionary history. But with a firm conception of what distinguishes the luck present in Coin in the Head from the luck in God’s Coin in hand, the kind of evolutionary luck that evolutionary theory tells us is present in ethical belief does not by itself imply the presence of epistemic luck. If debunkers wish to claim that it is luck which is doing the work in their arguments, they will

---

20 Obviously the prospects are better here than they are in §2.3: in the present case, we have at least guaranteed that the disposition $D$ is a part of the token belief-forming process in both cases. The earlier argument did not hold this fixed. Premise (3) only holds that evolutionary processes could have produced a different and false beliefs; the presence of the same disposition is not guaranteed.

21 Think, for example, of the suggestion in §3.1 that a wishful thinking mechanism produces religious belief.
need to marshal specific (and no doubt controversial) evolutionary explanations of ethical belief which fit the contours outlined in §3. Some, but by no means all, evolutionary mechanisms will serve the debunker’s needs.

This does not mean that evolutionary debunking arguments can be safely ignored, even if the relevant evolutionary mechanisms are not forthcoming. Debunkers might instead retract their suggestions that the evolutionary luck needed for arriving at true ethical beliefs is central to their argument. They can scale back their use of luck-based rhetoric, and focus on other means of debunking ethical belief instead. Nothing I have said here rules out this strategy. But in closing it will be worth pointing out how the debate over debunking arguments can be expected to play out given such a move. Debunkers will have to do more than simply remove the term ‘luck’ and its cognates from their writings. In addition, supposed analogies with non-evolutionary cases where epistemic luck explains the intuitive verdicts will need to be discarded. It is worth revisiting Joyce’s Napoleon pill case in this setting: since the presence of epistemic luck explains why one’s Napoleon belief isn’t knowledge (and, if the need for luck is known, provides a defeater) it will no longer serve as a motivating analogy for the debunker’s thought that ethical beliefs explained by evolutionary forces are similarly suspect. For we know that the epistemically important features of the Napoleon case are not present in the evolutionary case, and so any intuitive support for the debunker’s verdict from the analogy goes missing.

The debunker thus faces a choice point when giving sufficient attention to the different kinds of luck. She can stick rhetoric which claims that luck does the work in explaining the force of her debunking of ethical belief, and try to show that the empirical details of our best theories are not too unfriendly. Or she can give up on the suggestion that it is luck that does the debunking work; in this case she will find motivating analogies for her argument much harder to come by. Either one of these options may turn out to be viable. But regardless of the outcome, extended reflection on the role of luck in evolutionary debunking arguments points to some directions which are dead ends for debunkers, sheds light on directions in which a successful argument might be developed, and imposes significant constraints on what would constitute success for the debunker.

References


