Dan Baras
Hebrew University

Justin Clarke-Doane
Columbia University

[This is the penultimate draft of paper that is forthcoming in *Philosophy and Phenomenological Research*.]

Modal Security

In this paper we critically examine the following thesis:

Modal Security: If evidence, E, *undermines* our belief that P, then E gives us reason to doubt

that our belief is sensitive or safe.[1]

Modal Security is proposed as a necessary condition on undermining defeat. It is not put forth as an analysis of undermining, since it does not even include a sufficient condition for a belief's being undermined. It says that if evidence undermines (rather than rebuts) one's belief, then one gets reason to believe that her belief is unsafe or insensitive (or both).

Modal Security has been widely discussed of late (Berry, unpublished ms.; Faraci 2019; Korman and Locke, forthcoming; Jonas 2016; Schechter 2018; Tersman 2016; Warren and Waxman, unpublished ms.; Woods 2018, Klenk, unpublished ms., Schafer 2017). If it is true, then influential arguments against moral, mathematical, logical, theological and modal realism would seem to fail (see Section 3). Such arguments include Genealogical Debunking Arguments (Joyce 2001, Ruse 1986, Street 2006), as well the Benacerraf-Field Challenge

---

[1] This is Clarke-Doane's (forthcoming) most recent formulation. Note that previous formulations varied in some details. In Clarke-Doane (2015) the principal is supposed to apply to all defeaters, not just undermining defeaters. In addition, while here it is formulated as a principle about individual beliefs, in Clarke-Doane (2015), Clarke-Doane (2016a) and (2016b), it is supposed to apply to beliefs of a kind.

(Benacerraf 1973, Field 1989, 25–30). Consequently, Modal Security would seem to have broad ramifications for the realism-antirealism debate across philosophy.

There are alternatives to Modal Security in the literature.[2] An alternative proposal of interest says that learning that there is no "explanatory connection" between our belief and the corresponding fact is undermining, even if this gives us no reason to doubt the modal security of our belief (Korman and Locke, forthcoming; Faraci 2019). In Section 10, we will show that a natural argument for this alternative actually serves to support Modal Security.

The structure of the article is as follows. In the next three sections we explain some of the motivations behind Modal Security and why it matters whether or not it is true. We describe the role it plays in so-called Genealogical Debunking Arguments against moral realism, as well as structurally similar arguments against realism about other domains, like mathematics and logic. We will be brief in these sections, as they summarize arguments developed in detail elsewhere. Our contributions to the debate come in the sections that follow. We raise a number of potential counterexamples and theoretical challenges to Modal Security. Some of these are developments of objections that have been raised in the literature, but most are new. We develop each of these objections into a valid argument and then examine the argument. One of the aims of the paper is to expose the weakness of these objections. Another is to reveal how the debate over Modal Security interacts with core problems in epistemology — including the generality problem, and the distinction between direct and indirect evidence.

---

[2] Interestingly, John Pollock's original definition of undercutting defeaters may imply Modal Security. According to Pollock "If believing P is a defeasible reason for S to believe Q, M* is an undercutting defeater for this reason if and only if M* is a defeater and M* is a reason for S to doubt or deny that P would not be true unless Q were true" (Pollock & Cruz 1999, p. 196). The modal condition that Pollock suggests, although not directly a modal relationship between belief and truth, as safety and sensitivity are, arguably has the implication that an undercutting defeater must call into question sensitivity. This holds if we assume that the belief forming method and the reason to believe Q covary.

1. Modal Security: Some Details

Sensitivity and safety play the dominant role in Modal Security, so we start with a definition of each. A lesson from the literature is that they must be defined relative to a belief forming method (Pritchard 2008).

Sensitivity: Our belief that P is sensitive iff had it been that ¬P, we would not still have

believed that P, had we used the method that we actually used to determine whether P.

Safety: Our belief that P is safe iff we could not have easily had a false belief as to whether or

not P, using the method that we actually used to determine whether or not P.

Note that there are two ways that our belief that P could fail to be safe. First, it could happen that P could have easily been false and we still believed that P (using the method that we actually used to determine whether or not P). Second, it could happen that we could have easily believed ¬P while P was still true. Hence, if we know that P's truth is modally robust, and if we have reason to believe that we could not have easily believed that ¬P (assuming that we actually believed P), then we have reason to believe that our belief that P is safe.

Although the above formulation of safety is adequate for many purposes, it actually needs to be complicated slightly. As Williamson points out, one's belief should only count as safe if "one avoids false belief in *every case similar enough* to [P] (Williamson 2002, 124, our emphasis)." In other words, our belief that P should *not* count as safe if we could have easily had a false belief as to whether Q, where Q is any proposition similar enough to P (using the method that we actually used to determine whether P). This is so even if we could not have easily had a false belief as to whether P *per se*. Imagine, for example, that while the calculating rules we appeal to when computing tips are unreliable in general, they give the

correct answer to a particular question, e.g., "what is 20% of $9.98?". In that case, our belief

that 20% of $9.98 is $2.00, which we formed using those rules, is unsafe. This is so even

though we could not have easily had a false belief as to whether 20% of $9.98 is $2.00 using

those rules. So, Safety should really be formulated:

Safety: Our belief that P is safe iff we could not have easily had a false belief as to whether or

not Q, where Q is any proposition similar enough to P, using the method that we

actually used to determine whether or not P.

What counts as a proposition "similar enough"? We do not have a helpful answer to provide

(Williamson's view appears to be that no informative analysis can be given). But this is an

instance of a kind of problem that pervades epistemology. An analogous question can be

raised about methods. How to individuate methods is the notorious "generality" problem.

Although it was initially posed as a challenge to process reliabilism (Conee and Feldman

1998), it is increasingly thought to plague a much broader set of epistemological theories

(Adler & Levin 2002; Bishop 2010; Baras, ms.). In fact, it is arguable that any serious

epistemological theory will face a version of it.[3] So, if there is a reason to reject Modal

Security in favor of some alternative account of undermining, this does not seem to be it.

However, the fact that safety and sensitivity must be relativised to belief forming methods,

and the fact that we know of no settled account of how belief forming methods should be

individuated, will play a role in the discussion that follows. As we will see, these facts

contribute to the difficulty of composing compelling counterexamples to Modal Security.

2. Motivations for Modal Security

Having said a bit about the content of Modal Security, let us turn to its motivation. A

preliminary thought runs as follows. Suppose a belief that P is initially justified. How can

new evidence defeat that justification? One way is simple: by being evidence that P is false.

---

[3] The most well-known and closest-to-home theories that require individuating belief forming methods are sensitivity and safety theories of knowledge. Let us explain briefly why. Start with sensitivity. The simplest formulation of sensitivity says that if p were false, then S would not have believed that p. According to Robert Nozick's tracking account of knowledge, sensitivity is a necessary condition for knowledge. Nozick himself already noticed that if sensitivity is not relativized to methods, then it falls prey to counterexamples such as "the grandma case". In that case, grandson is brought to grandma; she sees that he looks healthy and judges that all is well with him. This is quite a reliable way of forming a belief about her grandson's health. Is it sensitive? If grandson were sick, his parents would not bring him to grandma, and they would tell her that all is well in order not to worry her. Therefore, if grandson were not healthy, grandma would still believe that he is. Therefore, according to the simple formulation of sensitivity, the belief is insensitive, and if sensitivity is a necessary condition for knowledge, then grandma does not know that grandson is healthy. This result is counterintuitive. Nozick suggests, and this is a very intuitive move to make, that sensitivity should therefore be relativized to a belief-forming method: if p were false and the belief-forming method remained stable, then S would not believe p (Nozick 1981, 179; Pritchard 2008). In the grandma case, if grandson were sick, then grandma would form her belief based on parents' testimony, not based on looking at her grandson. Therefore, it is not a relevant counterfactual scenario. There are, of course, other problems with sensitivity and further epicycles that a sensitivity account must incorporate to match intuitive judgments. Our intention here is not to provide an exhaustive account of sensitivity but rather to claim that whatever the best sensitivity account turns out to be, we have good reason to believe that it will include a relativization to belief-forming methods or processes.

Safety based accounts face a similar challenge. The simplest formulation of safety says that S's belief that p is safe iff S could not have easily had a false belief about p. Again, this simple account fails, and a plausible move is to relativise to belief forming methods. Consider grandma again. Suppose that it could have very easily been the case that grandson is sick. If grandson were sick, she would believe parents' testimony that he's healthy and would have had a false belief. Does her current belief that grandson is healthy count as knowledge? Once again, the intuition is that it does. And a plausible way to revise sensitivity to get this result is to relativise to belief forming methods. Given that her current belief is formed using a very safe belief-forming method, namely looking at her grandson, her belief counts as safe and constitutes knowledge. The fact that she could have easily *relied on a different*, unsafe, belief forming method — namely parents' testimony — does not ruin this status of her belief.

That is what is called "rebutting" evidence. However, epistemologists commonly recognize a second type of defeater — namely "undercutting" or "undermining" defeaters (Pollock and Cruz 1999, 196) (we use "undermining" in what follows). How do undermining defeaters work? How could evidence defeat the justification of our belief that P without being evidence that P is false? Intuitively, it must show that there is some important epistemic feature that our *belief that P* lacks. Safety and sensitivity, along with truth, have emerged from recent epistemology literature as among the most important epistemological features of beliefs. Although they arose in the context of the analysis of knowledge, there is a growing recognition that they also play a role in epistemic justification.[4] Indeed, some propose a "translation" between conditions required for knowledge and conditions required for justification. The idea is that whenever you get evidence that a belief of yours does not satisfy the conditions for knowledge, that tends to defeat its justification. Conversely, if evidence E gives you no reason to doubt that a justified belief satisfies conditions for knowledge, then E cannot defeat its justification (Clarke-Doane 2016a, Section 2.8).[5] Only this latter "translation" principle will matter below. If it is correct, and if a (defeasibly) justified belief, which is both safe and sensitive, counts as knowledge, then non-rebutting evidence which fails to tell against its (defeasible) justification or its safety and sensitivity must fail to undermine it.

Another motivation for Modal Security is abductive. Paradigmatic underminers seem to conform to it. For example, if E is that we took a pill that, for some range of contents, P,

---

[4] Some examples for a safety condition on justification, see Handerson & Horgan (2001), Hirvelä (2017), Whiting (2018) as well as knowledge-first philosophers, such as Williamson (2002) and Kelp (2018), who believe that justification should be analyzed in terms of knowledge, and that safety is a necessary condition for knowledge. For sensitivity, see Braddock (2017), and, in a less-committed way, also Enoch et al (2012), as well as the critical discussion in White (2010, 580–581). In addition, the idea is present in much of the debunking literature, as we will see in section 3 below. [acknowledgment omitted]

[5] A similar translation scheme is defended by Lutz (forthcoming, sec 3.1) and Faraci & Elliott (unpublished ms.).

unpredictably gives rise to belief in P or belief in ¬P, then E seems to undermine our belief that P, or ¬P, respectively, by giving us reason to believe that we could have easily believed ¬P, or P, instead. Hence, given that whether P is true does not counterfactually depend on our belief as to P, our belief that P or ¬P fails to be safe. Something similar could perhaps be said of evidence E that there is widespread disagreement as to P — though whether such evidence is undermining is hotly contested. On the other hand, so long as the truths are contingent, and E is that we were "bound" to believe P or ¬P because of some constraining influence (such as a tendency to overrate ourselves) then E seems to undermine our belief in P or ¬P by giving us reason to doubt that had it been the case that ¬P or P, we would not still have believed P or ¬P, respectively. In other words, it seems to undermine our belief by giving us reason to doubt that it is sensitive. (What to say when the F-truths are necessary will be discussed.)

Safety and sensitivity conditions on knowledge or justified belief are rarely endorsed together. Typically, a theorist subscribes to either a safety condition or a sensitivity condition, but not both. Modal Security should not be misconstrued as saying that if E undermines our belief, then E gives us reason to doubt *both* the safety and the sensitivity of our belief (much less that if E gives us reason to doubt that both, then that undermines our belief). Modal Security says that if E undermines our belief, then E gives us reason to doubt *either* the safety or sensitivity of our belief. Equivalently, if we *fail* to get evidence that tells against the safety or the sensitivity of our belief, and we fail to get rebutting evidence, then our (defeasibly) justified belief is not undermined. Modal Security should, thus, be attractive to both safety and sensitivity theorists. Of course, such theorists may take Modal Security to be true because a stronger principle is true — namely, that if E undermines our belief, then E gives us reason to doubt its safety, or sensitivity, independent of considerations of sensitivity, or safety.

3. The Significance of Modal Security

Although Modal Security is weak in the sense that it may appeal to a variety of epistemologists, it is controversially strong in another sense. There is a case to be made that, if it is true, then influential arguments for moral, mathematical, logical, modal, and theological anti-realism must all be unsound. Let us briefly illustrate how the case is made in connection with so-called "Genealogical Debunking Argument" against moral realism.

Moral realism is the view, at first approximation, that there are (non-vacuous) mind-and-language independent moral truths that we can hope to discover. Recently, a style of argument has emerged that targets the justification of our moral beliefs, assuming that they are construed realistically. Arguments of this sort have come to be known as "Genealogical Debunking Arguments". Influential proponents of such arguments include Ruse (1986), Joyce (2001; 2006; 2016) and Street (2006; 2015). The following reasoning is paradigmatic.

> Nativism [the thesis that moral beliefs are evolutionarily innate] offers us a genealogical explanation of moral judgments that nowhere… presupposes that these beliefs are true… My contention...is that moral nativism… might well… render [moral beliefs] unjustified… In particular, any epistemological benefit-of-the-doubt that might have been extended to moral beliefs… will be neutralized by the availability of an empirically confirmed moral genealogy that nowhere...presupposes their truth (Joyce 2008, 216).

Debunkers' broadly empirical thesis is that the best explanation of why we have the moral beliefs that we do fails to imply their contents (or disquotational truth) — at least if the contents of our moral beliefs are understood as the realist claims they should be, as beliefs

about a mind-independent reality.[6] Because a thesis along these lines was pressed in Harman

(1977), we will call it *Harman's Thesis*. Debunkers' epistemological hypothesis is that

knowledge of Harman's Thesis *defeats* our moral beliefs. We will call this *Debunkers'*

Thesis.

Debunkers do not allege that knowledge of Harman's Thesis "rebuts" our moral beliefs —

i.e., that it gives us reason to believe the negation of their contents. It is not as if knowledge

of Harman's Thesis gives us reason to believe that it is *not* the case that slavery is morally

wrong, for example. Rather, debunkers claim that knowledge of Harman's Thesis *undermines*

our moral beliefs, or gives us reason to give up our moral beliefs, without bearing on their

contents (at least not "directly" — more on this in Section 7 and Section 8). So, the key

question surrounding such arguments is how — if at all — this could be.[7] And this is where

Modal Security comes in. Given common assumptions, Modal Security implies that

knowledge of Harman's Thesis cannot undermine our moral beliefs, realistically construed.

Before we explain why, let us note that this result ramifies. The soundness of any argument

with the structure of Genealogical Debunking Arguments aimed at realism about a domain

meeting two conditions turns on Modal Security in the same way (Clarke-Doane 2016a;

Clarke-Doane 2016b). The conditions are that the truths in question are modally robust, and

that there is an explanation of our having the corresponding beliefs that we do which shows

that we could not have easily had opposite ones (i.e., that we could not have easily believed

---

[6] We call it a "broadly" empirical thesis because the question of exactly what assumptions one must make in order to explain some observable phenomenon is not the kind of question empirical scientists tend to investigate — much as the question of exactly what axioms one must assume in a mathematical proof is not the kind of question mathematicians, as opposed to mathematical logicians, tend to investigate. (The premise should really be that *any* (true) explanation of our having the moral beliefs that we have fails to imply their truth. But this will not matter below. See [REDACTED])

[7] For discussion of what epistemic principles might be at work in evolutionary debunking arguments, see Vavova (2014), Bogardus (2016) and Sinclair (2018).

that ¬P, if we believe that P). Such arguments include the Benacerraf-Field Challenge for mathematical realism. Consequently, if the argument we are about to review works, and if Modal Security is true, then a range of epistemological arguments for anti-realism fail.[8]

The argument begins with the observation that moral truths factor into two kinds — what we will call "explanatorily basic" and non-basic kinds. Explanatorily basic truths take the form of conditionals, 'For all x, if F(x), then M(x)', where 'M' is a moral predicate, and 'F' is a non-moral one. But they are not just any conditionals of this form. They are necessary truths of this form.[9] Why assume that there are such conditionals? Because the moral strongly supervenes on the non-moral. There can be no change in the distribution of moral properties absent a change in the distribution of non-moral ones. But, then, if F(x) is a (perhaps infinitely conjunctive) non-moral predicate true of x, and M(x) is a moral properties true of x, then at least the following is necessary, by supervenience: 'For all x, if F(x), then M(x)'. On the other hand, non-basic moral truths are contingent. Some are atomic, like "Hitler is wicked". But they could be of any form.[10]

With this distinction in mind, let us first explain why Harman's Thesis does not seem to give us reason to doubt the safety of our moral beliefs. Evidence that our belief that P is unsafe is evidence that we could have easily had a false belief as to whether P (as well as contents similar enough to P). There are two ways it could happen that we could have easily had a false belief as to whether P. First, it could be that we could have easily believed ¬P, while P was the case. Second, it could be that it could have easily been the case that ¬P, while we believed that P. Now Harman's Thesis would give us reason to believe that it could have

---

[8] The arguments that we sketch are elaborated in detail elsewhere (see Clarke-Doane (2015), (2016a), (2016b), and (forthcoming, Ch. 4), for details)
[9] Most commonly the necessity involved is supposed to be metaphysical necessity. However, Fine (2002) and Rosen (manuscript) argue that a different type of modality is involved, normative necessity.
[10] Any (first-order) consistent form, that is.

easily been the case that ¬P, while we believed that P, if it gave us reason to believe that ¬P (assuming that what is actually the case could have easily been). But that would make Harman's Thesis a rebutting defeater, and all sides agree that it is not. So, if Harman's Thesis gives us reason to doubt the safety of our moral beliefs, then if gives us reason to doubt the following conditional: *if* our moral beliefs beliefs are true, *then* they are safe. But Harman's Thesis gives us no reason to believe that the moral truths could have easily been different from what they actually are. It has nothing to say about the extent to which those truths are contingent. Nor, however, does it give us reason to think that our beliefs could have easily been different. Maybe they could have been. But Harman's Thesis is no reason to think so. On the contrary, debunkers themselves commonly suggest that we were evolutionarily "bound" to have at least some of the moral beliefs we do have "independent of their truth" (Street 2006, 116).[11] But, then, for any moral belief that we possess, that P, Harman's Thesis fails to give us reason to believe that it could have easily been that ¬P, or that we could have easily believed that ¬P. So, assuming that no non-moral belief is similar enough to a moral belief in the sense of safety, it fails to give us reason to doubt that our moral beliefs are safe.

What about sensitivity? Evidence that our belief that P is insensitive is evidence that had it been that ¬P, we still would have believed that P. If P is some moral proposition that we believe, then Harman's Thesis would be such evidence if it were evidence for ¬P — since then the closest world in which ¬P is the actual world, and we believe that P there. But, again, that would make Harman's Thesis rebutting evidence, and we are assuming that it is not. So, if Harman's Thesis gives us reason to doubt that our moral beliefs are sensitive, then it gives us reason to doubt that if they are true, then they are sensitive. However, the contents of our

---

[11] Of course, if the true explanation of our moral beliefs is the sort of thing that could easily have been different, then we quickly get a reason to think our moral beliefs could easily have been different. (See, e.g., Barkhousen (2016).) Our point is merely that the fact that the explanation in question fails to imply the contents (or disquotational truth) of our moral beliefs is neither here nor there.)

explanatorily basic moral beliefs are necessary, if true, and — at least if the necessity in question is absolute necessity — beliefs in necessary truths are vacuously sensitive on a standard semantics. Even if the relevant counterfactuals are not vacuous, however, parity of reasoning suggests that if our explanatorily basic moral beliefs are not sensitive and are thereby undermined, then our explanatorily basic beliefs quite generally are. For instance, it seems equally true that had the truths which fix the supervenience of the phenomenal on the physical been different, our beliefs would have been the same. Hence, on pain of quite general skepticism, knowledge that our explanatorily basic moral beliefs are not (non-vacuously) sensitive cannot undermine them (Clarke-Doane forthcoming, Section 5.7).

Meanwhile, our contingent moral beliefs which are the subject of typical moral discussion — such as our belief that, e.g., Hitler is wicked — seem to be sensitive, if true, whether or not the explanatorily basic moral truths are necessary. Had A not been M, where A is a particular person, action, or event, and M is a moral property, then A would have been different in non-moral respects—since even if there are worlds in which the explanatorily basic moral truths are different and A is not M, these worlds are presumably more 'distant' from the actual world than worlds in which the explanatorily basic moral truths are the same and A is not M. So, Harman's Thesis gives us no reason to doubt that our contingent moral beliefs are sensitive either.[12]

---

[12] For analogous reasons, Harman's Thesis does not seem to give us reason to doubt that the probability that our moral beliefs (realistically construed) are true is high, contra Street (2006, 129). Either the probability in question is epistemic or it is objective. Whether the epistemic probability that our moral beliefs are true is high, given Harman's Thesis, is what is in dispute. But Harman's Thesis does not give us reason to doubt that the *objective* probability that our moral beliefs are true is high. For any explanatorily basic moral truth, P, presumably Pr(P) = 1, given that such truths are metaphysically necessary. (We *could* require that Pr(P) = 1 only if P is necessary in an even stronger sense – e.g., conceptually necessary. But then it would turn out that virtually every proposition of interest, including, e.g., all atomic mathematical propositions, have equal claim to being objectively improbable.) Also, for all that has been argued, it may be that Pr(we believe that P) $\simeq$ 1. However, then, it may be that Pr(P & we believe that P) $\simeq$ 1, by the probability calculus. Since (P & we believe that P) implies (our belief that P is true), it may be that Pr(our belief as to whether P is true) $\simeq$ 1 as well.

Of course, the argument above does not show that Harman's Thesis can never, for any agent, give one reason to doubt that her moral beliefs are safe or sensitive. It is possible, for instance, to imagine an agent to whom a benevolent and omniscient being whispered in her ear the conditional "if Harman's Thesis is true, then your moral beliefs are unsafe". For such an imaginary agent, Harman's Thesis would indeed be a reason to doubt that her moral beliefs are safe. Rather, the argument merely aims to show that Harman's Thesis does not give *us* reason to doubt the safety or sensitivity of our beliefs, realistically construed.

The upshot is that if Modal Security is true, then Harman's Thesis does not undermine our moral beliefs — i.e., Debunkers' Thesis is false. It is worth emphasizing that canonical formulations of debunking arguments conform to Modal Security. They suggest that Harman's Thesis gives us reason to doubt the safety or sensitivity of our moral beliefs, realistically construed (Clarke-Doane (2016a). For instance, Joyce writes, "moral judgments are the output of a non-truth-tracking process….[where] the intuition at the heart of truth-tracking is that beliefs may or may not be *sensitive* to the facts which they represent (Joyce 2016, 147, emphasis in original)." And Braddock, Mogensen, and Sinnott-Armstrong seem to invoke safety when they write, "different instantiations of the process of cultural group selection have produced divergent normative systems, which nonetheless solve the same design-problem: namely, that of getting human societies to function as adaptive corporate units. In this way, one and the same process type may, through its various instantiations, easily result in divergent moral systems (2012)." However, in light of the arguments that we just surveyed, debunkers increasingly defend Debunkers' Thesis by denying Modal Security. This latter line of argument is what brings us to examine Modal Security more closely.

4. Argument 1: From Unreflective Beliefs

Having sketched the content, motivation, and significance of Modal Security, let us turn to our main topic — its assessment. We think that a good way to assess the principle is to try and come up with counterexamples to it. So, the first three arguments that we will discuss will involve purported counterexamples. By now, a number of alleged counterexamples to the principle have been proffered and we will draw on, and strengthen, some of them. Counterexamples to Modal Security would be cases in which the justification of an agent's belief is undermined by evidence E, but E is not reason to doubt its safety or sensitivity.

Let us begin with a simple potential counterexample.[13] Many of our beliefs were formed unreflectively. Given our limited resources, one might think that we are justified in having many of those unreflective beliefs. However, once we reflect, we may start thinking about whether we really have good reasons to maintain those unreflective beliefs. We may discover, with regard to some of the beliefs, that in fact we do not really have a good reason to maintain them and, therefore, that we should give them up. Consider the following case:

Example 1: Sara believes that incest is morally impermissible, even if contraception is used, no feelings are hurt and there are no other foreseeable bad consequences. This just always seemed to her correct, unreflectively. Inspired by an ethics course, she reflects on this belief and realizes that she can't think of any reason to believe one way or other on this issue. She therefore becomes agnostic about the moral status of such incest.[14]

---

[13] Woods's (2018) objection to Modal Security seems to us in the spirit of the argument we formulate here, although it isn't fully explicit.

[14] The example is inspired by Haidt (2001).

Sara's reduction of credence seems justified, despite her having gotten no evidence against the contents of her beliefs. But, apparently, there is no need to appeal to the lack of safety or sensitivity of Sara's beliefs in explaining why. Her belief is undermined because she realized that she lacks reason to believe it! Is this not a counterexample to Modal Security? The argument can be summed up as follows:

(1) Prior to her reflection, Sara was epistemically justified in her belief about incest.

(2) Upon reflection, Sara's belief loses its epistemic justification.

(3) If Sara's belief loses its epistemic justification, then this is in virtue of an undermining defeater.

(4) If the epistemic justification of Sara's belief is undermined, then it is not in virtue of a reason to doubt the safety or the sensitivity or Sara's belief.

(5) If a belief is undermined but not by a reason to doubt the safety or sensitivity of the belief, then Modal Security is false.

(6) Therefore, Modal Security is false.

This argument has some *prima facie* plausibility. But we will argue that it is quite weak on inspection. Moreover, the ways in which an advocate of Modal Security can resist this simple argument will serve as a useful point of departure for handling more sophisticated ones.

The main problem with the argument is that premise (4) is question-begging. Even given (1) — (3), why suppose that the justification of Sara's belief is not undermined thanks to a reason to doubt its safety or sensitivity? Once Sara discovers that she has no reason to hold her belief, that *is* a reason to believe that it is unsafe and insensitive, given typical background beliefs. Believing things for no reason is a very unsafe and insensitive way to form beliefs!

15

We grant that Sara may not think in terms of safety and sensitivity explicitly. Perhaps the only thought that crosses her mind is that she has no reason to hold the particular belief. But this does not show that Sara's belief was undermined in virtue of something besides a reason to doubt its safety or sensitivity. Why, after all, is forming a belief without reason a problem? The obvious answer is that holding beliefs for no reason is not a very reliable way of forming them. And this may be just another way of saying that the belief is unsafe or insensitive.

Even if (4) were true, however, the above argument would not be compelling. To say that Sara has no reason to have a belief could mean different things. It could mean that she has no epistemic reason. Perhaps she only has the pragmatic reason that she is cognitively limited, and has to believe something. But, in that case, (1) is false — unless pragmatic reasons can also be epistemic. Alternatively, it could mean that Sara has no reason to have the belief *other than the fact that its content seems to her to be true*. The "seeming true" supplies the epistemic justification.[15] But, then, upon reflection, the content of Sara's belief either continues to seem true to her, or not. If it does not, then the justification of her belief may be lost, but not, it would seem, thanks to a defeater (contra (3)). A defeater is an epistemic reason, not a lack of one (Pollock & Cruz 1999, p. 196). On the other hand, if the content of Sara's belief *still* seems true to her upon reflection, then her belief has not lost its justification, contra (2). This is so even if Sara mistakenly believes that her belief lost its justification. Sara can be wrong about whether her belief is justified.

So, this first argument is quite weak. But there are more promising lines of attack.

---

[15] Proponents include Huemer (2007) and Tucker (2010). For discussion, see Tucker's edited volume (2013) and Hawthorne & Lasonen-Aarnio's critique (unpublished manuscript).

5. Argument 2: From Evolutionary Debunking Arguments

Here is a slightly more sophisticated example:

Example 2: Morris, also a philosophy student, has many initially justified moral beliefs. For a

course on debunking arguments, he is assigned Joyce's *The Evolution of Morality*. He

reads the book carefully and becomes convinced that all of his moral beliefs are

unjustified. He therefore becomes a moral skeptic.

Suppose that arguments mentioned above (section 3) are sound and that Joyce has,
accordingly, not actually provided any evidence for the insensitivity or unsafety of Morris's
beliefs. Nevertheless, it does seem like Morris's reduction of credence is reasonable, given
that he did the best he could to assess the evidence and the arguments. A nice feature of this
example is that it tries to refute Modal Security by its own lights, or at least by the lights of
the papers with respect to which the principle was conceived to begin with. Clarke-Doane
argues both that Joyce hasn't given us reason to doubt the sensitivity and safety of our moral
beliefs, and that only such a reason could defeat those beliefs. If Argument 2 is sound, one of
these claims must be false.

The structure of this argument is similar to the previous one:

(1) Prior to reading Joyce, Morris was epistemically justified in his moral beliefs.

(2) After reading Joyce, Morris's moral beliefs lose their epistemic justification.

(3) If Morris's beliefs lose their epistemic justification, then it is in virtue of an

undermining defeater.

(4) If the epistemic justification of Morris's beliefs is undermined, then it is not in virtue

of a reason to doubt the safety or the sensitivity or Morris's belief.

(5) If a belief is undermined but not by a reason to doubt the safety or sensitivity of the belief, then Modal Security is false.

(6) Therefore, Modal Security is false.

As mentioned, a nice feature of this argument is that premise (4) seems to be supported by advocates of Modal Security. But, as we will argue, deeper reflection on the argument suggests that whether this is so depends on more fine grained details of the case.

We must distinguish between different ways in which Joyce's book may have influenced Morris. One way in which the book may have influenced Morris is tby claiming that the evidence supports the conclusion that our moral beliefs are insensitive or unsafe. Indeed, as we mentioned in Section 3, Joyce explicitly suggests that the evolutionary "genealogy" of our moral beliefs shows that they are not sensitive. In an earlier book, he says:

> Suppose that the actual world contains real categorical requirements—the kind that would be necessary to render moral discourse true. In such a world humans will be disposed to make moral judgments...for natural selection will make it so. Now imagine instead that the actual world contains no such requirements at all—nothing to make moral discourse true. In such a world, humans will still be disposed to make these judgments … for natural selection will make it so...[D]oes the truth of moral judgments...play a role in their usefulness?....I believe the answer is "No." (2001, 163, emphasis in original).[16]

---

[16] We need not assume, with Joyce, that morality presupposes categorical requirements. (Note that, unlike the previous quote, Joyce is not strictly concerned with sensitivity in this quotation, since sensitivity operates at the level of individual beliefs. He is concerned with something more like skepticism about the external world. He is asking what we would have believed had the contents of our atomic moral beliefs been systematically false. It could be that, for any one of our atomic moral beliefs, that P, had it been that ¬P, we would not have believed that P, even while it is also true that had there been no atomic moral truths at all, we would have believed that P. It could be that the closest worlds in which ¬P is a world in which there are atomic moral truths. But the

18

On this interpretation, the book functions as an expert's testimony for lack of sensitivity or safety. Testimony for P can, of course, be a reason to believe P even when P is false. Thus, Joyce's book can be a reason for Morris to believe that his moral beliefs are insensitive or unsafe, even if none of the evolutionary information in the book (minus Joyce's testimony for lack of sensitivity) is evidence that our moral beliefs are insensitive or unsafe. Misleading testimony for insensitivity can undermine. And that is consistent with Modal Security. So, in this case, premise (4) is false.

A second way in which Morris might become convinced by the book is that Morris reads all the evolutionary claims and becomes convinced that they imply that his beliefs are insensitive or unsafe. Once again, if the arguments in Section 3 are sound, then Morris's inference is invalid. What should we say of an invalid inference? You may think that if Morris thinks that the inference is valid, then he is justified in believing the conclusion (assuming the premises are sufficiently supported). In that case, he has a reason to believe that his beliefs are insensitive or unsafe, and, again, the example does not actually violate Modal Security (again, premise (4) is rejected). Alternatively, you may think that Morris is not justified in believing the conclusion of such an invalid argument. But, in that case, his reduction of credence is not justified and this is not actually a case of epistemic defeat. That is, premise (2) is false.[17]

---

observation that had there been no atomic truths of a kind, our atomic beliefs of that kind would have been the same overgeneralizes. As Nozick (1981, 227–29) emphasizes, we would have believed that we have hands had there been no atomic perceptual truths — i.e., truths about the objects of ordinary perception — too. The best that we can hope for is that the closest worlds in which we do not have hands are worlds in which we do not believe that we do (using the method that we actually used to determine this). Again, in later work, Joyce appears to recognize this, and focuses on sensitivity *per se*.)

[17] Some philosophers think that if a subject believes that she has an undermining defeater of a certain belief of hers, then it is an undermining defeater. Michael Bergmann (2006, 263–268) explicitly defends this view, and it is also implied by Scott Sturgeon's (2014, 216) analysis of an example. (For discussion, see Casullo (2018) and Klenk (2019)). If this view is correct, then the mere fact that Morris thinks that Joyce's book defeats his beliefs guarantees that it does. This view is incompatible with Modal Security. A person can irrationally believe that

What if Joyce himself had claimed in his book that insensitivity and unsafety are not the only ways in which beliefs can be defeated — that is, what if Joyce claimed that Modal Security was false? Well, then, if Modal Security is true, Morris would be wrong about his epistemic condition! Morris's moral beliefs would not have lost their justification, even if his belief about their epistemic justification lost its justification. In that case, premise (2) would again be false. The situation is analogous to one in which a philosophy 101 student reads Descartes and comes away believing that justification requires infallibility. He leaves the class disbelieving that he has hands. Intuitively, however, his disbelief was not epistemically required, even if his belief that it is was itself required. Similarly, in the case of Morris.[18]

6. Argument 3: From Modally Stable Beliefs

Even if the first two arguments against Modal Security are not persuasive, it is commonly supposed that there is a broad class of counterexamples to the principle. These are cases where our beliefs are the products of a constraining influence, but their contents are necessary, if true (unlike the cases discussed in Section 2). Thus, both the contents and the beliefs are modally stable. For concreteness, we will present two such cases. The advantage of the first case is that it is easier to have the intuition that the relevant beliefs are undermined. However, it is also easier for us to show that it is not actually a counterexample to Modal Security. Things go the other way around for the second example. While it is a bit

---

she has a defeater of a certain belief, without having any reason to doubt that the belief is safe or sensitive — or without having any reason at all. We believe that this prediction is a reason to doubt Bergmann and Sturgeon's view rather than a reason to doubt Modal Security. On our view, such an irrational belief cannot be a defeater. (We are thus in agreement with Casullo (2018)). We thank an anonymous referee for urging us to consider this issue.)

[18] The Descartes example is inspired by Barnett (ms.). We say "required" rather than "justified" because the claim that Morris's previous moral beliefs remain justified does not entail that Morris's current disbelief is unjustified. If the epistemic uniqueness thesis is true, that is, that there is a unique set of justified beliefs for any evidential state, then Morris can't be both justified in maintaining his previous beliefs and in becoming skeptical. However, the uniqueness thesis is controversial (White 2005; Schoenfield 2012; 2018; Kelly 2013).

more challenging for us to respond to, it is also more difficult to imagine and therefore more difficult to have a clear intuition about. We believe that all examples of modally stable beliefs fall into one of these two categories.

The first case is the following:

> *Example 3:* Jan is defeasibly justified in forming beliefs about which numbers are prime by consulting a computer. Perhaps its results to date have been verified by direct computation, or he has been told by an otherwise trustworthy mathematician that this computer is reliable. Jan consults the computer to form beliefs about which numbers are prime. Jan then gets evidence, E, that the machinery has been stuck in the last 10 cases. As it happens, the last 10 numbers considered have been prime.

We share the intuition that E defeats Jan's belief that the last 10 numbers are prime. Given that E does not seem like a reason to believe that those numbers are not prime, if E is a defeater, it is an undermining defeater. But, given that those numbers could not have been composite, and that the machine, stuck as it is, could not have easily said that they were not prime, isn't this a counterexample to Modal Security? For reasons similar to those described in section 3, E does not seem (at first) like a reason to doubt the safety or sensitivity of the relevant beliefs.

It is not. E is actually a reason to doubt the safety of the beliefs. Recall our definition of safety says that for a belief P to be safe, it is not sufficient for it to be the case that the agent could not have easily had a false belief about P. Rather, it must also be the case that for any content Q similar enough to P, the agent could not have easily had a false belief. Now Jan as we likely imagine him could have easily entered composite numbers into the machine after it got stuck. If he would have done so, the machine would have falsely said that they were

21

prime numbers and Jan would have had a false belief. The reason is that if the machine had considered composite numbers last, it still would have called them prime.

The lesson is that in order to concoct a potential counterexample to Modal Security of the type we are after in this section, namely in which the beliefs and the truths are modally stable, we must add even more modal stability to our subject's beliefs:[19]

Example 4: Aya is defeasibly justified in forming beliefs about which numbers are prime by consulting a computer. Aya consults the computer to form beliefs about which numbers are prime. As it happens, the numbers the computer outputs are all prime numbers. Moreover, these events are *maximally deterministic*. At every metaphysically possible world, Aya's counterpart forms beliefs about which numbers are prime by consulting the computer's counterpart and at every possible world, the computer's counterpart delivers the same answers as at the actual world. Aya then gets evidence, E, that the computer has been generating numbers using a random number generator.

This example is hard to imagine and that is in itself a reason to doubt that we have any relevant intuitions. Still, some people claim they have the intuition that upon learning about the mechanism responsible for her beliefs, Aya's prime-number beliefs lose their justification, despite the fact that Aya got no evidence that the truths could have been different or that she could have easily had different beliefs.[20]

A valid argument can be formulated using the same structure as the previous ones:

---

[19] The example is inspired by Faraci (2019).

[20] See Roland & Cogburn (2011); Klenk (ms.); Schechter (2018); Korman & Locke (forthcoming); Lutz (forthcoming) for an expression of this intuition. In section 10 we will discuss the further intuition these authors have that the mere fact that one learns that there is no "connection" between the beliefs and the truths explains why the beliefs in such examples are undermined.

(1) Prior to learning of the random generator, Aya was epistemically justified in her beliefs.

(2) After learning of the random generator, Aya's beliefs lose their epistemic justification.

(3) If Aya's beliefs lose their epistemic justification, then this is in virtue of an undermining defeater.

(4) If the epistemic justification of Aya's beliefs is undermined, then it is not in virtue of a reason to doubt the safety or the sensitivity or Aya's belief.

(5) Example 3 describes a metaphysically possible case.

(6) If a belief in a metaphysically possible case is undermined but not by a reason to doubt the safety or sensitivity of the belief, then Modal Security is false.

(7) Therefore, Modal Security is false.

Before outlining the problems with this argument, note a feature of all examples like the above. If Aya's beliefs are perfectly stable across possible worlds, that implies that she cannot disbelieve P in an extremely strong sense — not just psychologically or even physically. She was fated to believe P in something like the sense that Oedipus was fated to marry his mother. This adds a complication. If ought implies can in the epistemic realm, then arguably if Aya cannot but believe P, her belief remains justified and premise (2) is false. To get around this problem, a modification of the example is needed.[21] So, let us stipulate that there *are* possible worlds in which Aya suspends judgment about P. There are merely no possible worlds in which she can believe ¬P. This amendment solves the problem, because

---

[21] This was suggested by David Faraci in personal correspondence.

in such worlds Aya would not have a false belief about P or ¬P and therefore those worlds would not threaten the safety of Aya's actual beliefs.

Even with this amendment of the example, the argument is not compelling. First, premise (4) is doubtful. Aya *does* arguably get reason to doubt the safety of her beliefs. Once again, recall that, when introducing sensitivity and safety, we noted that the principles must be relativized to belief forming methods and that it is not sufficient that the specific belief must be true in relevant possible worlds, rather, beliefs with similar contents must be true as well. (It will be sufficient for us to focus on safety in what follows, but the same can be said for sensitivity). So, to fully specify *Example 4* such that it is a potential counterexample to Modal Security, we need to say what the method in question is and check whether this method would yield false beliefs about similar contents. What method did Aya use to determine which numbers are prime? There is more than one possibility here. It could be the method of <consulting a computer that gives modally stable answers>. But, then, it could also be the method <trusting a random number generator to answer mathematical questions>. Generally speaking, the latter is not a safe belief forming method! In nearby possible worlds in which Aya uses a (non-modally-stable) random number generator to determine which numbers are prime, she has many false beliefs with contents similar to her current beliefs about which numbers are prime. The upshot is that it may be that we have the intuition that in the example the evidence in question is undermining because we individuate methods and count beliefs as similar in such a way that Aya's beliefs qualify as unsafe.[22]

---

[22] Does not the lack of a principled individuation of belief-forming methods suggest that Modal Security is less dialectically effective as part of a response to debunking arguments than one might hope? After all, there will always be a way of individuating belief forming methods such that the actual way in which we form our moral beliefs (say) turn out unsafe. It does suggest this. Since debunking arguments are not the topic of this paper, we will not discuss this issue in detail. However, we wish to make the following two remarks. First, this is not yet a triumph for debunkers because, for the same reason, there will also always be a way of carving up methods such that the given beliefs come out safe, if true. Second, Modal Security still implies that Harman's Thesis is

Even if we set this objection aside, however, the case seems incoherent (and if so, premise 5 is false). We cannot just stipulate metaphysical necessities! How could it be *metaphysically impossible* for a person to form mathematical beliefs in any other way than by consulting a random number generator? That is a paradigmatically metaphysically contingent fact, on any known articulation of the concept of metaphysical necessity. Moreover, what is a random number generator that gives prime outputs in *all metaphysically possible worlds*? We could certainly write down an algorithm, and program a machine to follow it, which, in fact, identifies prime numbers, but which we *mistakenly believe* is a random number generator. But, in that case, direct confirmation that the initial values are all prime would undercut our belief that the algorithm generates random numbers. If Aya were really to get (misleading) evidence that her number generator was a random number generator, then she would get evidence that its outputs could have easily been false, because they could have easily failed to be prime. But, then, premise (4) would be false, and the example would fail to refute Modal Security.

In general, a tension plagues such counterexamples. Either the worlds in which we have different, and so false, beliefs are "distant" (or nonexistent), or not. If so, then it is hard to see why, epistemically, we should care about them. If not, then there is nothing to preclude us from saying that the subject's beliefs are undermined by evidence of their lack of safety. Hence, while it is easy to gesture at cases where the truths in question are necessary, and someone was "bound" to have the beliefs she has, such cases are problematic on inspection.

They are either unimaginable, not undermining, or, may, in fact, involve evidence of lack of safety.

7. Argument 4: An Accuracy Based Redundancy Worry

We have considered potential counterexamples to Modal Security. As far as we can tell, all other alleged counterexamples developed in the literature are variations on the above. But there are other problems with the principle. We turn to them now. Even if there are no clear counterexamples to Modal Security, there are theoretical difficulties with the principle.

Consider the following attractive idea about epistemic justification. The only sense of epistemic justification of a belief that we should be interested in is such that is tightly linked with the expected accuracy of the belief. That is, a belief that P is epistemically justified for subject S to the extent that S should expect P to be true. How does this expected accuracy conception of epistemic justification reflect on Modal Security? There is an argument that it implies Modal Security, but also renders it redundant.

Modal Security is a principle about epistemic undermining. While it is an open question how to precisely delineate the distinction between undermining and rebutting, both are, by definition, types of epistemic defeaters. And an epistemic defeater is something — in this paper we assume that it is *evidence* — that reduces epistemic justification. The expected accuracy approach to epistemic justification implies that any defeater must be a reason to doubt that the belief is true. And this includes undermining defeaters. Hence, we get:

Accuracy Based View of Defeaters: In order for a belief that P to be defeated by evidence E, it
must be the case that E is a reason to doubt that P is true.

On a simple reading of both the accuracy based view and Modal Security, the accuracy based view of defeaters implies Modal Security. That is because any reason to doubt that P is true is also a reason to doubt that the belief that P is safe and sensitive, because a false belief is unsafe and insensitive (and in the worst way).[23]

One nice result of the Accuracy Based View that provides a nice explanation for cases in which evidence of unsafety does *not* defeat. Evidence of unsafety only has a defeating effect if it somehow gives one reason to doubt that the belief is actually true. Suppose, for instance, that you learn that a mad scientist could have very easily envatted your brain. He actually flipped a die and decided that only if the die lands six you'll be saved from envatment. Lucky for you, the die landed six. Once you learn that this is what happened, is your belief that you have hands defeated? Of course not.[24] The reason is that we are ultimately interested in whether P is true, not in whether the belief that P is safe.

Is this a good result for Modal Security theorists? That depends on our expectations. If we are only interested in whether Modal Security is true, then showing that it is implied by an attractive conception of epistemic justification is no doubt a good thing. However, if we are also interested in whether Modal Security is useful or explanatory, then this result has negative consequences. The reason is that according to this view, safety and sensitivity don't

---

[23] You may worry that Accuracy implies that there can only be rebutting defeaters, not undermining defeaters. This would be true if all reasons to doubt p are rebutting and not undermining. However, this need not be so. For one, rebutting defeaters are reasons to believe ¬P and it is not obvious that every reason to doubt P is a reason to believe ¬P. Some, for instance, are reasons to be suspend judgment about P, which is arguably something else. This question depends on how precisely the distinction between undermining and rebutting defeaters is understood. The original distinction comes from Pollock (Pollock and Cruz 1999, 196). However, Pollock's definition has problems, and some alternative proposals have their difficulties as well (Sturgeon 2014; Kotzen 2019; Pryor 2013). We will have more to say about these issues in Section 8.

[24] There are complications with regards to the particular example. You may think that a BIV counts as forming beliefs using different methods than non BIVs and therefore, the information in the example is not evidence for unsafety of the actual belief that P. It will be impossible to come up with an example in which this sort of response is not available so long as we lack agreement as to how methods should be individuated. Still, the point seems plausible to us.

really matter. It just so happens that any reason to doubt that a belief is true must also technically be a reason to doubt that the belief is safe. This point can be put in a stronger way. We need not check whether E is a reason to believe that the agent would have had a false belief about P in any possible world other than the actual one. This is also why Modal Security is not useful. It can mislead us to think that we need to examine many possibilities that we don't actually need to.

The argument, then, is that Modal Security lacks theoretical interest. It goes like this:

(1) Accuracy: In order for a belief that P to be defeated by evidence E, it must be the case that E is a reason to doubt that P is true.

(2) Accuracy implies Modal Security.

(3) If (1) and (2), then Modal Security is true but unexplanatory and misleading.

(4) Therefore, Modal Security is true but unexplanatory and misleading.

In our response, we argue that Modal Security should be interpreted in such a way that (2) is false. But we also note that (1) and (3) can be resisted.

According to the interpretation we suggest, we must distinguish between direct and indirect reasons for belief. While hard to define precisely, the distinction can be illustrated by using paradigmatic examples. Suppose that Abe, an otherwise trustworthy source, informs you that P and you thereby form the belief that P. How can the justification of this belief be defeated? One way would be if Ben, a source you know to be reliable, informs you that ¬P. In that case, you get direct evidence against P. Another way to defeat the belief is by repudiating Abe's reliability. In this case, Ben doesn't tell you ¬P, he only tells you that Abe is unreliable. That is indirect evidence against P, and it can be cashed out in terms of safety or sensitivity. Ben's

claim that Abe is unreliable implies that forming a belief according to Abe's testimony is unsafe or insensitive.[25]

This pair of examples is also a paradigmatic illustration of the distinction between rebutting and undermining defeaters. And, indeed, it is natural to suppose that the distinction between rebutting and undermining defeaters should be cashed out using the direct / indirect distinction.[26] Here is a preliminary thought. Rebutting defeaters are direct reasons to doubt that the belief is true, while undermining defeaters are direct reasons to doubt that the belief is safe or sensitive. Note that undermining defeaters can still be defeaters according to the accuracy based view. This would be the case if they are also *indirect* reasons to doubt that the belief is true. Modal Security could, then, be formulated more precisely as follows.

Modal Security$_2$: If evidence, E, *undermines* our belief that P, then E gives us direct reason to doubt that our belief is sensitive or safe.

Modal Security$_2$ escapes the theoretical worry of this section. It is no longer implied by the accuracy based conception of defeaters. The accuracy based view of defeaters requires that defeaters be reasons to doubt the belief, but they need not be *direct* reasons to doubt the belief. According to this interpretation, when the term "reason" appears in the definition of Modal Security, it means direct reason.

Nevertheless, Modal Security$_2$, faces other problems, as we will now demonstrate.

---

[25] One possibility is to analyze unreliability in terms of a lack of correlation between the source and the truth across possible worlds. If so, unreliability just is lack of safety and sensitivity. However, even if you analyze unreliability differently, for instance as merely a lack of reason to trust the source, still that is arguably also a reason to doubt the safety and sensitivity of that source. [Acknowledgement omitted]

[26] We are not the first to make this suggestion. See Kotzen (2019, 213–214). Kotzen's work also reveals how difficult it is to account for this distinction in the Bayesian framework.

8. Argument 5: Against Modal Security$_2$

The first thing to note about Modal Security$_2$ is that no positive reason was provided in Section 2 to accept it. Indeed, our argument in that section systematically obscured the distinction between direct and indirect evidence. Why, after all, couldn't E undermine our belief that P by giving us *indirect* reason to doubt that our belief is safe and sensitive?

Take the example used above to demonstrate the direct / indirect distinction. We can use this case (and any other) to construct an example in which there's an *indirect* reason to doubt safety. Suppose that, to begin with, your belief that Abe is reliable was fully based on the testimony of Ben. (Assume that you would have otherwise had most reason to be indifferent as to whether Abe is trustworthy). Now suppose you receive a new testimony from Cas, who's otherwise justifiably considered trustworthy, that Ben is unreliable. Now that is not direct reason to doubt that Abe's testimony is safe or sensitive. If it were, then the case in which Ben testifies that Abe is unreliable would not count as a paradigmatic example of an indirect reason. For the same reason, however, it cannot be a direct reason to doubt that P is true. But it is a defeater of Bp just as much as the case in which Ben testified that Abe is unreliable. If so, however, then Modal Security$_2$ must be false.

(1) If Modal Security$_2$ is true, then an indirect reason to doubt that a belief is safe or sensitive can't be undermining.

(2) An indirect reason to doubt that a belief is safe or sensitive can be undermining.

(3) Therefore, Modal Security$_2$ is not true.

We believe this argument is sound and the example refutes Modal Security$_2$. However we can modify Modal Security such that it retains the original thought, retains the distinction

between direct and indirect reasons (and theoretical import), and handles the problem of indirect reasons for believing that a belief is unsafe or insensitive. Here's how:

Modal Security$_3$: If evidence, E, *undermines* our belief that P, then E gives us direct reason to doubt that our belief is sensitive or safe *or* E undermines the belief that <the belief that P is safe and sensitive>.

Modal Security$_3$ takes the form of a recursive definition. It says that in order for evidence, E, to undermine our belief that P, there must be a chain of undermining evidence which bottoms out in a direct reason to doubt safety or sensitivity. Here's how the modified version of the principle handles the counterexample. Cas's testimony *is* a direct reason to believe that Ben is unreliable. That is a direct reason to believe that any belief based on Ben's testimony is unsafe or insensitive. The belief formed on the basis of Ben's testimony was that Abe is reliable. Hence, Cas's testimony undermines the belief that <the belief based on Abe's testimony is safe and sensitive>. So, it undermines the belief based on Abe's testimony.

To be sure, Modal Security$_3$ still depends on the viability of the direct / indirect distinction. And one could reasonably doubt that this is at all determinate.[27] But this is a problem for

---

[27] One reason is the following. Consider a paradigmatic pair of cases that exemplify the direct / indirect distinction. Suppose I have witnesses who testify that they were at the scene of the crime, in Tiberias, and saw you point your gun at me and then saw you taking my money. You have two ways of defeating their testimony: One way is to prove that you are not the robber, say, if you prove that you were in Sepphoris while the crime took place in Tiberias. This would be an example of rebutting defeat, understood as a direct reason to believe that you are innocent. Alternatively, you can prove that the witnesses could not have seen the crime, say, if you prove that at the time of the crime the witnesses were in Sepphoris while the crime took place in Tiberias. This latter possibility would be an instance of undermining defeat, understood as an indirect reason to believe that you are innocent. A paradigmatic feature of such reasons is that it tends to lead to neutrality rather than to a contrary belief. The judge now has no reason to believe that you are guilty, but neither does she have reason to believe that you are innocent. So far so good.

But consider again the evidence that you were in Sepphoris while the crime was in Tiberias. Is that direct evidence that you did not commit the crime? Not quite. At the very least, not if direct evidence is supposed to be evidence for P without relying on an inference involving additional premises. In this case, we have to rely on our background knowledge that the crime was in Tiberias and that it would have been impossible for you to commit the crime if you were in Sepphoris at the same time, and that the evidence (say, another pair of witnesses) are reliable with regards to your location at a time. Now you may think in response that a reason can

anyone who wishes to countenance the common-sense contrast between undermining from rebutting evidence. We have a pair of distinctions — the rebutting/undermining and the direct/indirect distinction — which seem to stand or fall together. Of course, nothing that we have said shows that these distinctions are genuine. We merely contend that *for those of us who accept these distinctions*, the redundancy worry for Modal Security is not compelling.

9. Argument 6: From Justification

Suppose that we are right, and that none of the above objections to Modal Security is decisive. There is still a worry. The worry is that Modal Security seems to imply an asymmetry between epistemic justification and defeat. It makes sense that what counts as defeat, and in particular undermining, is closely related to what it is that makes a belief justified. Now, when it comes to the justification of a belief, it seems insufficient that you have no reason to doubt that a certain belief is unsafe or insensitive. Suppose I believe some mathematical claim because a website says that it is true. Suppose further that I have no reason to believe that this method is unsafe or insensitive. At the same time, suppose I have no positive reason to believe that this method is safe or sensitive. Is my belief justified? We think it is not. Now, if the lack of reason to question safety or sensitivity isn't enough to justify a belief, then something more is needed for justification. And if something more is needed, then it would seem that, contrary to Modal Security, that *something more* can be lost

---

be direct even if it requires auxiliary premises to support the relevant inference. That seems fine. But the question then arises why in the second paradigmatic example the evidence is considered indirect. The fact that we have no reliable witnesses for your involvement in the crime is a reason to believe that you were innocent. After all, most people are innocent of this crime. So, the 'undermining' evidence, together with some auxiliary premises and inferences, is also "direct" evidence that you have not committed the crime, if the addition of auxiliary premises does not demote the evidence from counting as direct. If there's any difference between the paradigmatic cases in terms of directness of the evidence, it seems more a difference in degree than a sharp difference — and a difference that's liable to be indeterminate at that. If it's only a difference in degree, then whether some evidence satisfies the third reading of Modal Security will sometimes be indeterminate.

without getting a reason to doubt safety or sensitivity. We can summarize the argument as follows:

(1) A lack of reason to doubt that a belief is safe or sensitive is insufficient for epistemic justification of the belief.

(2) **If** a lack of reason to doubt that a belief is safe or sensitive is insufficient for epistemic justification of the belief **then** epistemic justification can be defeated without a reason to doubt the safety or sensitivity of the belief.

(3) **If** epistemic justification can be defeated without a reason to doubt the safety or sensitivity of the belief **then** Modal Security is false.

(4) Therefore, Modal Security is false.

Since premise (3) is a straightforward implication of the definition of Modal Security, an advocate of Modal Security must reject either premise (1) or (2). Indeed, we will argue against both (1) and (2).

Premise (1) was supported by the website example. However, the website example is misleading. When we imagine such an agent who *unjustifiably* believes a mathematical claim on the ground that some website says that it is true, we normally imagine an agent who has a reason to believe that her belief is unsafe. We imagine an agent consulting a sham website that is widely known to be a sham. Believing things on the ground that a sham website says that they are true is clearly an unsafe method of belief formation. And it is difficult to imagine a similar case in which the agent *lacks* reason to doubt the safety or sensitivity of the belief in question. For instance, if we were to instead imagine an agent who believes a mathematical conjecture on the ground that he dreamed that it was true, then, again, he would

have reason to doubt the safety of his belief insofar as he knows that believing things based on dreams is an unsafe method. Without any such case, premise (1) is unsubstantiated.[28]

Moreover, premise (2) can be contested. Even if epistemic justification can be lost without having a reason to doubt its safety or sensitivity, it does not follow that the justification was undermined, or even defeated. Imagine that some version of dogmatism is true. That is, if it seems to S that P, that gives S defeasible justification to believe P. Now suppose it just stops seeming to S that P. At that point, S has lost his defeasible justification, and if there are not other considerations supporting P, S should stop believing that P. But this does not show that S's justification has been defeated. S has not — on any ordinary conception, at least — received new evidence. Rather, we have an exceptional case in which S *loses* evidence. It is similar to a case in which you forget a piece of evidence and therefore lose justification, without getting new evidence. Instead of receiving a defeater, you've lost a justifier. Modal Security says nothing about such a case.[29]

10. "Connection"

It might be thought that we have failed to do justice to the chief suspicion lurking in the background of objections. The worry is that Modal Security implies that one's beliefs can be secure from undermining, even upon learning that they bear no connection to the truth. This is just counterintuitive to many people. Korman & Locke (forthcoming), Lutz (forthcoming)

---

[28] To be clear, the suggestion here is not that epistemic justification *just is* lack of reason to doubt safety or sensitivity. We do not propose an analysis of epistemic justification. The denial of premise (1) entails only a sufficient condition for epistemic justification.

[29] An editor pointed out that one might receive misleading evidence that it does not seem to her that P, even though it actually does, and that such evidence could undermine the justification provided by the seeming. However, we believe that this is an impossible scenario. Even if it is possible to receive misleading evidence that it does not seem to one that P which is not evidence that her belief that P is unsafe or insensitive (this is questionable), if it actually does seem to her that P, we believe that such evidence would not be undermining, for reasons akin to those discussed in Sections 4 and 5. We think that while she might believe that her belief is undermined, so long as she hasn't actually lost the seeming, evidence that she has is not undermining.

and Faraci (2019), for instance, have each suggested that if one learns that there is no "explanatory" connection between the belief and the truth, then that *in itself* is undermining — regardless of the modal security of the belief. We have dealt with thought experiments of the type they use to motivate their view in Section 6. But some will complain that they just have the intuition that learning of a lack of explanatory connection is undermining in itself.

Bare intuitions are difficult to argue with. But we suggest that the intuitions at issue are misleading. In typical cases, a lack of explanatory connection is evidence of a lack of *causal* connection with a causally efficacious fact — and this is a reason to doubt that the belief is safe and sensitive. Therefore, in the cases we ordinarily imagine, a reason to believe that there is no explanatory connection between the belief and the truth is undermining, but not in a way that proves Modal Security wrong. For example, if one were to learn that there is no explanatory connection between one's perceptual belief that there is water in one's glass and the fact that there is (suppose the belief is the result of an hallucination of a glass, for instance), then this would indeed seem to undermine one's belief. *But the reason why* is that, in the case of glasses of water, hallucinations and most other belief forming methods that do not involve a causal connection to the fact, are not safe or sensitive belief forming methods. We suggest that intuitions about the need for an explanatory connection are shaped by typical cases.

To bolster our suggested analysis of these intuitions, consider: *why* is it that evidence of a lack of explanatory connection between the belief and the truth undermines the justification of the belief in the typical cases that we imagine? An attractive answer is that, if there is no such connection between the truth and the belief, then the belief is formed very unreliably. It matters whether the truth that there is water in one's glass is implied by some explanation of

our coming to believe this because, *in such a case*, this is *predictive* of epistemically valuable property – namely, reliability. What are the chances of a belief formed without an explanatory connection to the truth to be true? If you accept this sort of response, however, then the response can actually be seen as confirming Modal Security rather than disconfirming it. A lack of modal security is one very natural way to analyze unreliability.

Importantly, explanatory connection and modal security come apart in the cases of primary interest to debunkers — namely, truths that would be causally inert. There is a temptation to see hyperintensional explanatory connections of the sort Harman's Thesis targets as "just like" causation, except that they are applicable to necessary truths as well. But this is a serious mistake. On the one hand, that there is such a connection is no evidence for sensitivity and safety in the case of causally inert truths. One's (defeasibly justified) belief in any given logical truth will be insensitive and unsafe despite the fact that, as a logical truth, it is implied by every explanation at all — and so certainly by the explanation of your coming to believe it — if you formed it by consulting an authority who decided on its truth by flipping a coin, for instance. In the other direction, we mentioned in Section 3 that beliefs in causally inefficacious supervenient truths generally have the modal security of beliefs in a causally efficacious subvenient truths, because both co-vary with the causally efficacious facts that shaped our beliefs (see Clarke-Doane (2015) and (2016b) for details). So, the cases of primary concern *vis a vis* debunking arguments are *exactly* the cases where the intuition that connection matters comes apart from the intuition that reliability (or modal security) does. Rough and ready appeals to lack of "connection" obscure this essential complication.

We do not deny any role for hyperintensional "connections" between our beliefs their contents. It may be that in order to explain how, e.g., our moral beliefs *got to have the content*

they have we must cite such a connection. But, of course, if such a connection is required to give our beliefs content, then our moral beliefs are not undermined by Harman's Thesis *a fortiori* — since there is not actually anything that we believe. Modal Security speaks to how our moral and related beliefs might be undermined *assuming that we have (remotely determinate) such beliefs to start*.

In sum, we concede that there are notions of connection such that connection can come apart from modal security. What Modal Security denies is that such a notion of connection matters epistemically, in and of itself. In a slogan: it is epistemically preferable to have a safe and sensitive belief which fails to be "connected" to the truth, than to have an insafe and insensitive belief that is.

11. Conclusion

We have considered a range of objections to Modal Security. We have argued that, while some are substantial, and compel the defender of Modal Security to refine the principal's formulation, none refutes the basic idea. There are problems surrounding Modal Security. But the most serious ones point to problems for virtually all epistemologists — concerning the individuation of methods, and the distinction between direct and indirect evidence.

References

Adler, Jonathan, and Michael Levin. 2002. "Is the Generality Problem Too General?" *Philosophy and Phenomenological Research* 65 (1): 87–97. doi:10.1111/j.1933-1592.2002.tb00184.x.

Baras, Dan. unpublished ms. "Do Extraordinary Types Cry Out for Explanation?"

Barkhausen, Max. 2016. "Reductionist Moral Realism and the Contingency of Moral

Evolution." *Ethics* 126 (3): 662–89. https://doi.org/10.1086/684708.

Barnett, David. unpublished ms. "Higher-Order Evidence is the Wrong Kind of Reason".

Bergmann, Michael. 2006. *Justification without Awareness*. Oxford University Press. https://doi.org/10.1093/0199275742.001.0001.

Berker, Selim. 2018. "A Combinatorial Argument against Practical Reasons for Belief." *Analytic Philosophy* 59 (4): 427–70. https://doi.org/10.1111/phib.12140.

Berry, Sharon. Unpublished ms. "Coincidence Avoidance and Formulating The Access Problem," 1–24.

Bishop, Michael A. 2010. "Why the Generality Problem Is Everybody's Problem." *Philosophical Studies* 151 (2): 285–98. doi:10.1007/s11098-009-9445-z.

Bogardus, Tomas. 2016. "Only All Naturalists Should Worry About Only One Evolutionary Debunking Argument." *Ethics* 126 (3): 636–61. https://doi.org/10.1086/684711.

Braddock, Matthew. 2017. "Debunking Arguments from Insensitivity." *International Journal for the Study of Skepticism* 7 (2): 91–113. https://doi.org/10.1163/22105700-20171195.

Casullo, Albert. 2018. "Pollock and Sturgeon on Defeaters." *Synthese* 195 (7): 2897–2906. https://doi.org/10.1007/s11229-016-1073-5.

Clarke-Doane, Justin. forthcoming *Morality and Mathematics*. Oxford: Oxford University Press.

———. 2015. "Justification and Explanation in Mathematics and Morality." In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 10:80–103. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198738695.003.0004.

———. 2016a. "What Is the Benacerraf Problem?" In *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity*, edited by Fabrice Pataut, 17–43. Dordrecht: Springer. doi:10.1007/978-3-319-45980-6_2.

———. 2016b. "Debunking and Dispensability." In *Explanation in Ethics and Mathematics*, edited by Uri D. Leibowitz and Neil Sinclair, 23–36. Oxford University Press. doi:10.1093/acprof:oso/9780198778592.003.0002.

Conee, Earl, and Richard Feldman. 1998. "The Generality Problem for Reliabilism." *Philosophical Studies* 89 (1). Springer: 1–29. doi:10.1023/A:1004243308503.

Darwin, Charles. 1871. *The Descent of Man*. London: John Murray.

Dougherty, Trent, ed. 2011. *Evidentialism and Its Discontents*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199563500.001.0001.

Enoch, David, Levi Spectre, and Talia Fisher. 2012. "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge." *Philosophy & Public Affairs* 40 (3): 197–224. https://doi.org/10.1111/papa.12000.

Faraci, David. 2019. "Groundwork for an Explanationist Account of Epistemic Coincidence." *Philosophers ' Imprint* 19 (4): 1–26.

Faraci, David and Aaron Elliott. Unpublished ms. "Non-Naturalism and the 'Third-Factor' Gambit"

Fine, Kit. 2002. "The Varieties of Necessity." In *Conceivability and Possibility*, edited by John Hawthorne and Tamar Szabó-Gendler, 253–81. Oxford: Clarendon Press.

Haidt, Jonathan. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (4): 814–34. doi:10.1037//0033-295X.108.4.814.

Henderson, David, and Terence E. Horgan. 2001. "Practicing Safe Epistemology." *Philosophical Studies* 102 (3): 227–58. https://doi.org/10.1023/A:1010396124016.

Harman, Gilbert H. 1977. *The Nature of Morality*. Oxford: Oxford University Press.

Hawthorne, John, and Maria Lasonen-Aarnio. Unpublished ms.. "Not So Phenomenal!"

Jonas, Silvia. 2016. "Access Problems and Explanatory Overkill." *Philosophical Studies*, doi:10.1007/s11098-016-0807-z.

Joyce, James M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65 (4): 575–603.

Joyce, Richard. 2001. *The Myth of Morality*. Cambridge University Press. doi:10.1017/CBO9780511487101.

———. 2006. *The Evolution of Morality*. The MIT Press.

———. 2008. "What Neuroscience Can (and Cannot) Contribute to Metaethics." *Moral Psychology*.

———. 2016. "Evolution, Truth-Tracking, and Moral Skepticism." In *Essays in Moral Skepticism*, 142–58. Oxford University Press. doi:10.1093/acprof:oso/9780198754879.003.0008.

Kelly, Thomas. 2013. "How to Be an Epistemic Permissivist." In *Contemporary Debates in Epistemology*, 2nd ed.

Kelp, Christoph. 2018. *Good Thinking: A Knowledge First Virtue Epistemology*. New York: Routledge.

Klenk, Michael. Unpublished ms. "Modal Security and the Non-Modal Accidentality Problem for Moral Objectivism."

———. 2019. "Objectivist Conditions for Defeat and Evolutionary Debunking Arguments." *Ratio*, https://doi.org/10.1111/rati.12230.

Korman, Daniel Z, and Dustin Locke. forthcoming "Against Minimalist Responses to Moral Debunking Arguments." In *Oxford Studies in Metaethics*.

Kotzen, Matthew. 2019. "A Formal Account of Epistemic Defeat." In *Themes from Klein*, edited by Branden Fitelson, Rodrigo Borges, and Cherie Braden, 213–34. Cham: Springer. https://doi.org/10.1007/978-3-030-04522-7.

Nozick, Robert. 1981. *Philosophical Explanations*. Harvard University Press.

Pettigrew, Richard. 2016. *Accuracy and the Laws of Credence*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198732716.001.0001.

Pollock, John Leslie, and Joseph Cruz. 1999. *Contemporary Theories of Knowledge*. 2nd ed. Rowman & Littlefield.

Pryor, James. 2013. "Problems for Credulism." In *Seemings and Justification*, edited by Chris Tucker. Oxford: Oxford University Press.

Roland, Jeffrey, and Jon Cogburn. 2011. "Anti-Luck Epistemologies and Necessary Truths." *Philosophia* 39 (3): 547–61. doi:10.1007/s11406-010-9295-0.

Rosen, Gideon. forthcoming. "What Is Normative Necessity?" In *Metaphysics, Meaning and Modality: Themes from Kit Fine*, edited by M. Dumitru. Oxford University Press.

Ruse, Michael. 1986. *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*. Oxford: Basil Blackwell.

Schafer, Karl. 2017. "Review of: Explanation in Ethics and Mathematics : Debunking and Dispensability, Uri D. Leibowitz and Neil Sinclair (Eds.), Explanation in Ethics and Mathematics: Debunking and Dispensability, Oxford University Press, 2016." *Notre Dame Philosophical Review*, 257.

Salmon, Wesley C. 1998. *Causality and Explanation*. Oxford University Press. doi:10.1093/0195108647.001.0001.

Schechter, Joshua. 2018. "Is there a Reliability Challenge for Logic?" *Philosophical Issues*, September. doi:10.1111/phis.12128.

Schoenfield, Miriam. 2012. "Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief." *Noûs*. doi:10.1111/nous.12006.

———. 2018. "Permissivism and the Value of Rationality A Challenge to the Uniqueness Thesis." *Philosophy and Phenomenological Research*. https://doi.org/10.1111/phpr.12490.

Sinclair, Neil. 2018. "Belief Pills and the Possibility of Moral Epistemology." In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 13:98–118. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198823841.003.0005.

Sinnott-Armstrong, Walter. 2006. *Moral Skepticisms*. Oxford University Press. https://doi.org/10.1093/0195187725.001.0001.

Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (1): 109–66. doi:10.1007/s11098-005-1726-6.

———. 2008. "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth

Worrying About." *Philosophical Issues* 18 (1): 207–28.
https://doi.org/10.1111/j.1533-6077.2008.00145.x.

———. 2015. "Does Anything Really Matter or Did We Just Evolve to Think So?" In *The Norton Introduction to Philosophy*, edited by Gideon Rosen, Alex Byrne, Joshua Cohen, and Seana Valentine Shiffrin, 685–93. Norton.

Sturgeon, Scott. 2014. "Pollock on Defeasible Reasons." *Philosophical Studies* 169 (1): 105–18. doi:10.1007/s11098-012-9891-x.

Tersman, Folke. 2016. "Explaining the Reliability of Moral Beliefs." In *Explanation in Ethics and Mathematics*, 37–57. Oxford University Press. doi:10.1093/acprof:oso/9780198778592.003.0003.

Tucker, Chris, ed. 2013. *Seemings and Justification*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199899494.001.0001.

Vavova, Katia. 2014. "Debunking Evolutionary Debunking." In *Oxford Studies in Metaethics*, 9:76–101. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198709299.003.0004.

Warren, Jared, and Daniel Waxman. Unpublished ms. "Reliability , Explanation , and the Failure of Mathematical Realism."

Wedgwood, Ralph. 2002. "The Aim Of Belief." *Noûs* 36 (s16): 267–97. https://doi.org/10.1111/1468-0068.36.s16.10.

White, Roger. 2005. "Epistemic Permissiveness." *Philosophical Perspectives* 19 (1): 445–59. doi:10.1111/j.1520-8583.2005.00069.x.

———. 2010. "You Just Believe That Because..." *Philosophical Perspectives* 24 (1): 573–615. https://doi.org/10.1111/j.1520-8583.2010.00204.x.

Whiting, Daniel. 2018. "Knowledge, Justification, and (a Sort of) Safe Belief." *Synthese*. https://doi.org/10.1007/s11229-018-01905-z.

Williams, Bernard. (1973). Deciding to believe. In Bernard Williams (Ed.), *Problems of the Self* (pp. 136–51). Cambridge, MA: Cambridge University Press

Williamson, Timothy. 2002. *Knowledge and Its Limits*. Oxford University Press. doi:10.1093/019925656X.001.0001.

Woods, Jack. 2018. "Mathematics, Morality, and Self-Effacement." *Noûs* 52 (1): 47–68. doi:10.1111/nous.12157.