philosophies

MDPI

Article

The Safety Versus Sensitivity Debate Revisited

Duncan Pritchard

Department of Philosophy, University of California, Irvine, CA 92697, USA; dhpritch@uci.edu

Abstract

It is widely accepted that there is a modal condition on knowledge. There are two main candidates in the literature: *safety* and *sensitivity*. My goal in this paper is to revisit this debate in the context of *anti-risk epistemology*. This approach is meant to offer an independent way of determining the modal condition on knowledge (i.e., as opposed to simply pitching modal conditions against one another and seeing which one accommodates the wider range of cases). Interestingly, while anti-risk epistemology favors safety over traditional formulations of sensitivity, it also leads to a very specific way of formulating the safety condition on knowledge. Given that sensitivity might also be reformulated along anti-risk lines, a natural question to ask is where this leaves the debate between safety and sensitivity. I will be suggesting that the most plausible anti-risk formulation of sensitivity is equivalent to the corresponding formulation of safety. If that is right, then ultimately the moral to draw from anti-risk epistemology is not that it favors safety over sensitivity but that it leads to a formulation of the anti-risk condition on knowledge whereby this distinction collapses.

Keywords: anti-risk epistemology; epistemology; safety; sensitivity

1. Introduction

It is standardly held that if there is a modal condition on knowledge, then that condition is either to be understood in terms of *safety* or *sensitivity*[1]. The general tendency in this debate has been to pitch both conditions against one another to see which of them handles cases better. I have argued elsewhere, however, that a better approach is presented by what I used to call *anti-luck epistemology* but these days refer to as *anti-risk epistemology*. As we will see, this approach not only provides us with an independent rationale for treating safety as the anti-luck/risk condition on knowledge but also motivates a particular formulation of that principle.

My aim here is to revisit the debate between safety and sensitivity in the light of anti-risk epistemology. Given that anti-risk epistemology leads to a particular rendering of the safety condition on knowledge, how might this approach similarly refine our thinking about the sensitivity condition on knowledge? In particular, while anti-risk epistemology favors safety, so construed, over traditional formulations of sensitivity as the anti-risk condition on knowledge, what happens if we instead focus on an anti-risk formulation of sensitivity? My contention is that once safety and sensitivity are formulated in this way, then sensitivity does not any longer diverge from safety. If that is right, then anti-risk epistemology ultimately offers us a way of moving past the safety/sensitivity debate altogether.

Academic Editor: Joseph Ulatowski

Received: 16 June 2026
Revised: 15 July 2026
Accepted: 15 July 2026
Published: 17 July 2026

2. The Safety Versus Sensitivity Debate

Let us start off with the two principles in their most general form, which we will christen 'simple safety' and 'simple sensitivity'. In both cases, we are interested in a subject, *S*, who in the actual world truly believes that *p*:

Simple Safety

If *S* were to believe that *p*, then *p* would be true.

Simple Sensitivity

If *p* were not true, then *S* would not believe that *p*[2].

Both principles concern subjunctive conditionals[3]. In terms of the usual possible worlds semantics that are applied to such conditionals, we can recast these principles as follows[4]. Simple safety states that insofar as the subject believes that *p* in close possible worlds, then *p* is true, while simple sensitivity states that in the closest possible worlds in which *p* is false, the subject does not believe that *p*. In both cases, we are thus capturing a particular modal profile that applies to the subject's true belief across a particular set of possible worlds.

The corresponding indicative conditionals to simple safety and simple sensitivity would contrapose, but as is widely noted, subjunctive conditionals do not contrapose. Given the rendering of the particular pairing of subjunctive conditionals we have just offered for simple safety and simple sensitivity, we can easily see why. In short, this is because each formulation might concern a different set of possible worlds.

Consider first simple safety. Since one believes that *p* in the actual world, the closest possible worlds where one believes that *p* will obviously be possible worlds that are very similar to, and hence close to, the actual world. Simple safety is thus inherently concerned with nearby possible worlds, what is known in the literature as the 'modal neighborhood'.

In contrast, to evaluate whether a subject's belief satisfies simple sensitivity, we need to consider the closest possible world where what is believed is no longer true. Crucially, there is no inherent reason why such a possible world should be in the modal neighborhood. Some truths are *modally robust*, such that they are only false in far-off possible worlds. For example, there is no close possible world where the sun did not rise this morning. Accordingly, evaluating whether a true belief in this proposition is simple sensitive would require us to consider the far-off possible world where the sun did not rise. Given that different possible worlds could be relevant to the evaluation of these principles, it should be no surprise that they do not contrapose. A belief might satisfy one of these principles but not the other.

That the two principles do not contrapose entails that they are distinct epistemic principles. Nonetheless, it is uncontentious that, in general at least, when a subject's belief is simple safe, it is also simple sensitive, and vice versa. Consider a standard case of perceptual knowledge. Using their reliable perceptual abilities in appropriate conditions, a subject forms the true belief that the traffic light is green. The belief is both simple safe and simple sensitive. In the closest possible worlds where the subject continues to believe that the traffic light is green, it will continue to be green, and in the closest possible worlds where the traffic light is not green, but red or amber, the subject will not believe that it is green (but will rather believe that it is red or amber). Most ordinary cases of knowledge involve both simple safe and simple sensitive beliefs, which is why a *prima facie* case could be made for either principle being a necessary condition on knowledge.

Similarly, it is also uncontentious that, in general at least, when a subject's belief is simple *un*safe it is also simple *in*sensitive, and vice versa. Consider a standard Gettier-style case, such as the sheep example[5]. A farmer looks into his field and sees what appears to be a sheep there.

Accordingly, he forms the belief that there is a sheep in the field. His belief is true, but not because he is looking at a genuine sheep. In fact, what he is looking at is merely a sheep-shaped object, such as big hairy dog or a large boulder. Hidden from view behind the sheep-shaped object, however, is a real sheep. The farmer's belief is both true and, on standard accounts of justification anyway, justified, as he has good reasons for thinking that what he is looking at is a sheep, and yet it does not amount to knowledge because it is just a matter of luck that his belief is true. Cases like this thus demonstrate that justified true belief does not suffice for knowledge[6].

Notice that the farmer's belief is both simple unsafe and simple insensitive. Given that it is just a happenstance that there is a genuine sheep in the field, there are close possible worlds where the farmer would continue to believe that there is a sheep in the field (because the sheep-shaped object is still present) where the belief is no longer true (because the sheep has wandered out of the field). It is thus simple unsafe. Similarly, in the closest possible worlds where the farmer's belief is false—i.e., where the sheep has wandered out of the field, but everything else stays the same—the farmer would continue to believe that there is a sheep in the field because he would still be looking at the same sheep-shaped object. It is thus simple insensitive. In general, the lucky true beliefs in standard Gettier-style cases are both simple unsafe and insensitive. This means that both principles are in the running to be an anti-Gettier principle on knowledge that would explain why knowledge is lacking in these cases.

There are two key examples in the literature that are thought to illustrate that simple safety and simple sensitivity come apart in ways that bear on their potential as conditions on knowledge. The first concerns the denials of radical skeptical hypotheses, such as that one is a brain-in-a-vat (BIV) being 'fed' deceptive experiences by supercomputers. What is significant about such a hypothesis for our purposes is that its negation is held to take us to far-off possible worlds. Its truth is thus modally robust in the sense outlined above. At least if the world is roughly the way that we take it to be, then there are no close possible worlds where one is a BIV.

Consider now one's true belief that one is not a BIV. This belief is simple safe. If the possible world where one is a BIV is far-off, then inevitably in all close possible worlds where one continues to believe that one is not a BIV, this belief will continue to be true. Yet, this belief also seems to be simple insensitive. The closest not-*p* possible world—i.e., the far-off possible world where one is a BIV—is precisely a world where one continues to believe that one is not a BIV (as one is now the victim of the undetectable deception undertaken by the supercomputers).

Of course, that one's belief that one is not a BIV is simple safe does not suffice to make it knowledge, given that safety is, at most, a necessary condition on knowledge. Nonetheless, it at least leaves it open that anti-skeptical propositions like this might be known[7]. In contrast, if sensitivity is a condition on knowledge, then that such propositions are simple insensitive would entail that they cannot amount to knowledge. Potentially, at least, this could represent a significant concession to radical skepticism. Alternatively, if one is already convinced by the thought that one cannot know the denials of radical skeptical hypotheses, then one may take this to be grounds for thinking that simple sensitivity is a necessary condition on knowledge[8].

A second case where simple sensitivity and simple safety are held to come apart concerns our inductive beliefs. Consider the famous 'rubbish chute' case[9]. A subject who lives in a high-rise condominium puts his trash into the rubbish chute. There is nothing to indicate that the chute is malfunctioning, and it is usually reliable. Accordingly, the subject believes that his trash is now in the basement. The basis for this belief is inductive, given that the subject has not seen the trash in the basement himself but has rather inferred this

belief from his past experiences. Nonetheless, this looks like an excellent inductive basis for belief. Indeed, we would usually attribute the subject with knowledge that his trash is in the basement.

Interestingly, however, while the subject's belief is simple safe it is not simple sensitive. Given how we have described the case, we would expect the trash to be in the basement not just in the actual world but also in all close possible worlds too. Accordingly, in the closest possible worlds where the subject continues to believe that the trash is in the basement, this belief will continue to be true. This case of inductive knowledge is thus not in tension with simple safety as a condition on knowledge.

In contrast, this belief appears to be simple insensitive. The closest possible world in which the rubbish is not in the basement is presumably the non-close possible world where something happens to the chute during the trash bag's descent to prevent it from reaching the basement. For example, a workman on a lower level closes the chute in order to carry out repairs. In this possible world, however, the subject would continue to believe the target proposition regardless on the same inductive grounds. The belief is thus simple insensitive and hence, insofar as simple sensitivity is a condition on knowledge, in tension with this case of inductive knowledge.

The general problem posed by inductive knowledge for simple sensitivity arises because although the target proposition becomes false, the inductive basis for the belief is unchanged. This is possible even if the inductive basis is excellent. In contrast, if the belief were based on reliable perception, and conditions are otherwise normal, then one would expect that as the target proposition becomes false so the perceptual basis for the belief would change accordingly and hence the subject would no longer believe it. If the subject's basis for believing that the trash is in the basement is seeing it there, for example, then in the closest possible world where it is not in the basement, he would not believe that it was because he would be able to see that it is not there. The upshot is that there are cases of inductive knowledge that pose a challenge to simple sensitivity being a necessary condition on knowledge.

3. Basis-Relativity

We have reviewed the main contours of the debate between simple safety and simple sensitivity as necessary conditions on knowledge. Interestingly, however, these formulations are not credible ways of understanding these principles. Moreover, adding in the necessary detail has an impact on how we should think about the debate between these two principles.

Let us start with an important qualification to both principles, which is that they need to be understood in a basis-relative fashion. In our formulations of simple safety and simple sensitivity above, we are effectively asking whether the actual belief formed satisfies these principles, but this is not quite right. What is actually being treated as safe or sensitive is the *basis* for that belief, the conditions that, in fact, gave rise to that belief. What we are interested in is whether that basis for belief, which in the actual world led to a belief that p, is safe or sensitive. We thus get the following basis-relative versions of safety and sensitivity, where the basis in question is the actual basis on which the subject formed their belief that p:

Basis-Relative Safety

If S were to believe that p on the same basis, then p would be true.

Basis-Relative Sensitivity

If p were not true, then S would not believe that p on the same basis.

We can illustrate the importance of basis-relativity via the following example[10]. Consider a grandmother who has a highly reliable ability, as many grandmothers do, to tell

whether her grandchild—her grandson, say—is unwell insofar as she is in close proximity to him. In the actual world, after getting a good look at him, she forms the true belief that her grandson is unwell. Imagine, however, that if her grandson were unwell, then her children, not wishing to worry her, would keep him away from her. Moreover, they would falsely tell the grandmother that her grandson is well, and she would believe them. Whether we understand safety or sensitivity in a basis-relative way has an important bearing on whether we treat the grandmother as forming a safe or sensitive belief.

Consider simple sensitivity first, which is not construed in a basis-relative way. In the closest possible world where what the grandmother believes is false and her grandson is unwell, her children will lie to her about him being well, and she will believe them. Accordingly, she will continue to believe that her grandson is well even though this is no longer the case. Her belief thus fails simple sensitivity.

Once we construe sensitivity in a basis-relative way, however, then this changes the verdict. Our interest now is whether, in the closest possible world where what is believed is no longer true, the grandmother continues to believe it *on the same basis as in the actual world*. So construed, her belief is sensitive, as although she continues to believe the target proposition (now false) in this possible world, she does not believe it on the same basis as in the actual world. In particular, rather than believing it on the basis of her inspection of her grandson at close quarters, she is instead forming this belief on the basis of the (false) testimony from her children.

We can make similar points about safety. Consider simple safety, which is not basis-relative. There are close possible worlds where the grandmother continues to believe that her grandson is well where this is no longer true—i.e., those close possible worlds where the children falsely tell the grandmother that her grandson is well while keeping him out of sight. The grandmother's belief thus fails simple safety. Once safety is construed in a basis-relative way, however, then our verdict changes. Given the reliability of her actual basis for forming her belief, then in all close possible worlds where she continues to believe that her grandson is well *on this same basis*, her belief continues to be true.

In both cases, the basis-relative version of safety and sensitivity seems to be capturing what we are interested in when we ascribe knowledge to subjects. That is, we are specifically interested in whether the particular basis that the subject used in forming their beliefs generates the target modal profile. Cases where exceptions arise to this modal profile due to a switch in the basis employed do not seem relevant. Modifying safety and sensitivity along basis-relative lines thus seems independently plausible, in that it captures our intuitions about what it is that we are trying to evaluate with these principles, while also delivering versions of these principles that can accommodate a wider range of our epistemic intuitions.

Interestingly, opting for basis-relative formulations of these principles may remove one of the key divergences between them that we highlighted above. We noted that that while one's beliefs in the denials of radical skeptical hypotheses, such as the BIV hypothesis, will satisfy simple safety, they will not satisfy simple sensitivity. In the case of safety, this is a direct result of the fact that the believed proposition is modally robust, in that it is only false in far-off possible worlds. Accordingly, such a belief would be not only simple safe but also basis-relative safe too.

Consider now whether this belief satisfies basis-relative sensitivity. It is actually unclear how we should understand the basis for one's belief that one is not the victim of a radical skeptical scenario, like the BIV scenario[11]. The crux of the matter, however, is that once we start to consider possible worlds that are radically different from the actual world, then we cannot take it as given that the same basis for belief is even available. For example, suppose we consider the basis for one's belief that one is not a BIV to be perceptual. In that case, however, then the closest possible world where what is believed is false—the

BIV world—it is *not* true that one believes that one is not a BIV on the same basis as in the actual world. Remember that a BIV's experiences are not perceptual, but rather artificially generated by the supercomputers that the BIV is hooked up to. Accordingly, if that is the right way to think about the basis in this case, then one's belief that one is not a BIV can satisfy basis-relative sensitivity[12]. One potential difference between safety and sensitivity might thus disappear once we adopt basis-relative versions of these principles.

Once we opt for basis-relativity, then this naturally leads to a further revision of how we should formulate these modal principles. For if what matters to an evaluation of safety or sensitivity is the actual basis for one's belief, then what is the motivation for fixating on the fact that this basis resulted in a belief *that p* in the actual world? Why not instead consider a range of potential outputs that this basis could have resulted in, rather than focusing on the one output from that basis that actually occurred?

The distinction in play here may initially seem somewhat obscure, but we can bring its theoretical importance into sharp relief by considering its relevance for safety. Consider the following refinement of basis-relative safety:

Basis-Relative Safety II

If *S* were to form a belief on the same basis, then that belief would be true.

Notice that this formulation of safety does not focus on the proposition actually believed, but rather on the doxastic output of that same basis[13]. To illustrate the rationale behind this formulation, suppose that our subject's basis for belief is an instance of testimony from someone they believe to be highly reliable. In the actual world, the testimony concerns a true proposition that is modally robust, such as that the sun rose this morning. If our assessment of safety involves focusing on the actual proposition believed, then this basis for belief is inevitably safe. Since there is no close possible world where the target proposition is false, there is obviously no close possible world where the subject continues to believe this proposition but believes falsely. The problem, of course, is that this tells us nothing about the epistemic pedigree of the basis for belief in play. The safety of the basis is instead following from the fact that the truth of the proposition actually believed is modally robust. (The same goes for other beliefs in modally robust propositions, such as the denials of radical skeptical hypotheses that we have just discussed.)

This same basis for belief could well have resulted in different doxastic outputs, however. Let us imagine two extremes in this regard. In the one scenario, the informant who is believed to be highly reliable is indeed highly reliable. In fact, of the range of instances of testimony they could have provided in this case, all of them are true. In the other scenario, the subject is completely misled about the reliability of the informant, who is in fact making up their answers. In both scenarios, the informant gives the modally robust true answer. For each version, we can ask whether this basis is safe in the sense that it could have easily resulted in the subject forming a false belief on that basis.

Consider first the unreliable informant who is merely making up their answers. Although the answer they give in the actual world happens to not only be true but also modally robust, given that they are making up their answers, there will be close possible worlds where their answers are false. Accordingly, although there is no close possible world where the subject continues to believe on the same basis the same proposition as in the actual world and believes falsely, there are close possible worlds where she forms a belief on this basis and believes falsely. We can thus capture the sense in which this basis for belief is unsafe, even though it happens to lead to a modally robust true belief in the actual world.

Consider now the version of the scenario where the basis for belief is the testimony of the highly reliable informant. On the old formulation that fixated on the actual proposition believed, then the basis is safe by default, on account of how there is no close possible

world where this proposition is false. This formulation of safety thus does not discriminate between epistemically good and bad bases for belief. The new formulation fares much better. While the belief remains safe, this is not now simply a consequence of the truth of the actual proposition believed being modally robust. Instead, it will take into account the other propositions that could have been doxastic outputs of this basis in close possible worlds. Since they are all true, it remains the case that there is no close possible world where the subject continues to form a belief on the same basis and yet believes falsely as a result.

Again, we seem to have a principled way of understanding safety that generates a more plausible set of verdicts, at least insofar as safety is a necessary condition on knowledge. Since this way of understanding safety can ensure that our beliefs in modally robust truths are not automatically safe, it can also explain how safety might apply to necessary truths. These are truths that are not just true in all close possible worlds, but in all possible worlds *period*. Accordingly, it is often thought that safety cannot be employed to make sense of our knowledge of such truths. If there is no possible world where they are false, then *a fortiori* there is no close possible world where they are false either, and hence we know in advance that there cannot be a close possible world where the subject continues to form a belief in a necessary truth on the same basis and yet believes falsely.

On the more refined formulation of safety just offered, however, a basis for belief is not automatically safe just because it results in a necessary truth in the actual world. Consider, for example, a subject using a calculator in order to arrive at arithmetical beliefs, where they would accept whatever answer the calculator would generate. As before, we can imagine two versions of this scenario, one where the calculator is highly reliable and one where it is broken and generating random answers. In both cases in the actual world, the calculator gives a correct answer, which, since it is mathematical truth, is necessarily true.

On the refined formulation of basis-relative safety, where the basis for belief is the malfunctioning calculator, it is unsafe. While there is (obviously) no close possible world where what is actually believed is false, there are close possible worlds where this basis for belief results in falsehoods, and of course in those worlds the subject will continue to form their belief on this basis. In contrast, where the basis for belief is the properly functioning calculator, it is safe. Significantly, however, this is not merely a result of the actual belief formed being necessarily true. For what matters now is that there are close possible worlds where the same basis results in different doxastic outputs that are also true. In short, this is a basis for belief that could not have easily led to the subject forming a false belief[14].

We noted earlier that sensitivity is also understood along basis-relative lines in the contemporary literature. Interestingly, however, the same move away from focusing on the specific proposition believed is not made, even though, as we just noted with safety, it seems to be a natural consequence of opting for a basis-relative formulation. Such a refinement would also generate similar advantages. It is hard to see how sensitivity could be applied to necessary truths, for example, at least unless one wished to argue that all beliefs in necessary truths are automatically sensitive regardless of the epistemic pedigree of the basis employed, on the grounds that there can be no not-*p* possible world that could be plugged into the antecedent of the sensitivity counterfactual conditional.

Here is a formulation of basis-relative sensitivity that is not output-specific[15]. First, we specify the range of propositions that the basis could generate beliefs in. We then consider whether, for each of these propositions, if it were false, the subject would not believe it on this basis. This formulation would deal with the problem of beliefs in necessary truths being sensitive by default, regardless of their epistemic pedigree. Consider again the case of the malfunctioning calculator that happens to produce a belief in a necessary truth. This same basis could generate beliefs in falsehoods (indeed, necessary falsehoods). Moreover,

in cases where it does, the subject would still believe it on the same basis, given that they are unaware of the fact that the calculator is malfunctioning. The belief would thus fail sensitivity on this formulation, which is what we want.

Notice, however, that this formulation still faces the problem of inductive knowledge. Consider the rubbish-chute example again, where the subject believes that the rubbish is in the basement on an inductive basis. Since the subject actually believes this proposition, then clearly the proposition will fall within the target range of propositions that the basis could generate beliefs in. There is, however, a non-close possible world where this proposition is false and yet the subject continues to believe it on the same basis. This alerts us to the fact that even though basis-relativity restricts the scope of possible worlds that are relevant to assessments of sensitivity (as we saw in our discussion of the BIV case), the fact remains that non-close possible worlds can still fall within that scope. This is why although safety and sensitivity tend to converge once we opt for basis-relative (and non-output specific) formulations, there is also still some divergence.

4. Anti-Risk Epistemology

In the last section, we considered the case for a basis-relative formulation of safety along with how this leads, in turn, to a further refinement of the principle so that it is no longer focused on the actual proposition believed. In this section, I want to step back and examine a general rationale that has been offered to guide our thinking about safety. This motivates the two refinements just noted, but it also motivates other refinements too. Crucially, this rationale is meant to provide an independent basis for preferring safety over traditional formulations of sensitivity, as opposed to the more standard approach of simply comparing how they each fare across a range of cases. As we will see, my ultimate interest is how this rationale plays out once we consider its implications for the sensitivity principle too.

The rationale I have in mind is what I have elsewhere called *anti-luck* or *anti-risk epistemology*[16]. Since anti-luck epistemology came first, let us start with that. The guiding idea behind this proposal is that a significant element of the project of defining knowledge is concerned with the elimination of a particular kind of epistemic luck. For example, Gettier cases, lottery cases, and radical skepticism all trade on the idea that knowledge excludes a kind of luck. Accordingly, part of the project of analyzing knowledge is to identify its anti-luck condition. Usually, the way this is done is by formulating epistemic conditions and testing them out against our intuitions. Anti-luck epistemology, in contrast, involves a different theoretical approach. First, we define luck. To this end, I have argued for a *modal account of luck*[17]. Next, we define the specific sense in which knowledge is incompatible with luck. This is what I have called *veritic luck*, or luck that one's belief is true, given how it was formed[18]. Putting these two elements together should then reveal the nature of the anti-luck condition on knowledge.

These days, I argue that we should replace anti-luck epistemology with anti-risk epistemology. The basic idea is the same, it is just that rather than having a theory of luck at the heart of the proposal, we instead have a theory of risk. Since risk and luck are such closely related notions, this new approach generates many of the same conclusions as anti-luck epistemology. Indeed, I have argued that luck and risk are both modal notions with a very similar profile. Where they differ, primarily, is in terms of perspective. Think about someone who survives a plane crash. Looking backwards, after the event, they are lucky to be alive. Looking forwards, before the event, they are at a high risk of dying.

If that is right, then we can just as well think of the anti-luck condition on knowledge in terms of an anti-risk condition. Indeed, arguably the reason we want to eliminate veritic

https://doi.org/10.3390/philosophies11040122

luck from knowledge is because veritically lucky beliefs are at a high risk of being false. That would suggest that it is risk that is playing the primary explanatory role here.

With the foregoing in mind, consider the *modal account of risk*[19]. Call the *risk event* the target unwanted outcome. According to the modal theory of risk, in determining the *level* of risk, we are interested in the modal closeness of the risk event. Imagine one is about to do a parachute jump, and one is assessing the level of risk that the parachute might not open. On the modal theory of risk, if there is a close possible world in which one undertakes this jump and the parachute does not open, then it is a high-risk activity. In contrast, if there no close possible world where this occurs, then it is a low-risk activity.

In order to see the significance of this way of thinking about risk, notice that it generates very different conclusions to the *probabilistic account of risk*[20]. On this theory of risk, what determines the level of risk is the likelihood of the risk event occurring. For example, if the probability of the chute not opening is high, then so is the level of risk involved, while if the probability is low, then the level of risk is low too. What could easily occur is not the same as what is likely to occur, however. There could be a flaw in the design of the parachute that means that it could easily fail to open even though there is a high likelihood of it opening. A parachute that opens 99% of the time could still be such that it could easily fail to open. On the probabilistic theory of risk, the level of risk is low, but on the modal theory of risk, the risk is high. I suggest that our intuitions in this regard favor the modal account[21].

We can apply the modal theory of risk to the epistemic case. This will be concerned with the veritic epistemic risk associated with forming beliefs on a particular basis, where the target epistemic risk event is this basis leading to error (i.e., false belief)[22]. So construed, a basis for belief is high-risk if it would generate false belief in close possible worlds. Conversely, a basis for belief is low-risk if there is no close possible world where it generates false belief. We thus get *anti-risk epistemology*. The anti-luck condition on knowledge turns out to also be—perhaps even more fundamentally be—an anti-risk condition. While knowledge excludes high levels of veritic epistemic risk, it is compatible with low levels of veritic epistemic risk.

The observant reader will have spotted that anti-risk epistemology seems to be straightforwardly leading to a version of the safety condition for knowledge. What is important for knowledge, from a modal point of view, is that one's true belief is formed on a basis that could not easily lead to a false belief. This is an important result in itself, as it means that there is a rationale for safety as the anti-luck/risk condition on knowledge that is independent of the usual motivations offered, whereby support for safety is marshalled by showing that it is better than alternatives (like sensitivity) at dealing with a range of cases. It is also significant that anti-risk epistemology does not just lead to the safety condition on knowledge but also gives us a way to understand that principle.

For example, we are led to a basis-relative formulation of safety not merely because it generates the right intuitive result across cases but because that is the formulation suggested by anti-risk epistemology. It is the manner in which the actual belief is formed that is being assessed for epistemic risk. Relatedly, the formulation we are led to is not fixated on the particular proposition that the subject happens to believe on that basis in the actual world. Although we can motivate this point by appeal to cases (as we did above), this aspect of the formulation flows directly from anti-risk epistemology. The concern is simply whether that basis could easily lead to error, since that would make it a high-risk basis, and not the artificially restricted concern that the basis could lead to error about the proposition actually believed.

There are several other advantages to motivating safety via anti-risk epistemology, but let me just focus on one more here[23]. One issue that has been raised about the formulation of safety concerns exactly how much error it allows within the general modal neighborhood.

Does it exclude all error, or only error in the closest possible worlds? The general worry is that the former might be too restrictive and the latter not restrictive enough, but on what principled basis is the defender of safety to draw the relevant line?[24]

Anti-risk epistemology is able to speak directly to this concern. Risk is a degree notion. The upshot of anti-risk epistemology is that, from the perspective of knowledge at any rate, the closer the relevant possible world where the error occurs, the greater weight it carries in terms of its potential to undermine knowledge. We are completely intolerant of high epistemic risk, which is knowledge-undermining, and completely tolerant of low epistemic risk, which can happily co-exist with knowledge. That means that knowledge is incompatible with error on the same basis in close possible worlds but compatible with error on the same basis in far-off possible worlds. Inevitably, there will be a penumbral range where the possible worlds are not especially close but also not especially far-off either. Anti-risk epistemology predicts that as the possible worlds get further out, so the strength of our intuition that knowledge is undermined begins to weaken, until we get to the point where our intuition is that the epistemic risk is so remote that it does not undermine it at all. The penumbral cases will thus be cases where the intuitions are not strong either way. We should thus expect conflicting intuitions in such cases as to whether the level of epistemic risk in play undermines knowledge. In this way, anti-risk epistemology can not only accommodate the penumbral range where it is unclear whether the level of epistemic risk is knowledge-undermining but can also explain why it arises, as it is a consequence of risk coming in degrees[25].

5. Revisiting the Safety–Sensitivity Debate

In the last section I showed how anti-risk epistemology can be used to motivate not just the safety condition on knowledge, but also to generate a specific formulation of this condition. Where does anti-risk epistemology leave the sensitivity principle?

We also saw in the last section that anti-risk epistemology motivates the idea that it is only close possible worlds where one's basis leads to error that involve the high level of epistemic risk that would be knowledge-undermining. If high levels of epistemic risk only apply when one's basis leads to false belief in close possible worlds, then that should have a bearing on how we think about sensitivity. Recall that when we formulated a basis-relative version of sensitivity above that was not output specific, we still ended up with a formulation that allowed non-close possible worlds to be relevant to sensitivity assessments. Suppose that we now, on anti-risk grounds, restrict the non-output specific basis-relativity formulation of sensitivity to close possible worlds?

Here is what such a formulation would look like. First, we specify the range of propositions that, *in close possible worlds*, the basis could generate beliefs in. We then consider whether, for each of these propositions, if it were false, the subject would not believe it on this basis. This formulation would retain the advantage of dealing with the problem of beliefs in necessary truths being sensitive by default, regardless of epistemic pedigree. Consider the malfunctioning calculator example again. The point is not merely that this basis for belief could generate beliefs in falsehoods, but that it could do so in close possible worlds.

Crucially, however, this formulation would also avoid the problem of inductive knowledge. As we have noted in our discussion of the rubbish-chute case, there is no close possible world where the believed proposition is false. There is thus no close possible world where the subject's inductive basis for belief leads to a false belief. It follows that the belief is sensitive on this formulation, at least in the sense that it cannot be insensitive.

Indeed, the sense in which the belief is sensitive parallels the sense in which it is safe (i.e., that it cannot be unsafe). This is no coincidence. It reflects the fact that this version

of sensitivity that is restricted to close possible worlds collapses into the safety principle. For what it demands is that there is no close possible world where one forms a belief on the same basis and believes falsely. However, that is just the safety principle. Bringing in anti-risk epistemology to our understanding of sensitivity thus leads to a formulation of this principle that no longer diverges from the safety principle.

The upshot is that once we follow-through on the methodology supplied by anti-risk epistemology, then we are able to dispense with the safety/sensitivity contrast altogether, at least insofar as that distinction is meant to capture competing accounts of the anti-risk condition on knowledge. It initially seemed as if this contrast survived the adoption of anti-risk epistemology, but that is because we were contrasting an anti-risk formulation of safety with traditional formulations of sensitivity. So long as we consistently apply anti-risk epistemology to both principles, then we end up with a formulation of sensitivity that converges on safety. We noted above that the general trend in modal epistemology has been towards convergence on this score, in that improved formulations of these principles are not so divergent. One could thus plausibly consider anti-risk epistemology the culmination of this process.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: I am grateful to two anonymous reviewers from *Philosophies* for detailed comments on a previous version of this paper. Thanks also to Lars Bo Gundersen. This paper was written while a Senior Research Associate of the *African Centre for Epistemology and Philosophy of Science* at the University of Johannesburg.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1 In order to keep our discussion manageable, in what follows I will take it as given that there is a modal condition on knowledge. There might be reasons to reject this presupposition, for example if one is suspicious of modal principles in general or if one rejects the anti-luck intuitions that usually motivate modal conditions on knowledge. For an example of the latter view, see Hetherington [1,2]. For a response, see Prichard [3,4]. See also Olsson [5], who presents the case for a pure reliabilist theory of knowledge that rejects Gettier-style counterexamples.

2 For some of the main defenses of the safety condition, see Sainsbury [6], Sosa [7,8], and Williamson [9]. See also Luper [10,11]. For a survey of contemporary work on the safety condition, see Rabinowitz [12]. As regards sensitivity, the *locus classici* are Dretske [13] and Nozick [14], ch. 3. See also Dretske [15]. For a prominent contemporary defense of the sensitivity condition, see Becker [16]. See also Black and Murphy [17]. For discussions of related proposals, see Black [18,19], Gundersen [20–22], Roush [23], Zalabardo [24], and Melchior [25]. For some useful surveys of the contemporary debate regarding safety and sensitivity, see Pritchard [26] and Black [27,28].

3 On some construals of these subjunctive conditionals, that *S* truly believes that *p* in the actual world is entailed (in which case this stipulation I just offered in this regard is redundant), but we do not need to take a stance on this issue here.

4 See, especially, Lewis [29,30] and Stalnaker [31].

5 Originally due to Chisholm [32].

6 The *locus classicus* in this regard is, of course, Gettier [33].

7 For an influential anti-skeptical defense of safety over sensitivity, see Sosa [7]. See also Pritchard [34,35].

8 Famously, this was the line taken by Dretske [13,15] and Nozick [14], ch. 3. They both avoided the potential radical skeptical consequences of this claim by denying the 'closure' principle that is usually used to extract radical skeptical conclusions from our inability to know the denials of radical skeptical hypotheses. For discussion of the closure principle and the idea that there might be a principled basis for rejecting it, see the exchange between Dretske [36,37] and Hawthorne [38].

9 Due to Sosa [7]. For an early statement of the problem posed by inductive knowledge for sensitivity, see Vogel [39].

[10] This example is originally due to Nozick [14], 179ff.

[11] Although it does not matter to our present discussion, my own view is that the reason why it is so difficult to identify a basis in this case is that they are hinge commitments, along the general lines set out in Wittgenstein [40]. See, especially, Pritchard [41], part 2.

[12] See Williams [42], ch. 8 for an early statement of this claim that our beliefs in the denials of radical skeptical scenarios may be sensitive provided that we opt for a basis-relative version of this principle. For a more recent development of this idea, see Black [18,19]. As Williams notes, although Nozick defends a basis-relative version of sensitivity, he misses this point because when it comes to the sensitivity of this particular anti-skeptical claim he interprets the notion of a basis in a purely internal fashion, such that so long as it seems to the subject like the same basis as the actual basis, then it is the same basis. That is not a credible way of thinking about bases, however, and in fact conflicts with Nozick's own rendering of this notion when he is discussing it independently of the problem of radical skepticism.

[13] This way of thinking about safety is sometimes referred to as 'global methods' safety. See, for example, Hirvelä [43] and Bernecker [44].

[14] For further discussion of this particular way of construing basis-relative safety so that it can handle modally robust truths, including necessary truths, see Pritchard [45,46]. For a related discussion, see Hirvelä [43].

[15] I am grateful to an anonymous reviewer for *Philosophies* for pressing me on this issue.

[16] For the main places where I develop anti-luck epistemology, see Pritchard [35,45–50] and Pritchard, Millar and Haddock [51], chs. 1-4. For the main places where I develop anti-risk epistemology, including how it diverges from anti-luck epistemology, see Pritchard [52–54], ch. 8.

[17] See, especially, Pritchard [35], *passim* and [55].

[18] For further discussion of veritic luck, see Pritchard [35], *passim*.

[19] For the development of the modal account of risk, along with an explanation of how luck and risk relate to one another, see Pritchard [54], *passim* and [56]. See also Navarro [57,58] for a related, but distinct, account of luck and risk and their epistemic import.

[20] This theory of risk is orthodoxy in the risk literature outside of philosophy. See Hansson [59].

[21] There is an extensive psychological literature on risk and luck that confirms this point. For a survey, see Pritchard and Smith [60]. See also Pritchard [54], *passim* and [56].

[22] At least, this is the epistemic risk event that is relevant for knowledge, which is why it is applicable to an anti-risk condition on knowledge. See, for example, Pritchard [52]. As argued in Pritchard [61], however, we can use this general anti-risk framework to develop different axes of epistemic risk where we vary the target epistemic risk event.

[23] For further discussion of the advantages of anti-risk epistemology, see Pritchard [52–54], ch. 8. In order to give a full account of the theoretical advantages of anti-risk epistemology, I would need to discuss how this proposal can account not only for the level of epistemic risk but also its *depth*. Very roughly, depth of risk relates not to the modal closeness of the risk event but to its preponderance across close possible worlds (i.e., the greater the preponderance, the greater the depth of risk). I discuss this distinction between level and depth of risk, and apply it to the epistemic case, in Pritchard [54], *passim*. For a related discussion that draws a similar distinction between level and depth of risk, see Hirvelä and Paterson [62].

[24] For an influential statement of this problem, see Greco [63].

[25] For more on this point, see Pritchard [52,53].

References

1. Hetherington, S. Knowledge Can Include Luck. In *Contemporary Debates in Epistemology*, 3rd ed.; Roeber, B., Steup, M., Turri, J., Sosa, E., Eds.; Blackwell: Oxford, UK, 2024; pp. 151–159.
2. Hetherington, S. On Whether Knowing Can Include Luck: Asking the Correct Question. In *Contemporary Debates in Epistemology*, 3rd ed.; Roeber, B., Steup, M., Turri, J., Sosa, E., Eds.; Blackwell: Oxford, UK, 2024; pp. 169–171.
3. Pritchard, D.H. There Cannot be Lucky Knowledge. In *Contemporary Debates in Epistemology*, 3rd ed.; Roeber, B., Steup, M., Turri, J., Sosa, E., Eds.; Blackwell: Oxford, UK, 2024; pp. 159–168.
4. Pritchard, D.H. Reply to Hetherington. In *Contemporary Debates in Epistemology*, 3rd ed.; Roeber, B., Steup, M., Turri, J., Sosa, E., Eds.; Blackwell: Oxford, UK, 2024; pp. 171–173.
5. Olsson, E.J. Gettier and the Method of Explication: A 60 Year Old Solution to a 50 Year Old Problem. *Philos. Stud.* **2015**, *172*, 57–72.
6. Sainsbury, R.M. Easy Possibilities. *Philos. Phenomenol. Res.* **1997**, *57*, 907–919. [CrossRef]
7. Sosa, E. How to Defeat Opposition to Moore. *Philos. Perspect.* **1999**, *13*, 141–154. [CrossRef]
8. Sosa, E. How Must Knowledge be Modally Related to What is Known? *Philos. Top.* **1999**, *26*, 373–384. [CrossRef]
9. Williamson, T. *Knowledge and Its Limits*; Oxford University Press: Oxford, UK, 2000.
10. Luper, S. The Epistemic Predicament. *Australas. J. Philos.* **1984**, *62*, 26–50. [CrossRef]

11. Luper, S. Indiscernability Skepticism. In *The Skeptics: Contemporary Essays*; Luper, S., Ed.; Ashgate: Aldershot, UK, 2003; pp. 183–202.
12. Rabinowitz, D. The Safety Condition on Knowledge. In *Internet Encyclopedia of Philosophy*; Dowden, B., Fieser, J., Eds.; 2010. Available online: https://iep.utm.edu/safety-c/ (accessed on 1 June 2026).
13. Dretske, F. Epistemic Operators. *J. Philos.* **1970**, *67*, 1007–1023. [CrossRef]
14. Nozick, R. *Philosophical Explanations*; Oxford University Press: Oxford, UK, 1981.
15. Dretske, F. Conclusive Reasons. *Australas. J. Philos.* **1971**, *49*, 1–22. [CrossRef]
16. Becker, K. *Epistemology Modalized*; Routledge: London, UK, 2007.
17. Black, T.; Murphy, P. In Defense of Sensitivity. *Synthese* **2007**, *154*, 53–57. [CrossRef]
18. Black, T. A Moorean Response to Brain-In-A-Vat Scepticism. *Australas. J. Philos.* **2002**, *80*, 148–163. [CrossRef]
19. Black, T. Defending a Sensitive Neo-Moorean Invariantism. In *New Waves in Epistemology*; Hendricks, V., Pritchard, D.H., Eds.; Palgrave Macmillan: London, UK, 2008; pp. 8–27.
20. Gundersen, L.B. *Dispositional Theories of Knowledge a Defence of Aetiological Foundationalism*; Routledge: London, UK, 2003.
21. Gundersen, L.B. Tracking, Epistemic Dispositions and the Conditional Analysis. *Erkenntnis* **2010**, *72*, 35364. [CrossRef]
22. Gundersen, L.B. Knowledge, Cognitive Dispositions and Conditionals. In *The Sensitivity Principle in Epistemology*; Becker, K., Black, T., Eds.; Cambridge University Press: Cambridge, UK, 2012; pp. 66–81.
23. Roush, S. *Tracking Truth: Knowledge, Evidence, and Science*; Oxford University Press: Oxford, UK, 2006.
24. Zalabardo, J. *Scepticism and Reliable Belief*; Oxford University Press: Oxford, UK, 2012.
25. Melchior, G. *Knowing and Checking: An Epistemological Investigation*; Routledge: London, UK, 2019.
26. Pritchard, D.H. Sensitivity, Safety, and Anti-Luck Epistemology. In *The Oxford Handbook of Scepticism*; Greco, J., Ed.; Oxford University Press: Oxford, UK, 2008; pp. 437–455.
27. Black, T. Modal Epistemology. In *Routledge Encyclopedia of Philosophy*; Routledge: London, UK, 2018. [CrossRef]
28. Black, T. Modal and Anti-Luck Epistemology. In *Routledge Companion to Epistemology*; Bernecker, S., Pritchard, D.H., Eds.; Routledge: London, UK, 2011; pp. 187–198.
29. Lewis, D. *Counterfactuals*; Blackwell: Oxford, UK, 1973.
30. Lewis, D. *On the Plurality of Worlds*; Blackwell: Oxford, UK, 1987.
31. Stalnaker, R. *Inquiry*; MIT Press: Cambridge, MA, USA, 1984.
32. Chisholm, R. *Theory of Knowledge*, 2nd ed.; Prentice-Hall: Englewood Cliffs, NJ, USA, 1977.
33. Gettier, E. Is Justified True Belief Knowledge? *Analysis* **1963**, *23*, 121–123. [CrossRef]
34. Pritchard, D.H. Resurrecting the Moorean Response to the Skeptic. *Int. J. Philos. Stud.* **2002**, *10*, 283–307. [CrossRef]
35. Pritchard, D.H. *Epistemic Luck*; Oxford University Press: Oxford, UK, 2005.
36. Dretske, F. The Case Against Closure. In *Contemporary Debates in Epistemology*, 1st ed.; Sosa, E., Steup, M., Eds.; Blackwell: Oxford, UK, 2005; pp. 13–26.
37. Dretske, F. Reply to Hawthorne. In *Contemporary Debates in Epistemology*, 1st ed.; Sosa, E., Steup, M., Eds.; Blackwell: Oxford, UK, 2005; pp. 43–46.
38. Hawthorne, J. The Case for Closure. In *Contemporary Debates in Epistemology*, 1st ed.; Sosa, E., Steup, M., Eds.; Blackwell: Oxford, UK, 2005; pp. 26–43.
39. Vogel, J. Tracking, Closure, and Inductive Knowledge. In *The Possibility of Knowledge: Nozick and His Critics*; Luper-Foy, S., Ed.; Rowman & Littlefield: Totowa, NJ, USA, 1987; pp. 197–215.
40. Wittgenstein, L. *On Certainty*; Anscombe, G.E.M., von Wright, G.H., Eds.; Paul, D.; Anscombe, G.E.M., Trans.; Blackwell: Oxford, UK, 1969.
41. Pritchard, D.H. *Epistemic Angst: Radical Skepticism and the Groundlessness of Our Believing*; Princeton University Press: Princeton, NJ, USA, 2015.
42. Williams, M. *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism*; Blackwell: Oxford, UK, 1991.
43. Hirvelä, J. Global Safety: How to Deal with Necessary Truths. *Synthese* **2019**, *196*, 1167–1186.
44. Bernecker, S. Against Global Method Safety. *Synthese* **2020**, *197*, 5101–5116. [CrossRef]
45. Pritchard, D.H. Safety-Based Epistemology: Whither Now? *J. Philos. Res.* **2009**, *34*, 33–45. [CrossRef]
46. Pritchard, D.H. Anti-Luck Virtue Epistemology. *J. Philos.* **2012**, *109*, 247–279. [CrossRef]
47. Pritchard, D.H. Epistemic Luck. *J. Philos. Res.* **2004**, *29*, 193–222. [CrossRef]
48. Pritchard, D.H. Anti-Luck Epistemology. *Synthese* **2007**, *158*, 277–297.
49. Pritchard, D.H. In Defence of Modest Anti-Luck Epistemology. In *The Sensitivity Principle in Epistemology*; Black, T., Becker, K., Eds.; Cambridge University Press: Cambridge, UK, 2012; pp. 173–192.
50. Pritchard, D.H. Anti-Luck Epistemology and the Gettier Problem. *Philos. Stud.* **2015**, *172*, 93–111.
51. Pritchard, D.H.; Millar, A.; Haddock, A. *The Nature and Value of Knowledge: Three Investigations*; Oxford University Press: Oxford, UK, 2010.

52. Pritchard, D.H. Epistemic Risk. *J. Philos.* **2016**, *113*, 550–571. [CrossRef]
53. Pritchard, D.H. Anti-Risk Virtue Epistemology. In *Virtue-Theoretic Epistemology: New Methods and Approaches*; Greco, J., Kelp, C., Eds.; Cambridge University Press: Cambridge, UK, 2020; pp. 203–224.
54. Pritchard, D.H. *Tempting Fate: A Philosophical Guide to Risk, Luck, and a Meaningful Life*; Princeton University Press: Princeton, NJ, USA, 2026.
55. Pritchard, D.H. The Modal Account of Luck. *Metaphilosophy* **2014**, *45*, 594–619. [CrossRef]
56. Pritchard, D.H. Risk. *Metaphilosophy* **2015**, *46*, 436–461. [CrossRef]
57. Navarro, J. Luck and Risk: How to Tell them Apart. *Metaphilosophy* **2019**, *50*, 63–75. [CrossRef]
58. Navarro, J. Epistemic Luck and Epistemic Risk. *Erkenntnis* **2021**, *88*, 929–950. [CrossRef]
59. Hansson, S.O. Risk. In *Stanford Encyclopedia of Philosophy*; Zalta, E., Nodelman, U., Eds.; Stanford University: Stanford, CA, USA, 2022. Available online: https://plato.stanford.edu/entries/risk/ (accessed on 1 June 2026).
60. Pritchard, D.H.; Smith, M. The Psychology and Philosophy of Luck. *New Ideas Psychol.* **2004**, *22*, 1–28. [CrossRef]
61. Pritchard, D.H. Varieties of Epistemic Risk. *Acta Anal.* **2021**, *37*, 9–23. [CrossRef]
62. Hirvelä, J.; Paterson, N.J. A Unified Theory of Risk. *Philos. Q.* **2024**, 1–21. [CrossRef]
63. Greco, J. Worries about Pritchard's Safety. *Synthese* **2007**, *158*, 299–302.