Michael Blome-Tillmann
McGill University

Original Scientific Paper
UDC

NON-REDUCTIVE SAFETY*

Abstract: *Safety principles in epistemology are often hailed as providing us with an explanation of why we fail to have knowledge in Gettier cases and lottery examples, while at the same time allowing for the fact that we know the negations of sceptical hypotheses. In a recent paper, Sinhababu and Williams have produced an example—the Backward Clock—that is meant to spell trouble for safety accounts of knowledge. I argue that the Backward Clock case is, in fact, unproblematic for the more sophisticated formulations of safety in the literature. However, I then proceed to construct two novel examples that turn out problematic for those formulations—one that provides us with a lottery-style case of safe ignorance and one that is a straightforward case of unsafe knowledge. If these examples succeed, then safety as it is usually conceived in the current debate cannot account for ignorance in all Gettier and lottery-style cases, and neither is it a necessary condition for knowledge. I conclude from these troublesome examples that modal epistemologists ought to embrace a much more simple and non-reductive version of safety, according to which the notion of similarity between possible worlds that determines in which worlds the subject must believe truly is an epistemic notion that cannot be defined or reduced to notions independent of knowledge. The resulting view is shown to also lead to desirable results with respect to lottery cases, certain quantum phenomena, and a puzzling case involving a cautious brain-in-a-vat.*

1. Classical Safety

Since the turn of the century, a number of epistemologists have defended a necessary condition on knowledge that is familiar as the *safety condition*. Safety is meant to provide us with a plausible response to scepticism, by offering us an explanation of how we know both ordinary propositions and the negations of sceptical hypotheses, and thus by delivering a response to sceptical arguments that succeeds without giving up closure. Roughly, according to authors such as Ernest Sosa, Duncan Pritchard, and Timothy Williamson, a subject *S* knows *p* only if *S* could not have easily been wrong with respect to *p*. Even though Sosa has given up on safety in more recent writing, let us begin the discussion with his formulation of the principle, which represents the most straightforward and familiar way to articulate the general idea underlying safety. Here is Sosa's (1999: 146) definition of what I shall call *classical safety*:

* I am indebted to Sven Bernecker and Duncan Pritchard for discussion of earlier versions of this paper, and to an anonymous referee for very helpful comments and suggestions.

(ES) *S*'s belief that *p* is classically safe $=_{df}$
[if *S* were to believe *p*, then *p*].

Given (ES), a belief is classically safe iff it could not have been false easily. In terms of possible worlds, (ES) says that one's belief that *p* is classically safe just in case one believes *p* in a nearby world *w*, only if *p* is true in *w*. Sosa (1999) further defends the view that classical safety is a property of knowledge:

($SAFE_C$) Necessarily, *S* knows *p* only if:
[if *S* were to believe *p*, then *p*].

The main motivation of ($SAFE_C$) consists, according to Sosa (1999), in the fact that it accounts neatly for the fact that we lack knowledge in Gettier cases and lottery examples. In such examples, the explanation goes, we fail to know that *p* because our belief that *p* is not classically safe—our belief could have been false easily as there are many nearby ¬*p*-worlds in which we (falsely) believe that *p*.

Consider, for illustration, the following version of a lottery case, inspired by LJ Cohen (1977):

> *The Gatecrasher*:
> The organizers of the local rodeo decide to sue John for gatecrashing their Saturday afternoon event. Their evidence is as follows: John attended the Saturday afternoon event—he was seen and photographed on the main ranks during the rodeo. No tickets were issued at the entrance, so John cannot be expected to prove having bought a ticket with a ticket stub. However, while more than 1,000 people were counted in the seats, only 157 paid for admission. No further evidence is presented in court.

In the Gatecrasher example, the judge is epistemically rather well justified in believing that John gatecrashed, but she crucially does not know that proposition: for all the judge knows, John was one of the 157 honest fee-paying people in attendance. Thus, while the statistical evidence available to the judge can justify her belief that John gatecrashed,[1] it intuitively cannot ground her *knowledge* that he gatecrashed.

Next, note that Sosa's notion of classical safety provides us with an elegant explanation of this *prima facie* surprising datum. According to ($SAFE_C$), the judge does not know that John gatecrashed because there are numerous nearby possible worlds in which the judge believes falsely that John gatecrashed—namely, precisely those worlds in which John paid the entrance fee instead of climbing the fence. Thus, by requiring that knowledge be free from what many theorists have called *epistemic luck*,[2] ($SAFE_C$) seems to provide us with an elegant explanation of our intuitions—not only in the Gatecrasher example, but also in other lottery-style examples and Gettier cases.[3]

---

1 The probability that John gatecrashed given the judge's evidence is .843.

2 See, for instance, (Pritchard 2005).

3 Note also that the judge cannot justly impose liability on the basis of the statistical evidence available to her. See (Blome-Tillmann 2017a) for discussion.

2. Safe Ignorance: The Backward Clock

Attempts in the literature to discredit safety have usually aimed at producing instances of *unsafe knowledge*—that is, counterexamples to ($SAFE_C$) in which a subject intuitively knows that *p* even though her belief that *p* is classically unsafe.[4] In a recent paper, however, Neil Sinhababu and John Williams (2015) have taken a different route—namely, by producing a Gettier-style example of non-knowledge in which safety does not fail. Thus, according to Sinhababu and Williams, safety does not adequately capture the notion of epistemic luck at issue in Gettier examples and cannot explain why we fail to know in Gettier cases. I shall, in this section, briefly describe the example at issue and then show that it does not turn out problematic for some of the more sophisticated formulations of safety in the literature. In Section 3, I shall then offer a different example that in fact achieves Sinhababu and Williams' goal.[5]

Here is Sinhababu and Williams' example:

> *Backward Clock*:
>
> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. Unbeknownst to you, your clock is a special model designed by a cult that regards the hour starting from 4 pm today as cursed, and wants clocks not to run forward during that hour. So your clock is designed to run perfectly reliably backwards during that hour. At 4 pm the hands of the clock jumped to 5 pm, and it has been running reliably backwards since then. This clock is analogue so its hands sweep its face continuously, but it has no second hand so you cannot tell that it is running backwards from a quick glance. Awaking, you look at the clock at exactly 4.30 pm and observe that its hands point to 4.30 pm. Accordingly you form the belief that it is 4.30 pm. (Williams and Sinhababu 2015)

As Sinhababu and Williams point out, Backward Clock is problematic for classical safety, since your belief that it is 4.30 pm is, intuitively, not knowledge despite being classically safe. It is classically safe because, in nearby worlds in which it is not 4.30 pm when you look at the clock, you do not believe that it is 4.30 pm (in those worlds you (falsely) believe that it is 4.31 pm, 4.32 pm, 4.29 pm, etc.). However, intuitively, you could have easily believed falsely in Backward Clock, and that is why your belief, despite being classically safe, is not knowledge: it is, intuitively, true as a matter of mere luck. Consequently,

---

4 See, for instance, (Neta and Rohrbaugh 2004).

5 Adams and Clarke (2016) also criticize Williams and Sinhababu's example, but they do so by pointing out that it is not a counterexample to *sensitivity*. See also fn. 11 for the topic of sensitivity.

Backward Clock is a Gettier-style example in which classical safety cannot account for the absence of knowledge.

Sinhababu and Williams claim that their example is problematic for more sophisticated formulations of safety, too. To back up this claim they consider the following version of the safety principle, which they ascribe to Duncan Pritchard (2012):

($SAFE_B$) Necessarily, *S*'s knows *p* on basis *B* only if:
[*S* could not have easily formed a false belief on basis *B*].

Sinhababu and Williams argue that ($SAFE_B$)—let us call the principle *Basis Safety*—falls prey to their example, too, and to establish this conclusion they point out that, in Backwards Clock, the basis on which you believe that it is 4.30 pm

> must be that the hands point to 4.30 pm. That you look at the clock is not a sufficient basis for believing that it is 4.30 pm, as this leaves open where the hands are pointing. You need to see that the hands point to 4.30 pm to have grounds for believing that it is 4:30 pm. (Williams and Sinhababu 2015: 53)

While Sinhababu and Williams might be right that their example spells trouble for both ($SAFE_C$) and ($SAFE_B$), there are other versions of safety that clearly avoid the problem. Instead of formulating safety in terms of belief bases, for instance, we might—following some of Pritchard's earlier work—formulate it by appeal to *belief-forming methods*. Call the following principle *Method Safety*:

($SAFE_M$) Necessarily, *S* knows *p* via method *M* only if:
[*S* could not have easily formed a false belief via *M*].

According to ($SAFE_M$), a belief is safe just in case it was produced by a method that leads to true beliefs not only in the actual, but also in nearby worlds. Interestingly, this condition is not satisfied in Backward Clock. This is so because, in Backward Clock, you formed your belief that it is 4.30 pm by the method of *reading the clock in front of you*. By this method, however, you form, in a nearby world, the false belief that it is 4.31 pm when it in fact is 4.29 pm. In Backward Clock, there are, as a consequence, numerous nearby worlds in which the method that you actually apply leads to false beliefs. Thus, in Backward Clock, your belief that it is 4.30 pm is not method-safe and, given ($SAFE_M$), does not qualify as knowledge. Consequently, ($SAFE_M$) provides us with an effective response to the problem posed by Backward Clock.

As already mentioned, method-safety is inspired by some of Duncan Pritchard's earlier work on safety. In particular, in light of examples including necessary truths and other problem cases, Pritchard (2007a: 292, 2007b: 40, 2009: 34) proposes the following definition of what I shall call *weak safety*:

($SAFE_W$) Necessarily, *S* knows *p* only if:
[in most near-by possible worlds in which *S* continues to form her belief about the target proposition in the same way as in the actual world, and in all very close near-by possible worlds in which *S* continues to form her belief about the target proposition in the same way as in the actual world, the belief continues to be true.]"[6]

Pritchard's appeal to 'ways of forming a belief' in this passage clearly bears a strong similarity to the notion of belief-forming methods. Thus, both Methods Safety and Weak Safety seem to provide us with an attractive response to the Backward Clock example presented by Sinhababu and Williams.

There are further ways for the safety theorist to respond to Backward Clock that are worth mentioning here. Consider the following version of safety, which is a variant of ($SAFE_B$) and is inspired by Timothy Williamson's (2009b: 325) discussion of safety principles:

($SAFE_{B^*}$) Necessarily, *S*'s knows *p* on basis *B* only if:
[*S* could not have easily formed a false belief on basis *B* or a similar basis *B**].[7]

It is fairly straightforward to see why ($SAFE_{B^*}$) is not troubled by Backward Clock. For, in Backward Clock, there are numerous nearby worlds in which you believe a falsehood on a basis that is very similar to your actual belief's basis. For instance, in a nearby world in which you look at the clock at 4.29 pm, you believe, on the basis of looking at the clock and seeing the hands point to 4.31, the falsehood that it is 4.31 pm.

Let me sum up. While Sinhababu and Williams's objection to safety is effective with respect to classical safety as formulated by Ernest Sosa in the early days of modal epistemology, alternative and more sophisticated notions of safety are well-positioned to capture the sense in which our beliefs in Backward Clock are true as a matter of epistemic luck.

3. Safe Ignorance: The Opportunistic Gatecrasher

The general strategy pursued by Sinhababu and Williams—namely, to produce a Gettier-type example of epistemic luck that involves safe ignorance—is interesting, and I shall here attempt to produce an example that is better suited to achieve this goal. The case I have in mind is a variant of the lottery-style example mentioned in Section 1 of this paper. Consider what I shall call the *Opportunistic Gatecrasher*. The details in this example

---

6 For a predecessor of this definition, see (Pritchard 2005: 163).

7 Cp. also (Williamson 2009b: 325): "If in a case α one knows *p* on a basis *b*, then in any case close to α in which one believes a proposition *p** close to *p* on a basis [*b**] close to *b*, *p** is true."

are exactly as in the Gatecrasher example from Section 1, but we fill in the background story as follows:

> *The Opportunistic Gatecrasher*:
>
> John is on his way to the bowling alley to meet his friends, as he does on every Saturday afternoon. John would love to watch the rodeo, but he has not been able to afford the ever-rising entrance fee for many years now. This weekend, however, when he passes by the rodeo on his way to the bowling alley, John sees that a lot of people are climbing the fences. Seizing the opportunity to watch the rodeo for free, John decides to join in and gatecrashes.
>
> Realizing that something is at odds, the organizers of the rodeo decide to sue John for gatecrashing their Saturday afternoon event. Their evidence is as follows: John attended the Saturday afternoon event—he was seen and photographed on the main ranks during the rodeo. No tickets were issued at the entrance, so John cannot be expected to prove having bought a ticket with a ticket stub. However, while more than 1,000 people were counted in the seats, only 157 paid for admission. No further evidence is presented in court.

As in our initial example, the judge is, in the Opportunistic Gatecrasher, rather well justified in believing that John gatecrashed, but she crucially does not *know* that John gatecrashed. Again, the statistical evidence available to her cannot ground knowledge: for all the judge knows, John was one of the honest fee-paying attendees at the rodeo.

What is important about the Opportunistic Gatecrasher, however, is that this time (SAFE_C) cannot account for the datum that the judge does not have knowledge. To see this note that the judge's belief that John gatecrashed is classically safe: in all nearby worlds in which the judge believes that John gatecrashed, he in fact gatecrashed. And that is so because, if John had not gatecrashed, he would have gone bowling with his friends and, therefore, could not have been spotted or photographed at the rodeo. Thus, in those nearby worlds in which John does not gatecrash, the judge does not form the (false-in-those-worlds) belief that John gatecrashed. Consequently, the judge's belief that John gatecrashed is classically safe, true, and well-justified. But, crucially, it is not knowledge. The *Opportunistic Gatecrasher* is, therefore, a lottery-style example of problematic epistemic luck that cannot be accounted for by means of classical safety.

One might wonder at this stage whether the alternative and more sophisticated notions of safety discussed in the previous section are better suited to capture the notion of epistemic luck at play in the above example. Consider first Method Safety, reproduced here for convenience:

> (SAFE_M) Necessarily, *S* knows *p* via method *M* only if:
> [*S* could not have easily formed a false belief via *M*].

Is the judge's belief in the Opportunistic Gatecrasher method-safe? It is iff there is no nearby world in which the judge formed a false belief via the relevant method. But what is the relevant method? If the relevant method is *believing on the basis of photographic evidence documenting John's presence at the rodeo and the pertinent statistical evidence*, then the method is safe, since the judge does not have photographic evidence of John's presence at the rodeo in nearby worlds in which he did not gatecrash. Remember that, in those worlds where John did not gatecrash, he went bowling instead of attending the rodeo, and so was not photographed at the rodeo in the first place. If, however, the relevant method is *believing that* x *gatecrashed on the basis of photographic evidence of* x*'s presence and the pertinent statistical evidence*, then the judge's belief that John gatecrashed is not method-safe. And that is so because there are many nearby worlds in which a subject other than John is sued for compensation—and, importantly, in some of those worlds the organizers have picked a defendant for their lawsuit who paid the entrance fee and thus did not gatecrash. Since the judge believes, in those nearby worlds and on the basis of the relevant photographic and statistical evidence, that those fee-paying defendants gatecrashed, the belief at issue is not method-safe. Thus, depending on how we specify the belief-forming method at hand, the judge's belief either is or is not method-safe.

What about Williamson's version of safety ($SAFE_{B^*}$), also reproduced here?

($SAFE_{B^*}$) Necessarily, *S*'s knows *p* on basis *B* only if:
[*S* could not have easily formed a false belief on basis *B* or a similar basis *B**].

In the Opportunistic Gatecrasher the judge believes that John gatecrashed on the basis of the conjunction of photographic evidence of John's presence and the pertinent statistical evidence. Since John was picked at random, there are nearby worlds in which the judge believes of a fee-paying customer on a very *similar* (or even identical) basis that they gatecrashed. The judge's belief that John gatecrashed is, therefore, not basis*-safe, and ($SAFE_{B^*}$) offers us a plausible explanation of why the judge fails to know that John gatecrashed in the Opportunistic Gatecrasher.

While the mentioned principles ($SAFE_{M}$) and ($SAFE_{B^*}$) may very well both be able to handle the example as it was presented above, I take it that the case nevertheless illustrates an important point about safety. For we can fairly easily amend the example presented above to the effect that only gatecrashers are being sued by the organizers in nearby worlds. One way to insure this is by stipulating that the real reason for which the organizers sue John is because they do not like him very much, for reasons entirely independent of his propensity to gatecrash the rodeo. Once we have added such stipulations to the effect that there are no nearby worlds in which the organizers sue somebody other than John, the example illustrates the inadequacy of both Method Safety and Basis* Safety.

In summary, we can, with some imagination, construe lottery-style examples in which certain belief-forming methods are only applied in nearby worlds, if they lead to true beliefs, or, in Williamson's terminology, in which certain beliefs are only formed on a particular kind of basis, if the resulting beliefs are true. The Opportunistic Gatecrasher is, therefore, a fairly simple and straightforward lottery-style example of safe ignorance, giving rise to rather strong and clear intuitions.

4. Testimony and Unsafe Knowledge

While the previous section provided a lottery-style example in which safety cannot explain the absence of knowledge, I shall, in this section, produce a case of *unsafe knowledge* and thus aim to show that safety is not necessary for knowledge. While there are several attempts to produce examples of unsafe knowledge in the literature already, the example I propose here is attractive because of its comparative simplicity.[8] Consider the following case of testimonial knowledge, which I borrow from Jennifer Lackey:

> Chicago Visitor:
> Having just arrived at the train station in Chicago, Morris wishes to obtain directions to the Sears Tower. He looks around, approaches the first adult passerby that he sees, and asks how to get to his desired destination. The passerby, who happens to be a lifelong resident of Chicago and knows the city extraordinarily well, provides Morris with impeccable directions to the Sears Tower by telling him that it is located two blocks east of the train station. Morris unhesitatingly forms the corresponding true belief. (Lackey 2009: 29)

I assume that Morris acquires testimonial knowledge that Sears Tower is two blocks east of the train station in this example. What is important about the example in the present context, however, is that we can amend the case slightly to the effect that the passerby would have told Morris a lie, if he had asked for directions to a different location. Imagine, for instance, that the passerby is an overenthusiastic Democrat, who would have sent Morris in the wrong direction had he asked for directions to the Republican National Convention (RNC).[9] In this scenario, Morris's belief that Sears Tower is two blocks east of the train station is classically safe, and even basis safe, but it is neither method-safe nor basis*-safe. It is classically safe (basis safe) because there is no nearby world in which Morris believes falsely (on the basis of the passerby's testimony) that Sears Tower is two blocks east of the train station. However, it fails to be method-safe because the method of asking a passerby

8 See, for instance, (Comesaña 2005).

9 Of course, I do not mean to suggest that Democrats are prone to lying or deceiving as a political tactic.

for directions leads to false beliefs in nearby worlds in which Morris asks for directions to the RNC rather than for directions to Sears Tower. Similarly, Morris's belief fails to be basis*-safe, because, in those nearby worlds in which Morris asks for directions to the RNC, he believes a falsehood on a basis very similar to the basis of his actual belief that Sears Tower is two blocks east of the train station—namely, on the basis of testimony from the mentioned passerby.[10]

5. Non-Reductive Safety

One might wonder whether any of the above examples spells the end of safety accounts of knowledge. The outlook is, to my mind, not quite as bleak. Consider another principle, also defended by Timothy Williamson (2000: 147, 2009a), which defines what I shall call *Simple Safety*:

(SAFE_S) Necessarily, if one knows *p*, one could not easily have been wrong in a similar case.

Simple Safety offers us, I believe, a straightforward response to both the Opportunistic Gatecrasher and our amended version of the Chicago Visitor. In the Opportunistic Gatecrasher, the judge does not know that John gatecrashed because she could have easily been wrong in similar (even though far away) cases—namely, in precisely those cases in which John paid the entrance fee. Thus, even though the closest worlds in which John pays the entrance fee are overall rather dissimilar to John's actuality, they are nevertheless very similar to John's actuality *in those respects that are relevant for knowledge*. Call this type of similarity *epistemic similarity*. Then, a world *w* can be epistemically similar to a world *w′*, even though *w* is overall rather dissimilar (and thus 'far away') from *w′*. An analogous explanation can be given of our amended version of the Chicago Visitor. In the example, we do not count the case in which Morris asks for directions to the RNC as epistemically similar to the case in which he asks for directions to Sears Tower—one possible explanation being that Sears Tower is not a politically loaded venue, whereas the RNC is, thus potentially rendering a random passerby's testimony unreliable.

Can we give a more informative characterization of epistemic similarity? While it would be desirable to have a reductive account of the notion that allows us to explain in detail how epistemic similarity differs from the intuitive notion of overall resemblance, the demand of an explicit definition or analysis is misplaced. Firstly, it is, as the Gettier literature suggests, rather unlikely that any reductive definition or analysis of knowledge will be

10 Thanks to Sven Bernecker here, who has drawn my attention to Lackey's example (pc) in the context of safety. See also (Bernecker forthcoming) for critical discussion of safety principles similar to what I have called Basis*-Safety.

resistant to counterexample. Secondly, from a methodological point of view, it is perfectly sufficient to explicate a theoretical term of which one has an intuitive grasp—such as the notion of epistemic similarity—by relating it to other intuitive concepts in our theory—such as the notion of knowledge. ($SAFE_S$) does exactly that: it relates the concepts of knowledge and epistemic similarity to each other in a way that allows us to account for the fact that we do not have knowledge in Gettier examples, lottery cases, and the abovementioned examples—a feat that no other conception of safety has so far achieved.[11]

Do we have an intuitive grasp of the notion of epistemic resemblance? We can determine, for a vast array of examples, whether or not a given case qualifies as epistemically similar to the subject's actuality. As mentioned above, there is an intuitive sense in which worlds in which John pays the entrance fee to the rodeo are relevantly similar to his actuality—despite the fact that they are overall not very close to it. Similarly, it is intuitively plausible that worlds in which John is a brain in a vat do not qualify as epistemically similar to John's actuality. Our grasp of the notion of epistemic similarity closely tracks, in the relevant cases, our intuitions as to whether the subject knows. Thus, in the light of a more holistic or non-reductive approach to epistemological theory building, the demand for an explicit definition or analysis of the notion of safety or epistemic similarity appears unwarranted.[12]

---

11 The non-reductive account has further explanatory virtues. It can, for instance, also explain why a reliable eyewitness who saw John climb the fence does not fail to know that John gatecrashed: a reliable eyewitness could *not* have easily been wrong in a similar case.

12 It is worthwhile noting at this point that the *Opportunistic Gatecrasher* is as problematic for sensitivity accounts of knowledge as it is for classical or reductivist accounts of safety. Here is Nozick's (1981: 179ff.) formulation of sensitivity in terms of the ordinary language counterfactual conditional:

(SEN) Necessarily, *S* knows *p* via method *M* only if:
[if *p* were false, then *S* would not believe *p* via *M*].

Next, note that the following counterfactual conditional is true with respect to the *Opportunistic Gatecrasher*:

(A) If John had not gatecrashed, then the judge would not believe, by inferring from the evidence presented in court, that he gatecrashed.

(A) is true with respect to the *Opportunistic Gatecrasher* because the closest worlds in which John does not gatecrash are worlds in which he goes bowling and does not attend the rodeo. In those worlds John was not singled out by the organizers of the rodeo and, consequently, has never been taken to court. Thus, in the closest worlds in which John does not gatecrash, the judge does not believe falsely that John gatecrashed. The judge's belief that John gatecrashed is, as a consequence, sensitive but it is not knowledge. Sensitivity, accordingly, cannot account for the problematic type of epistemic luck we find in the example and does not provide us with an appropriate response to the challenge of lottery-style examples. Of course, many will consider sensitivity accounts of knowledge problematic for independent reasons. As Nozick (1981: 227–229) himself and many others have pointed out, sensitivity accounts of knowledge entail closure failure. For further criticism of sensitivity see (Blome-Tillmann 2017b).

6. Further Advantages: Lotteries and Cautious Brains-in-Vats

Before concluding, it is worthwhile noting two further potential advantages of the view proposed here.[13] First, consider the case of the *Cautious Brain in a Vat*—a variant of the *New Evil Demon* problem.[14] The problem arises from the observation that, intuitively, a brain in a vat (henceforth 'biv') doesn't know that it has less than three hands, despite the fact that that belief is perfectly safe in the classical sense. In all nearby possible worlds in which the cautious biv forms the belief that it has less than three hands, it is true that it has less than three hands. Method safety ($SAFE_M$) or similar basis safety ($SAFE_{B^*}$) cannot solve the problem either, if we think of the cautious biv as a thinker who would never believe that it has two hands, but only that it has less than three hands and similarly for all kinds of other beliefs ('I own less than four bicycles', 'My epistemology class has less than 21 students', 'I see at most one sunrise', and so on). The cautious biv's belief is thus both method safe and basis* safe, since beliefs on similar bases and formed by similar methods in similar worlds are also true. According to Simple Safety, however, the cautious biv's belief that it has less than three hands isn't safe, because worlds in which the cautious biv has three or more hands are, intuitively, epistemically similar to the cautious biv's actuality—despite the fact that they are overall rather dissimilar to the biv's actuality, and thus 'remote'. In contrast, my current belief that I'm not a biv is safe in the way proposed by Simple Safety, because possible worlds in which I am a biv are not only remote but also epistemically dissimilar to actuality.[15]

Second, Simple Safety is plausibly also helpful for dealing with lottery cases. As has been pointed out in the literature,[16] there is a tension between the safety theorists' claim that we don't know that my lottery ticket has lost (because winning the lottery is a very similar case) and her claim that we know all kinds of ordinary propositions about the external world. Consider, for instance, the proposition that the book is on the table. Given According to classical safety, a strong case can be made that we do not know that the book is on the table because there is a very nearby world in which the book has, due to an extremely unlikely quantum phenomenon, tunnelled through the table the very moment we turned away. In that nearby world we thus believe falsely, by means of the same method and on the very same basis as we actually

13 I am greatly indebted to an anonymous referee for this journal, who pointed out the following two advantages of Simple Safety.

14 See (Cohen 1984) for the New Evil Demon Problem.

15 Another response to the example might be to deny that the cautious biv fails to know that it has less than three hands. I shall, however, not pursue this strategy further here.

16 See, for instance, (Blome-Tillmann 2014: ch 5.2; Dodd 2012).

do, that the book is still on the table.[17] However, if epistemic similarity is a basic and irreducible notion, then we can easily uphold the idea that the mentioned world is epistemically rather dissimilar to our actuality (which, again, it intuitively is), whereas other safety theorists need an explanation of why winning the lottery is a close world, but one where the book tunnels through the table isn't.

7. Conclusion

Alleged counterexamples to safety principles in epistemology are often complex and convoluted, and usually give rise to diverging intuitions. I have here developed two novel, rather simple and intuitive examples that are problematic for the traditional, reductionist safety accounts familiar from the literature. I have further argued that an explanation of the data emerging from those cases comes at the price of abandoning the idea that safety can be reductively defined in favour of an account of safety in terms of *epistemic similarity*: only non-reductive accounts of safety seem immune to the problems outlined in this paper. Finally, I have argued that non-reductionism is far from problematic or explanatorily idle. To the contrary, once we abandon the reductionist dogma underlying much of 20th century epistemology, we have cleared the way for a fruitful, albeit non-reductive account of safety. Within the framework of a modal epistemology, simple safety can play an important explanatory role with respect to both responses to sceptical arguments and solutions to the Gettier and lottery problems for knowledge.

Appendix – Table of Safety Principles

Table of Safety Principles

S's belief *p* (which is based on basis *B* and formed via method *M*) is safe iff	
Name	Condition
Classical Safety	if *S* were to believe *p*, then *p*
Basis Safety	if *S* were to believe *p* on basis *B*, then *p*
Method Safety	if *S* were to believe some proposition $p^\star$ via method *M*, then $p^\star$
Similar Basis Safety	if *S* were to believe a similar $p^\star$ on a similar basis $B^\star$, then $p^\star$
Simple Safety	if *S* were to believe $p^\star$ in an epistemically similar case, then $p^\star$

17 Note that Pritchard's *weak safety* (SAFE_W), as discussed in Section 2, can plausibly handle this problem, too.

References

Adams, Fred and Murray Clarke (2016). „Beat the (Backward) Clock." Logos and Episteme **VII**(3): 353–361.

Bernecker, Sven (forthcoming). „Against global method safety." Synthese.

Blome-Tillmann, Michael (2014). Knowledge and Presuppositions. Oxford, Oxford University Press.

Blome-Tillmann, Michael (2017a). 'More likely than not' – Knowledge First and the Role of Bare Statistical Evidence in Courts of Law. Knowledge First – Approaches to Epistemology and Mind. A. Carter, E. C. Gordon and B. Jarvis. Oxford, Oxford University Press**:** 278–292.

Blome-Tillmann, Michael (2017b). „Sensitivity Actually." Philosophy and Phenomenological Research **94**(3): 606–625.

Cohen, L. Jonathan (1977). The probable and the provable. Oxford, Clarendon Press.

Cohen, Stewart (1984). „Justification and truth." Philosophical Studies **46**(3): 279–295.

Comesaña, Juan (2005). „Unsafe Knowledge." Synthese **146**(2): 395–404.

Dodd, Dylan (2012). „Safety, Skepticism, and Lotteries." Erkenntnis **77**(1): 95–120.

Lackey, Jennifer (2009). „Knowledge and Credit." Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition **142**(1): 27–42.

Neta, Ram and G. Rohrbaugh (2004). „Luminosity and the Safety of Knowledge." Pacific Philosophical Quarterly **85**(4): 396–406.

Nozick, Robert (1981). Philosophical Explanations. Oxford, OUP.

Pritchard, Duncan (2005). Epistemic Luck. Oxford, Clarendon.

Pritchard, Duncan (2007a). „Anti-luck epistemology." Synthese **158**(3): 277–297.

Pritchard, Duncan (2007b). Knowledge, Luck and Lotteries. New Waves in Epistemology. V. Hendricks and D. Pritchard. Basingstoke, Palgrave Macmillan**:** 28–51.

Pritchard, Duncan (2009). „Safety-Based Epistemology: Whither Now?" Journal of Philosophical Research **34**: 33–45.

Pritchard, Duncan (2012). „Anti-Luck Virtue Epistemology." Journal of Philosophy **109**(3): 247–279.

Sosa, Ernest (1999). „How to Defeat Opposition to Moore." Philosophical Perspectives – Epistemology **13**: 141–153.

Williams, John N. and Neil Sinhababu (2015). „The Backward Clock, Truth-Tracking, and Safety." Journal of Philosophy **112**(1): 46–55.

Williamson, Timothy (2000). Knowledge and Its Limits. Oxford, OUP.

Williamson, Timothy (2009a). „Probability and Danger." The Amherst Lecture in Philosophy **4**: 1–35.

Williamson, Timothy (2009b). Replies to Critics. Williamson on Knowledge. P. Greenough and D. Pritchard. Oxford, OUP**:** 279–384.