Philosophia
DOI 10.1007/s11406-013-9508-4

Yes, Safety is in Danger

Tomas Bogardus · Chad Marxen

Received: 29 July 2013 / Revised: 7 September 2013 / Accepted: 6 November 2013

Abstract In an essay recently published in this journal ("Is Safety in Danger?"), Fernando Broncano-Berrocal defends the safety condition on knowledge from a counterexample proposed by Tomas Bogardus (*Philosophy and Phenomenological Research*, 2012). In this paper, we will define the safety condition, briefly explain the proposed counterexample, and outline Broncano-Berrocal's defense of the safety condition. We will then raise four objections to Broncano-Berrocal's defense, four implausible implications of his central claim. In the end, we conclude that Broncano-Berrocal's defense of the safety condition is unsuccessful, and that the safety condition on knowledge should be rejected.

Keywords Methods of belief formation · Bogardus · Knowledge · Safety · Epistemic luck

Introduction

The safety condition on knowledge is the claim that a subject's true belief counts as knowledge only if her belief is also, in some important sense, *safe*. And a subject's belief is safe just in case the method she employed to arrive at that belief did not put her in serious epistemic danger, that is, serious danger of arriving at a false belief thereby. In other words, to say that a subject's belief that p is safe is to say that, were she to believe that p via this method, p would be true.[1] Or, alternatively, that not easily would she have believed falsely via that method.[2]

[1]See, for example, Duncan Pritchard (2005, 163): "If a believer knows that p, then in nearly all, if not all, nearby possible worlds in which the believer forms the belief that p in the same way as she does in the actual world, that belief is true." And Steven Luper (2006): "at time t, S knows p by arriving at the belief p through some method M only if: M would, at t, indicate that p was true only if p were true."

[2]See, for example, Ernest Sosa (1999, 142): "[A] belief by S is 'safe' if: as a matter of fact, though perhaps not as a matter of strict necessity, not easily would S believe that p without it being the case that p." And R.M. Sainsbury (1997, 907): "If you know, you couldn't easily have been wrong."

T. Bogardus (✉) · C. Marxen
Department of Philosophy, Pepperdine University, Malibu, CA, USA
e-mail: tbogardus@gmail.com

That *knowledge must be safe* is a trendy view among contemporary epistemologists, for a variety of reasons.[3] In spite of this popularity, Tomas Bogardus (2012) attempts to refute the alleged condition via counterexample. Here's the general recipe for whipping up such a counterexample: first, pick the most virtuous belief-forming method you can imagine, and have a subject form a belief via that method. In the original counterexample, called **Atomic Clock**, a subject named Smith formed a belief about the time on the basis of the world's most accurate clock. Second, add a twist of fate: put the method in danger of malfunctioning, but let the danger remain purely counterfactual. In the original example, Smith's clock was atomic, and it was imperiled by a nearby radioactive isotope. The isotope was due to decay at any moment, and were it to decay it would stop the clock (or even just slow it down significantly), rendering it unreliable.

Now, since that danger remains purely counterfactual—since the clock could have malfunctioned but in fact remained the world's most accurate clock; since things *could* have gone less well epistemically but *didn't*—it's quite tempting to allow that Smith knows the time on the basis of the clock. And yet, one might think, her belief in this scenario is not formed safely, for there are many nearby possible worlds in which she forms a false belief on the basis of that clock, worlds in which the isotope has decayed and the clock has stopped or slowed. It's false that, were Smith to believe via that clock, her belief would be true. Very easily could she have believed falsely via that clock: had the isotope decayed, which it easily might have, Smith easily might have believed falsely on the basis of the clock. And so **Atomic Clock** seems to be a counterexample to the alleged safety condition on knowledge.

But, as any timeshare owner can tell you, some things aren't how they seem. And, in his recent essay, Fernando Broncano-Berrocal (forthcoming) attempts to argue that **Atomic Clock** is not all it looks to be. He concedes that Smith does *know* the time on the basis of the clock. However, Broncano-Berrocal argues that, contrary to appearances, Smith's belief is formed *safely* in **Atomic Clock**. Why think that? Well, Broncano-Berrocal reminds us that safety applies to methods. So, evaluating whether the safety condition holds requires that we examine the method used by Smith in the actual world, and then compare the results of *that very method* in nearby possible worlds. And, according to Broncano-Berrocal, checking a broken clock is a different way of forming beliefs about the time than is checking a working clock. The method that Smith actually uses—checking the world's most accurate clock—is different from the method that would easily have led her astray, namely checking a clock that was stopped by that radioactive isotope. And so it's not true that the method Smith actually used could easily have misled her; *that* method is as good as they get. She easily might have used some other, inferior method (namely, checking a stopped clock), but that's neither

[3] Pritchard (2005, 147–52) argues that it can capture the intuitively attractive idea that knowledge is non-lucky true belief, the central dogma of popular anti-luck epistemologies. Sosa (1999) argues that the view that knowledge must be safe gives an excellent account of inductive and anti-skeptical knowledge. And, according to John Hawthorne (2004, 56 n. 17), the view seems poised to explain why the subject in standard Gettier-style cases lacks knowledge: the subject could so easily have been wrong: "Insofar as we withhold knowledge in Gettier cases, it seems likely that 'ease of mistake' reasoning is at work, since there is a very natural sense, in such cases, in which the true believer forms a belief in a way that could very easily have delivered error."

here nor there as far as the safety condition is concerned, according to Broncano-Berrocal.

So Broncano-Berrocal defends what one might call an *externalist* individuation principle for belief-forming methods. Even though they feel the same "from the inside," forming a belief about the time *via this clock while it's working* is different from forming such a belief *via this clock while it's broken*. And what distinguishes those methods is external to the believer, "upstream of experience," one might say. More carefully, here's Broncano-Berrocal's proposed individuation principle for belief-forming methods, which he calls "R4":

(R4) For any type of method of belief-formation m_1 and for any type of method of belief-formation m_2, $m_1=m_2$ if and only if

(i) m_1 and m_2 are globally reliable to the same degree with respect to the same field of propositions and the same range of circumstances,
(ii) they are both based on vision or olfaction or audition or taction or gustation or testimony or deduction or induction or memory etc., and
(iii) the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.

In favor of (R4), Broncano-Berrocal argues—rather convincingly, to our minds—that the principle gets the right result in a wide variety of cases, from the perspective of a safety theorist. That is, *if* knowledge requires safety, (R4) carves belief-forming methods at the right places. If one is interested in defending the safety condition from counterexamples like the one under discussion, as Broncano-Berrocal himself is, then one ought to defend something like (R4).

And so consider how (R4) might defuse **Atomic Clock**. According to Broncano-Berrocal, conditions (i) and (iii) are both violated in **Atomic Clock**. Let "m_1" refer to the method Smith actually used to tell the time, when the clock was functioning normally, and let "m_2" refer to the method Smith would have used had the isotope decayed and stopped the clock. Broncano-Berrocal claims that "[i]f Smith used m_1 in a range of different situations to form the belief that it is 8:22 am, her belief would be true in most of them. Obviously, the same cannot be said about m_2. Therefore, condition (i) does not hold."

And, further, Broncano-Berrocal claims that "[i]n the circumstances in which Smith uses m_2, a radioactive isotope has decayed disrupting the clock's sensor and stopping the clock. Therefore, those circumstances are not in the set of circumstances with respect to which m_1, Smith's actual method, is globally reliable. Therefore, condition (iii) does not hold." Therefore—Broncano-Berrocal concludes, using (R4)—$m_1 \neq m_2$. And therefore, **Atomic Clock** does not, contrary to appearances, feature a subject who knows via a method that could have gone awry. It features, rather, a subject who knows via a method (a working clock), but who easily might have failed to know via a *different* method that would have been indistinguishable to her "from the inside" (a broken clock). But that doesn't show that the method the subject *actually* used was unsafe.

To recap, here is Broncano-Berrocal's main argument for the conclusion that Smith's belief in **Atomic Clock** was, despite appearances, formed safely. First, the assumptions:

(1) **Safety**: For any subject S, S's belief that p formed in the actual world @ via a belief-forming method of type M is safe if and only if 1) it is true in @ and 2) in

nearly all, if not all, close possible worlds in which S forms the belief that p via a belief-forming method of type M, that belief is true.

(2) (R4) is true.

(3) If (R4) is true, then, despite the danger posed by the soon-to-decay isotope, in **Atomic Clock** Smith's belief formed via a method of type M—namely, *checking the world's most accurate clock while it's functioning properly*—is true in @, and in nearly all, if not all, close possible worlds in which Smith forms the belief that p via a belief-forming method of type M, that belief is true.

Now, the conclusions. From (2) and (3) we infer:

(4) In **Atomic Clock** Smith's belief formed via a method of type M is true in @ and in nearly all, if not all, close possible worlds in which Smith forms the belief that p via a belief-forming method of type M, that belief is true.

Finally, from (1) and (4) we conclude:

(5) In **Atomic Clock** Smith's belief is formed safely.

And if (5) is true, then obviously **Atomic Clock** is not a case of *unsafe* knowledge, and so not a refutation of the safety condition. In what follows, we'll offer four objections to principle (R4), i.e. four reasons to think that premise (2) is false. We'll conclude that Broncano-Berrocal's argument above is unsound. So, his defense of the safety condition from the threat of **Atomic Clock** is unsuccessful, and we should believe that **Atomic Clock** is what it seems to be: a counterexample to the alleged safety condition on knowledge.

Before we turn to objections, allow us to make a quick note about the stakes of this debate. Broncano-Berrocal aims to find the most plausible principle for individuating methods of belief formation, *given that knowledge requires safety.* We'll soon argue that Broncano-Berrocal's individuation principle is implausible. Insofar as one thinks Broncano-Berrocal succeeded in finding the best individuation principle given that knowledge requires safety, the implausibility of this principle counts against the safety condition on knowledge. If anything in the neighborhood of (R4) is the best that the safety theorist can hope for—and we're inclined to think that it is—then safety is in serious danger.

Objection 1: (R4) Entails There are no Unreliable Methods

Are there any unreliable ways of forming beliefs? No need to answer; of course there are. Making important financial decisions based on your dog's body temperature, for example. Or, when it comes to the details of important historical events or crucial public policy, *just guessing*.

We hope you'll agree that there are really bad ways to form beliefs. Pick the worst one you can think of, one which is reliable under no circumstances. Perhaps the method gets it right sometimes, but in no circumstances is it better than chance. Take tyromancy: divination based on observing cheese, especially when it is coagulating. Call that method "m_1," and imagine employing it to arrive at some

belief: you consult your coagulating cheese, and interpret its pattern as predicting that the Democrats will gain control of the United States House of Representatives in 2014. (Let's suppose that, for reasons having nothing to do with cheese, your belief is true.) Now imagine using that method again, at some later time, in similar circumstances, to answer the same question. On the second occasion, call it "m_2." Is $m_1=m_2$? By hypothesis, yes, it's the same terrible method: tyromancy. So far, so good; this all seems perfectly possible.

But let's ask (R4) whether that's possible. For your convenience, here is that principle again:

> (R4) For any type of method of belief-formation m_1 and for any type of method of belief-formation m_2, $m_1=m_2$ if and only if
>
> (i) m_1 and m_2 are globally reliable to the same degree with respect to the same field of propositions and the same range of circumstances,
> (ii) they are both based on vision or olfaction or audition or taction or gustation or testimony or deduction or induction or memory etc., and
> (iii) the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.

In the case we described, condition (i) of (R4) is satisfied: m_1 and m_2 are both reliable to the same (low) degree. And condition (ii) may well be satisfied. But how about condition (iii)? The circumstances in which the target belief is formed via m_2 are *not* in the set of circumstances with respect to which m_1 is globally reliable *because there are no such circumstances*. By hypothesis, m_1 is unreliable in every circumstance. So, (R4) entails that $m_1 \neq m_2$. But, intuitively, "m_1" and "m_2" both picked out the same unreliable method type. Insofar as one thinks there can be unreliable methods, then, (R4) is false. And since there clearly can be such bad methods, we should conclude that (R4) is false. And so premise (2) in Broncano-Berrocal's main argument is false, and his defense of the safety condition fails.

Allow us to respond to a possible objection. An objector may point out that, for all we've said, (R4) is consistent with there being unreliable methods, just none that may be *repeated*. All we've really shown—the objector may say—in the case of tyromancy above, is that m_1 and m_2 are *both* unreliable methods. (R4) may commit us to a surprising number of methods—every time we try to use tyromancy again, we end up using a distinct method—but not to *all* methods being reliable. So things aren't as bad as they look for (R4), the objector may conclude.

We have three responses, and the first is quick: isn't it essential to the concept of a belief-forming method that it be repeatable? And so wouldn't it be a rather heavy cost of a theory that it rules that no unreliable method can be repeated? We vote "yes" for both questions. Secondly, repeat the tyromancy example again, but instead of consulting the cheese twice, consult it only once. And call that one method you used by two names: "m_1" and "m_2." According to (R4), it's impossible that those two names should refer to the same method, for again m_2 is not used in circumstances in which m_1 (i.e., the method itself)

is globally reliable, because there are no circumstances. So there really can be no unreliable methods on (R4). Pick any unreliable method you like: (R4) rules that it is not self-identical.

Thirdly, if one cannot repeat a bad way of believing, then terrible methods like the one described above—divination based on observing cheese—will come out as safe. According to Broncano-Berrocal, a belief is formed safely by some method if and only if the belief is true, and in all (or nearly all) of the nearby possible worlds in which one uses *the same method*, the resulting belief is true. Now, we stipulated that, by the sheerest coincidence, you managed to form a true belief on the basis of that coagulating cheese. And, if, as principle (R4) would have it, methods like tyromancy are so finely individuated that they cannot be repeated, then it will be trivially true that *in all (or nearly all) the worlds in which one uses the same method, the belief comes out as true*, since there are no such worlds. But then, according to (R4), both conditions for safety are met by this terrible method, a paradigmatically unsafe method. So much the worse for (R4), we say. And so, again, premise (2) in Broncano-Berrocal's main argument is false, and his defense of the safety condition fails.[4]

Objection 2: (R4)'s Conditions are Insufficient

Things get worse for (R4): it's possible for two distinct types of belief-forming methods to meet conditions (i)–(iii) of (R4).[5] So, contrary to Broncano-Berrocal's claim, those conditions are not sufficient to show that a type of method m_1 is identical with a type of method m_2. To show this, we'll describe what are clearly two distinct methods of forming beliefs on a topic, which nevertheless satisfy conditions (i)–(iii).

[4] An anonymous referee helpfully encouraged us to consider another natural interpretation of condition (iii) in principle (R4), namely: the circumstances in which the target belief is formed via m_2 are in the set of circumstances that *fix* the global reliability of m_1 (to whatever degree it is reliable). There are two reasons to think this is the charitable interpretation of Broncano-Berrocal. First, it's plausibly the more general principle—applying to reliable and unreliable methods alike—that lies behind and explains Broncano-Berrocal's formulation of (R4), and it could easily be overlooked given the focus on generally *reliable* methods central to the debate. Second, it allows Broncano-Berrocal to avoid the objection of this section, since the unreliable method we described, when used on the second occasion—m_2—would indeed be used within that range of circumstances that fix the (very low) global reliability of the method used on the first occasion—m_1. And in that case our new interpretation of condition (iii) would be satisfied.

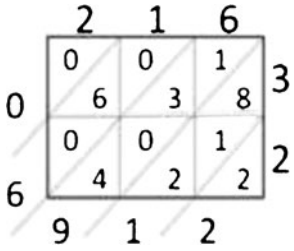
However, there is a strong reason *not* to prefer this alternative interpretation of condition (iii). Namely, it would take out some crucial legs from under Broncano-Berrocal's response to **Atomic Clock**. For, in that scenario, the method one would use were the isotope to decay would be used in circumstances that indeed *fix* the "global" reliability of the method one actually uses. The global reliability of the actual method used in **Atomic Clock** is determined by its likelihood of error on a range of propositions in a range of circumstances across logical space. If the unreliability of the method in circumstances is no barrier to those circumstances being included among those that fix the method's global reliability, and if the counterfactual conditions in **Atomic Clock** are sufficiently mundane (there's no more soon-to-decay isotope, the clock has either slowed down or stopped as clocks often do, etc.), then it's hard to see why those nearby counterfactual circumstances in which the isotope has decayed would not be among those that fix the global reliability of m_1. But then this interpretation of condition (iii) is satisfied in **Atomic Clock**, and Broncano-Berrocal loses one of his only two arguments for the conclusion that condition (iii) is not satisfied in **Atomic Clock.** So, this interpretation of condition (iii) may rescue Broncano-Berrocal from one of our four objections, but at the high cost of sacrificing *half* of his defense of the safety condition.

[5] This objection is due to Marxen.

Consider truth tables and truth trees, for example. Here we have two different methods of testing for logical validity, equivalence, consistency, etc. And yet they are, one may well think, equally reliable. There are also different ways to solve multiplication problems. Most of us learned one algorithm in school, the so-called "traditional" way of doing long multiplication. It looks a bit like this:

$$\begin{array}{r} 216 \\ \times \quad 32 \\ \hline 432 \\ +\,6480 \\ \hline 6912 \end{array}$$

But there are alternative methods which are just as useful, for example "lattice multiplication." The same problem as above looks like this using lattice multiplication:



More trivially, one might use a coin flip to make important financial decisions. Or one might use a standard six-sided die, allowing an odd result to count as the equivalent of "heads" and an even result to count as "tails." Here too we have two distinct methods that are equally reliable (namely, 50 % reliable).

But according to (R4) it's impossible that two distinct methods should be equally reliable. Here again is Broncano-Berrocal's (R4):

(R4) For any type of method of belief-formation m_1 and for any type of method of belief-formation m_2, $m_1 = m_2$ if and only if

(i) m_1 and m_2 are globally reliable to the same degree with respect to the same field of propositions and the same range of circumstances,
(ii) they are both based on vision or olfaction or audition or taction or gustation or testimony or deduction or induction or memory etc., and
(iii) the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.

Think about conditions (i)–(iii) with respect to a paradigmatic instance of traditional long multiplication (name it "m_1") and a paradigmatic instance of lattice multiplication

(name it "m_2"). Condition (i) is plausibly true: those two methods are globally reliable to the same degree with respect to the same field of propositions and the same circumstances.[6] And condition (ii) seems true: both methods are, at bottom, algorithms that rely on simple multiplication and deduction. And condition (iii) may easily be true in paradigmatic cases: the circumstances in which lattice multiplication is used may well be within the set of circumstances with respect to which traditional long multiplication is globally reliable: good lighting, a sharp pencil, a clean sheet of paper, no drugs or alcohol, etc.

But then (R4) would rule that $m_1=m_2$. And yet, intuitively, "m_1" and "m_2" picked out distinct methods: using lattices is a different way to solve multiplication problems than is the traditional method. Insofar as one thinks there can be distinct yet equally reliable methods, then, (R4) is false. And, since there can be distinct yet equally reliable methods, (R4) is false. But then premise (2) in Broncano-Berrocal's main argument is false, and so his defense of the safety condition fails.

Now, dialectically, Broncano-Berrocal does not need the conditions of (R4) to be sufficient. Mere necessity would serve in his main argument, since he wishes to run a modus tollens on an instance of (R4). Still, it's useful to show here that Broncano-Berrocal has not succeeded in giving us a plausible principle of individuation for belief-forming methods in the form of a biconditional, contrary to his claims. And our previous objection targeted the necessity of the conditions of (R4), as will our final objection. Cumulatively, then, our objections make serious trouble for (R4), and so serious trouble for Broncano-Berrocal's main argument, his defense of the safety condition. We turn now to our third objection.

[6] Perhaps you're a stickler here, and you wish to point out that there may well be some circumstances, however remote, in which these two methods would not be equally reliable. As a far-out but possible example, consider trying to do some multiplication problems in a room full of aggressive, muscular people who hate lattices, and who will rough you up should you try to draw any lattices. Your performance on those multiplication problems in that circumstance would likely be worse using lattice multiplication than using traditional long multiplication. Doesn't that show that condition (i) is *not* met in this example, and so that we don't have a counterexample to the sufficiency of the conditions of (R4)?

We respond: yes, that is a way to evade *this* counterexample, though it's much less clear that it can evade the example of coin and the six-sided die given in the text, or other examples such as tyromancy and tasseography (reading tea leaves). What's more, this evasion comes with a high cost: it raises a new threat to (R4). Suppose we individuate methods so finely that a method m_1 can be identical to a method m_2 only if their respective sets of ordered triples of the form <proposition, circumstance of evaluation, reliability>—for every proposition, every possible circumstance, and relativized to subjects—are *exactly* the same. Then we'd get bad results: methods that intuitively ought to come out as identical wouldn't, according to (R4). For example, suppose you use lattice multiplication in ideal conditions on Monday, and call that method m_1. Now suppose that, in virtue of that practice with lattice multiplication, you get slightly better at using it under those conditions and subsequently use it again on Tuesday, calling that method m_2. The set of ordered triples—relativized to you—associated with m_1 on Monday is therefore different from the set of ordered triples associated with m_2. With respect to some propositions, in those circumstances, the reliability of this method (for you) is higher on Tuesday than it was on Monday. And so, on this fine-grained principle of individuation (R4), $m_1 \neq m_2$. But clearly the method has remained the same: you've used lattice multiplication on both occasions. So (R4) is false on this strict construal of condition (i). Yet if we loosen up condition (i) so that methods can tolerate minor changes in their global reliability profiles, then the problems of this section will plague (R4). Therefore, the defender of (R4) has here a dispiriting dilemma.

Objection 3: A Fake Barn Dilemma for (R4)[7]

We believe there is something suspicious about Broncano-Berrocal's treatment of **Fake Barn Country**. As he understands this scenario, recall, "Henry forms the true belief that the object in front of him is a barn. Although the object is a genuine barn, Henry does not know it because the environment is populated with indistinguishable barn replicas that would easily have led him to form false beliefs in the same proposition."

Structurally, **Fake Barn Country** has much in common with **Atomic Clock**. In both cases, things go epistemically well for a subject, though they easily might have gone less well. In **Fake Barn Country**, Henry's eyes fall upon a real barn, though they easily could have landed on a fake barn. And in **Atomic Clock**, Smith looks at the world's most accurate clock, though she easily might have seen instead a broken clock (had the isotope decayed).

Given these structural similarities, we found it striking that Broncano-Berrocal argues that **Atomic Clock** features safe knowledge while **Fake Barn Country** features unsafe non-knowledge. Smith's belief in **Atomic Clock** comes out as safe, according to Broncano-Berrocal, since Smith's actual method—namely, checking a working clock—is different, according to (R4), from her counterfactual method had the isotope decayed—namely, checking a broken clock. Since the former method wouldn't easily go awry, and that's the method Smith actually used, Smith's belief is safe.

But why shouldn't the same go with respect to **Fake Barn Country**? For the love of consistency, why not say that Henry's actual method—namely, looking at a real barn—is different, according to (R4), from his counterfactual method—namely, looking at a fake barn—and that former method would not easily go awry, and therefore Henry's belief is safe? If in **Atomic Clock** Broncano-Berrocal appeals to a salient external factor to distinguish between checking a *working* clock and checking a *broken* clock, why shouldn't he do the same in **Fake Barn Country**, and distinguish looking at a *real* barn from looking at a *fake* barn? What principle justifies Broncano-Berrocal's different treatment of **Atomic Clock** and **Fake Barn Country**?

He is aware of this worry, and offers the following justification:

> One might be concerned about the apparent structural similarity between ATOMIC CLOCK and FAKE BARNS. In FAKE BARNS, the prototypical features that allow Henry to identify an object as a barn are shared both by genuine and fake barns. This is partly the reason circumstances with barn replicas belong to the set of circumstances with respect to which Henry's actual visual method is globally reliable (to belong to that set, the light conditions of the circumstances must be good as well, the distance must be appropriate, and so on). In ATOMIC CLOCK, one might argue, the prototypical features that allow Smith to read clocks are shared by the working and by the stopped clock (both read 8:22 am). Should then circumstances in which the clock is stopped be part of the set of circumstances with respect to which Smith's actual method is globally reliable? The answer is negative. Smith's actual method is truth-conducive because the clock is a *reliable indicator* of the time (according to Bogardus, it is the world's most accurate clock). Consequently, circumstances in which the clock is stopped are not the

[7] This objection is due to Marxen.

kind of circumstances with respect to which Smith's actual method is globally truth-conducive (or reliable).[8]

So, in Atomic Clock, Broncano-Berrocal wonders whether the method Smith would have used had the isotope decayed—m_2—is the same as the method she actually used, viz. looking at a working clock—m_1. In doing that, he consults condition (iii) of principle (R4). That condition tells him to check whether the circumstance in which m_2 was used is among those circumstances in which m_1 is globally reliable. Broncano-Berrocal says it *isn't* among those circumstances, because the circumstances in which m_1 is globally reliable don't include circumstances in which the clock has stopped, as it has in the circumstance in which m_2 is used. So, Broncano-Berrocal concludes, m_2 is distinct from m_1. And m_1 is therefore safe.

But doesn't consistency require the same treatment of Fake Barn Country? There, Broncano-Berrocal wonders whether the method Henry would have used had his eyes fallen on a fake barn—m_2—is the same as the method he actually used, viz. looking at a real barn—m_1. In doing that, he consults condition (iii) of principle (R4). That principle tells him to check whether the circumstance in which m_2 was used is among those circumstances in which m_1 is globally reliable. Shouldn't he say—as he did above, vis-à-vis **Atomic Clock**—that it is *not*, because the circumstances in which m_1 is globally reliable don't include circumstances in which the barn is fake, as it is in the circumstance in which m_2 is used? So, in the interest of consistency, shouldn't Broncano-Berrocal conclude that m_2 is distinct from m_1, and that m_1 is therefore safe, contrary to the obvious? If not, why the different treatment of the two cases?

We find in the quoted passage no principle that justifies Broncano-Berrocal's different treatments of **Atomic Clock** and **Fake Barn** Country. And we searched his essay in vain for a difference to justify that distinction. When analyzing **Atomic Clock**, Broncano-Berrocal assumes that the actual method is most aptly described as "the method of forming beliefs by looking at the working clock" and the counterfactual method is most aptly described as "the method of forming beliefs by looking at the stopped clock." In **Fake Barn Country**, however, Henry's method is never clearly described. The clearest Broncano-Berrocal gets is a reference to Henry's "ability to identify barns." Broncano-Berrocal says this ability is globally reliable for Henry, though it fails in the nearby world where Henry stands before a fake barn, which is why Henry's belief is unsafe. Well, in **Atomic Clock** Smith has an ability to read clocks. And this ability fails to deliver a true belief in nearby worlds where Smith stands before a broken clock. Shouldn't Smith's belief therefore be unsafe, just as Henry's is? Again, where is the distinction to justify Broncano-Berrocal's different treatment of these two cases?

As far as we can tell, Broncano-Berrocal unjustifiably treats similar cases differently here. To restore harmony between his views here, we suggest two options: he might rethink his treatment of **Atomic Clock**. Perhaps he should say instead that Smith's belief is *unsafe*, just as Henry's is in **Fake Barn Country**. But of course this line, combined with his view that Smith knows in **Atomic Clock**, would sink his beloved safety condition. That's the first horn of dilemma.

[8] We are grateful to an anonymous referee for encouraging us to interact directly with this quotation.

Alternatively, he might rethink his treatment of **Fake Barn Country**. Perhaps he should say instead that Henry's belief is safe, and perhaps even is knowledge. (Though this is an unpopular view, several prominent epistemologists have defended it—safety in numbers.) He can line up his analysis of **Fake Barn Country** with his analysis of **Atomic Clock**: Henry's actual method is looking at a real barn. That's an extremely reliable way for Henry to determine whether there's a barn in front of him. But he easily might have used a *different* method, indistinguishable to him "from the inside": looking at a fake barn. That's certainly not a reliable way for Henry to determine whether there's a barn in front of him. But since he used only that former method—which is ultra-reliable—his belief was formed safely, Broncano-Berrocal might say.

But suppose we change the story a bit so that, unbeknownst to Henry, the landscape before him is covered in fake barns, with only one real barn peeking out. On this line we're considering, Broncano-Berrocal would say that, should Henry's attention happen to find that one genuine barn—even after incorrectly taking 40 fake barns for real barns, let's say—Henry can know on the basis of his ultra-reliable, externalistically-individuated method of *looking at a real barn* that it is indeed a real barn. It's that *other* method he used 40 times before, *looking at a fake barn*, which is unsafe. *This* time, when by the sheerest coincidence he happens to glimpse the real barn in a sea of fakes, his belief is perfectly safe. Ignore those nearby possible worlds in which he looks at a fake; those aren't relevant to determining whether he easily might have been wrong *this* time. Consider instead only those possible worlds, scattered distantly through logical space, in which he luckily finds the diamond in the rough. Then you'll see that Henry formed his belief with no serious threat of error.

Does that sound implausible to you? It does to us, which is why we count this as the second horn of a menacing dilemma. But it has every penny of plausibility that Broncano-Berrocal's treatment of **Atomic Clock** has. That, in a quick turn, is why we recommend passing on Broncano-Berrocal's analysis of **Atomic Clock** and instead accepting it for what it seems to be: a genuine case of unsafe knowledge, and so a refutation of the safety condition on knowledge.

Let's turn finally to the conflicting intuitions at the bedrock of this debate. Below, in closing, we'll explain Broncano-Berrocal's admittedly attractive core intuition—the intuition driving his defense of the safety condition against **Atomic Clock**—and why we should reject it.

Objection 4: When can Methods Fail, According to (R4)?

Can belief-forming methods go awry due to the circumstances in which they're used? We think the answer is obvious: *of course* a method of forming beliefs could have some circumstances in which it is unreliable. Think of a simple electronic calculator, for example. It's a very accurate way of forming beliefs about (some subset of) mathematics here, at this low pressure, moderate temperature, and low humidity. But it's a very unreliable way of forming such beliefs at the bottom of the ocean, say, with high pressure, low temperature, and high humidity. So common sense delivers to us this datum: an ordinary calculator is a way of forming beliefs about mathematics that is reliable here, but unreliable deep under water.

Yet Broncano-Berrocal's (R4) can't make any sense of that datum; in fact it must deny that datum. To see this, consider the clearest case you can think of in which a method fails. Suppose, for example, that you form a belief about the product of a complicated multiplication problem on the basis of your electronic calculator here, dry, at sea-level. name that method "m_1." Now suppose you use your calculator again, but this time under water, where calculators like that are unreliable. On that second occasion, baptize the method you used "m_2." Is $m_1=m_2$? Intuitively, we'd say, yes of course: you've used the same method in two different circumstances (the first auspicious, the second not).

But what does (R4) say? Here for the last time is Broncano-Berrocal's principle of individuation for belief-forming methods:

> (R4) For any type of method of belief-formation m_1 and for any type of method of belief-formation m_2, $m_1=m_2$ if and only if

(i) m_1 and m_2 are globally reliable to the same degree with respect to the same field of propositions and the same range of circumstances,
(ii) they are both based on vision or olfaction or audition or taction or gustation or testimony or deduction or induction or memory etc., and
(iii) the circumstances in which the target belief is formed via m_2 are in the set of circumstances with respect to which m_1 is globally reliable.

In the case we described, condition (iii) is not satisfied. The circumstances in which the target belief is formed via m_2 are *not* in the set of circumstances with respect to which m_1 is globally reliable. By hypothesis, m_1 is unreliable under water, and m_2 was used under water. So, (R4) entails that $m1 \neq m2$. But, according to us, intuitively "m_1" and "m_2" both picked out the same method. Insofar as one thinks methods can fail due to inauspicious circumstances, then, (R4) is false.[9]

We've highlighted here the intuition that methods can fail, and shown how that intuition counts against (R4). And yet there is something attractive about (R4). Something seems right about allowing external factors to play a role in individuating methods of belief formation. Broncano-Berrocal provides a nice illustration of this temptation:

> Terry the taxi driver might be a very reliable way for you to go home, unless Terry is completely drunk, in which case Terry's taxi is a terribly unreliable way

[9] Just what does it take for circumstances to be among those in which a method is globally reliable? Broncano-Berrocal gives us a hint in **Fake Barn Country**: he says, of Henry's method, that for circumstances "to belong to that set, the light conditions of the circumstances must be good as well, the distance must be appropriate, and so on." And we are grateful to an anonymous referee for helpfully suggesting that, for Broncano-Berrocal, all and only a method's "normal" circumstances of use will fix the method's degree of global reliability.

As we've pointed out in this section, on this understanding of "global reliability," (R4) entails that no method can fail in abnormal circumstances. But it gets worse: (R4) will also entail that no method can even be *used* in abnormal circumstances, for any such use will violate condition (iii). We'll be forced to say, counterintuitively, that those brave astronauts who gazed back at Earth from the lunar surface used a method distinct from our vision. They didn't *see* anything up there, in fact. And if no method can be used in abnormal circumstances, just what are we to make of this distinction between "normal" and "abnormal" circumstances in which a method is used? On this interpretation, the latter category is necessarily empty, for every method: belief forming methods can only be used in normal circumstances, circumstances in which they are reliable. That's powerfully counterintuitive, and that's a serious strike against (R4).

of getting home. The degrees of reliability here are so different that we judge that these are different ways of getting home, even though one is also tempted to judge that they are instances of the same type of method (Terry's taxi) used in different circumstances (driving sober vs. driving drunk).

We believe Broncano-Berrocal has put his finger on something important here: a bedrock of conflicting intuitions. Suppose I use Terry's taxi to get home on Monday, when Terry's sober. And then suppose I use Terry's taxi again on Tuesday, when the demon drink renders Terry's service terribly unreliable. We admit to feeling the pull toward saying that on Monday and Tuesday I've used two different ways of getting home: one reliable, the other not. But, at the same time, we're also tempted by this alternative description of the situation: it's really the *circumstances* that are different here. On both occasions I used the same method of returning home—namely, Terry's taxi—but that method was reliable in Monday's circumstances and unreliable in Tuesday's.

So, which temptation should we indulge? Do we have here one method that fails in certain circumstances? Or two methods, one reliable and the other not? We believe that our arguments in this paper strongly suggest that we should lean toward the first option: we have here one method that fails. For if we follow Broncano-Berrocal we'll run into the objections above: there are no unreliable methods, no method can ever fail due to inauspicious circumstances, and there couldn't be equally reliable but distinct methods. We'll also have trouble explaining why truly troubling Fake Barn cases do not feature knowledge. In those cases, we'd like to point to *one* method that could easily have failed, but if we follow Broncano-Berrocal here there is no such method. Rather, there are two: one reliable the other not, and only the reliable method was used to arrive at the belief that there is a barn out there.

And so, Broncano-Berrocal offers the safety theorist a victory so ruinous it's tantamount to defeat. Instead, let's cleave to the conviction that methods can fail. Terry's taxi is the same method of returning home whether Terry's drunk or sober (though we suggest you use that method only when he's sober). And, in **Atomic Clock**, Smith's clock is the same method of telling the time whether the clock is running or stopped (though we suggest you use that method only when it's running).

But if there really is only one method in play in **Atomic Clock**, then Smith is indeed awash in a sea of nearby possible worlds where she forms false beliefs using the same method that delivered a true belief in the actual world. And so her belief really was formed unsafely, despite the fact that—as Broncano-Berrocal admits—Smith's belief is genuine knowledge. But then knowledge need not be safe, and the safety condition is false. For all Broncano-Berrocal says, then, one ought to take **Atomic Clock** for what it seems to be: a counterexample to the safety condition. And therefore safety really is in serious danger.

References

Bogardus, T. (2012). Knowledge under threat. *Philosophy and Phenomenological Research*. doi:10.1111/j.1933-1592.2011.00564.x.

Broncano-Berrocal, F. (forthcoming). Is safety in danger? *Philosophia*.

Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: Oxford University Press.

Luper, S. (2006). Restorative rigging and the safe indication account. *Synthese, 153*, 161–170.
Pritchard, D. (2005). *Epistemic luck*. Oxford: Oxford University Press.
Sainsbury, R. M. (1997). Easy possibilities. *Philosophy and Phenomenological Research, 57*, 907–919.
Sosa, E. (1999). How to defeat opposition to Moore. In J. Tomberlin (Ed.), *Philosophical perspectives 13: Epistemology*. Blackwell, 141–154.