



# A new solution to the safety dilemma

Dario Mortini<sup>1</sup>

Received: 19 April 2021 / Accepted: 16 February 2022 / Published online: 12 April 2022

© The Author(s) 2022

## Abstract

Despite the substantial appeal of the safety condition, Kelp (J Philos Res 34:21–31, 2009; Am Philos Q 53:27–37; Good Thinking. A Knowledge First Virtue Epistemology, Routledge, London, 2018) has raised a difficult challenge for safety-theoretic accounts of knowledge. By combining Gettier-style fake barn cases with epistemic Frankfurt cases, he concludes that no formulation of safety can be strong enough to predict ignorance in the former and weak enough to accommodate knowledge in the latter. In this note, my contribution is two-fold. Firstly, I take up Kelp’s challenge and I show that, once properly understood, safety successfully rises to it. Secondly, I draw a more general lesson on the safety condition: a satisfactory solution to Kelp’s challenge calls for a revision of the standard formulation of safety, which must be indexed to *both* methods and environments. My conclusion will be that an environment-relative version of safety not only meets Kelp’s challenge, but it also advances our understanding of the safety condition on knowledge.

**Keywords** Safety · Fake barn cases · Epistemic Frankfurt cases · Epistemic luck · Environmental luck · Higher-order environmental luck · Belief-forming methods · Environments

## 1 The hardest challenge for safety

According to the safety condition, knowledge requires protection from error: if one knows, then one could not have *easily* been wrong in similar cases. As a modal condition, safety is formulated by reference to possible worlds and indexed to methods of belief formation:

---

✉ Dario Mortini  
Dario.Mortini@glasgow.ac.uk

<sup>1</sup> Present Address: University of Glasgow, 69 Oakfield Avenue, Glasgow G12 8QQ, UK

**SAFETY.** In most or all close possible worlds in which S believes that  $p$  via the same method of belief formation M that S uses in the actual world,  $p$  is true.<sup>1</sup>

Safety has considerable explanatory power. As an anti-luck condition, it promises to solve the Gettier problem: safety delivers the correct verdict of ignorance in various Gettier-style cases. By denying knowledge of lottery propositions, safety deals also with another much debated issue in epistemology—the lottery problem.<sup>2</sup> Moreover, safety complies with the closure principle, and, unlike sensitivity, it does not open the door to a disturbing sceptical threat (cf. Pritchard, 2008, p. 446).

Despite such substantial appeal, the safety condition has now fallen on hard times. In various places, Kelp (2009) has challenged safety with a troubling conjunction of cases: Gettier-style fake barn cases and more ingenious epistemic Frankfurt cases. Begin with the former:

**FAKE BARN.** Barney, a reliable barn spotter, is driving through the countryside. He looks out of the window, sees a barn and comes to believe that he is looking at a barn. Whilst Barney's belief is true, unbeknownst to him, the structure he is looking at is the only real barn in an area filled with fake barns that are indistinguishable from real barns. (cf. Goldman, 1976)

Intuitively, Barney does not know that he's looking at a barn, and safety neatly explains why. Barney's belief could too easily have been false: in many similar close possible worlds where he forms the same belief via the same method (i.e., visual perception), Barney's belief is false because he's looking at one of the many papier-mâché copies of the real barn. Here, safety delivers the correct verdict. But now consider the following apparently similar case:

**FRANKFURT CLOCK.** Russell's arch-nemesis, a powerful demon, has an interest that Russell forms a belief that it is 8:22 by looking at the grandfather clock in the hallway when he comes down the stairs. Russell's arch-nemesis is prepared to do whatever it may take in order to ensure that Russell acquires a belief that it is 8:22 by looking at the grandfather clock when he comes down the stairs. However, Russell's arch-nemesis is also lazy. He will act only if Russell does not come down the stairs at 8:22 of his own accord. Suppose, as it so happens, Russell does come down the stairs at 8:22. Russell's arch-nemesis remains inactive. Russell forms a belief that it is 8:22. It is 8:22. The grandfather clock is working reliably as always. (Kelp 2016, p. 28)

Intuitively, Russell *does know* the time. Yet, his belief appears to be unsafe: while Russell's belief is true, it could too easily have been false. For sake of vividness, consider the possible worlds where Russell comes down the stairs a few minutes earlier. In these nearby worlds, Russell forms a false belief via the same method, which Kelp identifies as Russell's ability to read clocks (Kelp, 2016, p. 28). **FRANKFURT CLOCK**

<sup>1</sup> Prominent defenders of safety include Sainsbury (1997), Sosa (1999), Williamson (2000), and Pritchard (2005). For recent discussion of the safety condition on knowledge, see Bernecker (2020), Greco (2020b) and hirvelä (2019). See Rabinowitz (2011) for an overview.

<sup>2</sup> Consider Chisholm's sheep in the field Gettier-style case: in nearby worlds where there's no sheep behind the rock, the agent forms a false belief, hence no knowledge. Same with lottery-style propositions: since there's a similar nearby world where your ticket wins, your belief that your ticket is a loser could too easily be false (Pritchard 2012, pp. 253–254).

is offered as a straightforward instance of unsafe knowledge: Russell knows, but the belief in question is *unsafe*. In this case, safety delivers the wrong verdict.

Kelp's argument casts into serious doubts the success of safety as anti-Gettier condition on knowledge. To fully appreciate the significance of his challenge, we must consider FAKE BARN *in conjunction* with FRANKFURT CLOCK.<sup>3</sup> It looks as though no version of the safety principle will be *strong* enough to predict ignorance in Gettier-style cases like FAKE BARN and, at the same time, *weak* enough to accommodate knowledge in FRANKFURT CLOCK.<sup>4</sup> Kelp calls this "the safety dilemma". He draws the following pessimistic conclusion, worth quoting in full:

I'd like to add that meeting this challenge is by no means trivial. Most standard proposals in the literature are prone to fail it in one way or another. Modal accounts, such as safety and sensitivity-based accounts, fail it because the conditions they impose are too strong: they predict absence of knowledge in Frankfurt cases. Other accounts fail in that the proposed conditions are too weak to explain the absence of knowledge in fake barn cases. If there exists a workable account of knowledge in terms of justified true belief and an anti-Gettier condition that gets both fake barn cases and Frankfurt cases right, epistemology still awaits its discovery. (Kelp, 2016, p. 36)

Safety theorists have not met Kelp's challenge yet, at least not in Kelp's terms. To wit, Pritchard (2009: Sect. 3) tentatively defends a no-knowledge verdict in FRANKFURT CLOCK. Given the presence of environmental luck in *both* FAKE BARN and FRANKFURT CLOCK, he maintains that neither is a genuine case of knowledge. However, Pritchard's response is counterintuitive: the knowledge intuition in FRANKFURT CLOCK is plausible, and it's hardly explained away by considerations pertaining to environmental luck. Other safety theorists (e.g., Grundmann, 2018) have attempted to capture the knowledge intuition in FRANKFURT CLOCK by appealing to a more fine-grained individuation of belief-forming methods. As we will see, this strategy struggles to explain absence of knowledge in FAKE BARN. At present, Kelp's challenge still stands.

In this paper, I revisit Kelp's challenge and develop a better solution than the ones currently on offer. I shall argue that a properly understood safety condition successfully captures the intuitions of knowledge and ignorance in FRANKFURT CLOCK and FAKE BARN. The two vignettes display an important but hitherto overlooked difference in their respective *environments*. Once brought into focus, such difference has far-reaching consequences. Firstly, it provides a principled way out of Kelp's dilemma, and, in doing so, it also weakens one of the main motivations for the recent knowledge-first shift in the virtue-epistemological camp (e.g., Miracchi, 2015; Kelp, 2017, 2018). Crucially, Kelp raises this dilemma to motivate the move from traditional (or indirect)

<sup>3</sup> As an anonymous reviewer helpfully notices, Kelp's epistemic Frankfurt case closely resembles other counterexamples to the necessity of safety for knowledge moved by Neta and Rohrbaugh (2004), Comesaña (2005), Baumann (2008) and Bogardus (2014). However, Kelp takes these counterexamples one step further by conjoining them with fake-barn cases, thus posing an even more difficult challenge for the safety condition on knowledge.

<sup>4</sup> For the purpose of this paper, I assume the no-knowledge verdict in FAKE BARN to be correct. However, I also note that both epistemologists (Lycan 2006; Sosa 2010, pp. 472–473) and experimental philosophers (Colaço et al., 2014) have adduced arguments to challenge the correctness of the no-knowledge verdict in cases like FAKE BARN.

virtue epistemology to knowledge-first (or direct) virtue epistemology: with a principled way out of his dilemma on the table, knowledge-first virtue epistemology loses one of its main motivations.<sup>5</sup> Secondly, this key environmental difference is of independent interest to safety theorists: in fact, it further elucidates the safety condition on knowledge, which is best understood as relative not only to belief-forming methods but also to environments.

Having clarified my aims, I now proceed to criticise the two extant solutions to Kelp's dilemma. The former is *conciliatory*: by adopting a more fine-grained individuation of belief-forming methods, it attempts to reconcile safety with the knowledge verdict in FRANKFURT CLOCK. The latter is *revisionary*: by insisting on the presence of environmental luck, it defends a verdict of ignorance in both FRANKFURT CLOCK and FAKE BARN. Neither is ultimately successful. To do better, I close by outlining a refined environment-relative safety condition: unlike its conciliatory and revisionary rivals, this solution rises to Kelp's challenge.

## 2 Methods

To avoid counterexamples like Nozick's grandmother case (Nozick, 1981, p. 179) or Goldman's dachshund vignette (Goldman, 2009, p. 80), modal conditions on knowledge are usually indexed to belief-forming methods.<sup>6,7</sup> Could the safety theorist also meet Kelp's challenge by appeal to a more specific, fine-grained individuation of methods? Unfortunately, the answer is negative. In this section, I argue that while initially promising, this conciliatory strategy ultimately fails. In fact, by yielding a knowledge verdict in both FRANKFURT CLOCK and FAKE BARN, the conciliatory strategy simply doesn't distinguish between our two cases of interest.

Let's focus on methods. In assessing FRANKFURT CLOCK, Kelp comments as follows:

Intuitively, Russell knows that it is 8:22 in this case. After all, we may assume, he has the ability to read the clock and forms a belief via an exercise of this ability. *Moreover, the clock is actually functioning properly, and the reading is accurate.* At the same time, Russell's belief is not safe. After all, among the very close nearby worlds are worlds at which Russell comes down the stairs a minute

<sup>5</sup> See Kelp (2016 pp. 35–36, 2017, pp. 240–243, 2018, pp. 78–79) for the connection of the safety dilemma with knowledge-first virtue epistemology. The other main motivation for knowledge-first virtue epistemology stems from the difficulties of traditional virtue epistemology in dealing with Gettier cases (Miracchi 2015; Kelp 2017, pp. 229–230, 2018, pp. 47–52). However, Hirvelä (2019a) argues that, *pace* Kelp and Miracchi, traditional virtue epistemology can in fact handle these Gettier cases successfully. Special thanks to an anonymous reviewer for encouraging me to emphasise this point.

<sup>6</sup> Let me illustrate the main upshot of these counterexamples. In Nozick's vignette, the agent forms a false belief in nearby worlds, but the belief-forming method is crucially different (Granny is *told* that her grandson is well instead of *seeing* that he's well). These cases have led safety theorists to focus not only on the truth value of the target proposition in nearby worlds, but also on whether the agent's belief continues to be true in nearby worlds *given the way it was formed in the actual world*. For further discussion, see Pritchard (2008, pp. 444–446, 2013, pp. 158) and Rabinowitz (2011: Sect. 3).

<sup>7</sup> To be precise, Nozick's and Goldman's cases are offered in support of method-relative *sensitivity* rather than safety. However, similar considerations equally motivate a method-relative version of safety.

earlier or later. *At those worlds, the demon intervenes with the result that Russell takes a reading from an inaccurate clock and ends up with a false belief that it is 8:22. At the same time, Russell continues to form his belief in the same way as in the actual world, that is, by exercising his clock-reading ability.* Since safety does not tolerate false beliefs at very close nearby possible worlds at which the agent forms his belief *in the same way* as in the actual world, Russell's belief turns out unsafe. (Kelp, 2016, p. 28; emphasis mine)

As the quote suggests, Kelp identifies the relevant belief-forming method as the *general* ability to read a clock. To make his case, he moves from worlds where the clock is functioning properly and accurately to worlds where the clock is *not* functioning properly and it's *inaccurate*. However, in such worlds the belief-forming method is crucially different. One thing is to form beliefs on the basis of a properly functioning *accurate* clock; another is to form beliefs on the basis of a manipulated, *inaccurate* clock. If in lieu of the latter we assume the former belief-forming method and we then focus on the worlds where the clock is functioning properly, Russell does keep forming true beliefs. Once specified as the more fine-grained ability to read a *properly functioning* clock, the belief-forming method turns out safe: in this way, safety captures the knowledge verdict in FRANKFURT CLOCK.<sup>8</sup>

In fairness to the conciliatory strategy, the focus on worlds where the clock is functioning properly enjoys independent plausibility. If we agree with Kelp that the relevant belief-forming method is Russell's *ability* to read a clock, it makes sense to specify the relevant ability by including the fact that the clock is functioning properly. This is clearer if we think of abilities along the lines of Sosa's influential triple-S model (Sosa, 2010, pp. 465–467; 2017, pp. 191–192). For every ability, we can specify the innermost *seat* (or *skill*) in the agent's brain, nervous system and body, the agent's inner *shape* (are they drunk, sober, asleep, alert?) and the external *situation* the agent is in (are there any relevant external interferences or intervening factors?). Russell gains knowledge only through the exercise of a *complete* ability to read a clock: the complete ability is exercised only when *each* S is in play, that is, only when he is properly situated. The relevant situation here must include a properly functioning clock: in the worlds where Russell forms a false belief, the relevant ability is different due to the crucial change of situation (the clock stops). Once the belief-forming method is specified by reference to Russell's *complete* ability, his belief is safe.<sup>9</sup>

While *prima facie* promising, the conciliatory strategy is ultimately unsuccessful. Recall the key challenge of the safety dilemma: to accommodate presence of knowledge in FRANKFURT CLOCK *and* to capture the ignorance intuition in FAKE BARN. A more fine-grained individuation of methods achieves the former but crucially fails to deliver on the latter. To see why, let's consider again FAKE BARN. If we include the fact that the clock is *functioning properly* in FRANKFURT CLOCK, by parity of reasoning we will also have to concede that Barney is looking at the only *real* barn in FAKE BARN. Advocates of the conciliatory strategy will have to individuate a more

<sup>8</sup> See Grundmann (2018, pp. 10–12) for a brief defence of this line of response Kelp's epistemic Frankfurt case.

<sup>9</sup> Hirvelä (2019b, pp. 1178–1180) also defends a similar virtue-theoretic approach to individuate safe belief-forming methods.

fine-grained belief-forming method in *both* cases: so, just like Russell reads from a properly functioning clock instead of merely reading from a clock, Barney also sees the real barn instead of merely having a visual appearance of a barn.<sup>10</sup> But, once specified in the same way as Russell's, Barney's belief will be safe too: if he forms beliefs by looking at *that* real barn, then he will trivially form true beliefs in every close world in which he's looking at *that* very same barn. Otherwise put, this more specific belief-forming method yields a *de re* true belief in every close world.<sup>11</sup> A more fine-grained individuation of methods may reconcile safety with the knowledge verdict in FRANKFURT CLOCK, but it inevitably fails to explain ignorance in FAKE BARN.

This more fine-grained criterion of individuation of belief-forming methods spells trouble for the safety theorist because it's so specific to be *factive*, and thus yields only true beliefs: the clock is functioning properly and *accurately*; Barney sees that there's a barn and *in fact* there is a barn. I am not the first to raise this type of worries concerning such way of individuating belief-forming methods. For instance, Goldman (2009, p. 80) also argues that an analogous too fine-grained criterion of individuation of belief-forming methods raises trouble for Williamson's safety condition on knowledge, and so does Whitcomb (2008).<sup>12</sup> In a similar spirit, and more to the point given the cases of interest here, Ranalli (2014, pp. 1231–1234) insightfully observes that, if understood with reference to the agent's *actual* and *factive* belief-forming method, safety inevitably delivers a knowledge verdict in FAKE BARN. Here's why. Given factivity, it's logically impossible to see a barn *without* there not being a barn: accordingly, in all the nearby worlds where the agent sees a barn and believes there is a barn the agent forms a true belief. The safety condition is satisfied and the verdict in FAKE BARN *has* to be knowledge.<sup>13</sup> Similar problems arise also in the context of the safety dilemma: as a result, the conciliatory strategy leaves the dilemma unsolved.

For sake of clarity and to make the problem more vivid, I'd like to offer an additional case. A genuinely perceptual version of FRANKFURT CLOCK brings out even more clearly the difficulties that the conciliatory strategy incurs. Consider the following case:

**BOWL OF GRAPES.** You and your flatmate are both very hungry, so you go to the kitchen to have some grapes. You desperately want those grapes, but your

<sup>10</sup> On this point, terminological clarification is needed. Need a proper functioning accurate clock be as accurate as the veridical perception of the only real barn? Perhaps the clock's hands are slightly off and therefore slightly inaccurate, and yet the clock seems to be still functioning properly. But by 'properly functioning clock' I mean a properly functioning *accurate enough* clock. A properly functioning clock need not (and often will not) be *perfectly* accurate: even if the hands are slightly off, it will be accurate *enough* to read the correct time. This imperfect degree of accuracy will apply to the belief-forming method in FAKE BARN: properly functioning perceptual capacities need not (and often will not) be *perfectly* accurate (perhaps they won't discriminate some details of the shape or colour of the barn), but they're nonetheless accurate *enough* to single out the real barn. Parity of reasoning here requires two very fine-grained and ultimately factive methods: reading from a properly accurate clock and having an accurate perception of the real barn. Thanks to an anonymous reviewer for encouraging me to be more precise on this.

<sup>11</sup> For more discussion of *de re* barn beliefs, see Bernecker (2020, p. 1107), Hiller and Neta (2007, p. 312), Pryor (2004: 71), Hawthorne (2004, p. 56, fn. 17) and Brown (2000).

<sup>12</sup> See Williamson (2009, pp. 305–308) for a reply to Goldman's criticism.

<sup>13</sup> Special thanks to an anonymous reviewer for drawing my attention to this point.

flatmate has an interest that you don't eat them, so they're willing to do whatever it takes to eat the grapes first and replace them with fake grapes. However, your flatmate is slow, so you get to the kitchen before them. Everything is in order, and you see the bowl of grapes on the table, but had you arrived a few minutes later, you'd have seen the fake grapes.

This is a perceptual variation of FRANKFURT CLOCK.<sup>14</sup> To capture the knowledge verdict by appeal to safety, advocates of the conciliatory strategy must specify the belief-forming method in a fine-grained and factive fashion, and then include the *real* grapes in the relevant method: if the method is specified as merely seeing grapes, the belief turns out false in nearby worlds. But once the same belief-forming method is operating in another perceptual case like FAKE BARNS, then barn beliefs will also be trivially safe, and the two cases of interest are collapsed. In fairness, I should note that the advocate of the conciliatory strategy may still point out another principled difference between the two cases, and then individuate belief-forming methods in a fine-grained, narrow and ultimately factive manner in FRANKFURT CLOCK and in a more coarse-grained, non-factive manner in FAKE BARNS. However, absent any such difference and motivation, this individuation of belief-forming methods is clearly ad hoc.

Let's take stock. So construed, safety fails to distinguish between the two cases: both Russell's and Barney's beliefs are safe, and thus candidates for knowledge. This is especially troublesome once we appreciate that one of the main motivations for adopting a full-blooded safety condition is precisely to explain absence of knowledge in fake barn cases. If this minimal desideratum is not met, safety loses its appeal. Overall, relativizing safety to more fine-grained belief-forming methods does not make progress on the safety dilemma: a successful solution is to be sought elsewhere.

### 3 Environments

Pritchard's approach to Kelp's dilemma is *revisionary*: he defuses Kelp's challenge by sacrificing the knowledge intuition in FRANKFURT CLOCK. Pritchard maintains that such intuition mistakes mere cognitive achievement with knowledge: while Russell forms a true belief on the basis of his cognitive ability (that is, reading a clock), his belief falls short of knowledge courtesy of environmental luck. Pritchard writes:

What is being exhibited in this example is, I would claim, not knowledge at all, but rather a mere cognitive achievement. *In order to see this, one only has to note that this case is essentially a barn-façade-style case in which the agent's true belief is infected by environmental luck.* While nothing intervenes between [Russell's] cognitive ability and his cognitive success—he really does employ his cognitive abilities in order to gain his true belief about the time—he is in a

<sup>14</sup> For the case to go through, I am stipulating that experiences can have high level content: seeing that object *o* is a grape involves the concept GRAPE, which is plausibly a rich/high-level fruit concept. See Siegel (2021: Sect. 4.3 and Sect. 6) for relevant discussion. Thanks to an anonymous reviewer for urging me to clarify this point.

*very unfriendly environment from an epistemic point of view.* (Pritchard, 2009, p. 40; emphasis mine)

In what follows, I take issue with Pritchard's claim that the agent in FRANKFURT CLOCK is in an unfriendly environment from an epistemic point of view. I will argue that there's an important modal difference between *actually* unfriendly environments and only *potentially* unfriendly environments. While the former may be incompatible with knowledge, we have little reason to think that also the latter are. Since it will be crucial also to the refined formulation of safety that I outline in the next section, this modal difference between the two environments is worthy of special attention.

To get a better purchase on Pritchard's diagnosis, let's focus on environmental luck. In FAKE BARN, (veritic) environmental luck makes Barney's belief only luckily true for reasons having to do with specific *actual* features of the environment. The belief is true but it could easily have been false because of the environment in which it was formed (more precisely, because of the many papier-mâché copies of the real barn). Elsewhere too, Pritchard assumes that knowledge and environmental luck are incompatible (e.g., Pritchard, 2009, p. 38; Pritchard, 2012, p. 273).

If FRANKFURT CLOCK is also a genuine case of environmental luck, then it surely is a very unusual one. Canonical instances of environmental luck correspond to *actually* unfriendly environments: the possibility of error is not abstract, but it's rather concretely realised in some specific feature of the environment under consideration. We see no such feature in FRANKFURT CLOCK. Since the demon *could* but ultimately *does not* intervene, the environment is fine from an epistemic point of view. Granted, the environment *could* have been unfriendly, the demon *could* have intervened. However, absent such intervention, the environment is at best *potentially* unfriendly, not *actually* unfriendly. There is an overlooked but important modal difference between the environment in FAKE BARN and the environment in FRANKFURT CLOCK: while only the former is actually unfriendly from an epistemic point of view, the latter is not, it only could have been.<sup>15</sup>

This modal difference is starker when we consider other clear instances of environmental luck. Take Kvanvig's Comanche case. While the debate is different, the point to drive home is the same: *actually* unfriendly environments are incompatible with epistemic states like knowledge or understanding. Let's recall Kvanvig's case and pay special attention to its *structure*:

**COMANCHE.** Suppose you pick up a textbook on Native American History and read through a chapter documenting the Comanche dominance of the southern plains, until eventually you seem genuinely to understand why the Comanches dominated the southern plains. But suppose as well that while the book you happened to pick up is accurate, *most other books on this topic are full of errors.* If you had picked up one of these other books instead (and we can imagine that *they are all within easy reach!*), your beliefs about the Comanches would have been almost entirely false. (Kvanvig, 2003, pp. 197, 198; emphasis mine)

<sup>15</sup> Kelp (2018, p. 52) briefly notices this difference, but he doubts that it helps the safety theorist to meet his challenge. As I argue in the next section, this difference has far-reaching consequences for safety-based diagnoses of the cases.



Bracketing whether it is convincing, Kvanvig's case of environmental luck also features an *actually* unfriendly environment (*there are* many inaccurate books on the shelf) rather than a mere *potentially* unfriendly environment (there *could have been* inaccurate books on the shelf). The point is general, and applies to further clear-cut instances of environmental luck.<sup>16</sup> To deny knowledge in virtue of environmental luck, Pritchard needs the following version of Kelp's Frankfurt case:

**FAKE CLOCKS.** At 8:22, Russell comes down the stairs and checks the time by looking at one of the many clocks in the room. While most of them are *actually* stopped, Russell happens to look at the only properly functioning clock. As a result, he forms the true belief that it's 8:22.

I take this to be a more compelling instance of environmental luck. The environment is epistemically unfriendly because of an *actual* live possibility of error. At the same time, it's also clear that FAKE CLOCKS and FRANKFURT CLOCK are not structurally similar, and Pritchard needs such structural similarity to vindicate his no-knowledge verdict.

I have argued that the environment in FRANKFURT CLOCK is not a genuine instance of environmental luck; now I want to focus on why this kind of environment is after all compatible with knowledge. I have previously dubbed it 'potentially-unfriendly-but-actually-friendly'. Here's an explication to make further sense of it: we can think of the environment in FRANKFURT CLOCK as an instance of *higher-order environmental luck*. This is because the agent is merely lucky to not be unlucky: they find themselves in an environment that *could* have been unfriendly, but *actually* isn't. We see this difference at work by comparing the standard version of FAKE BARN with this crucially different variation of the case, which I slightly readapt from Carter (2016, p. 5):

**FAKE BARN\*.** Barney\* is driving in the countryside, and hits a fork in the road. If he goes left, he enters a normal epistemically friendly environment. If he goes right, he enters Fake barn country. He luckily takes a left, sees a barn and forms the belief that there is a barn.

This is a more explicit case of *higher order* environmental luck—a type of luck that doesn't undermine knowledge. After all, higher-order environmental luck is just an instance of *benign* luck. Barney\* may be lucky to be in a friendly environment and thus in a position to know that there is a barn. However, his belief is not true courtesy of luck. Similarly, the environment in FRANKFURT CLOCK is also affected by higher-order environmental luck. Russell may be equally lucky to have stepped into a friendly environment and to be in a position to know the time, but this doesn't prevent him from coming to know the time. These considerations tell against Pritchard's assessment: FAKE BARN and FRANKFURT CLOCK display a relevant modal difference in their respective environments. The former environment undermines knowledge, but the latter doesn't.

<sup>16</sup> Sosa (2010, p. 471) offers yet another case in point: "Take a subject who drinks from a cup out of several available on a table. *All the other cups*, let us say, contain a drug that much degrades 'subitizing' ability." This further instance of environmental luck also features an *actually* unfriendly environment (the cup *actually* contains the drug) rather than a mere *potentially* unfriendly environment (the cup *could have* contained the drug).

A quick summary of the cases covered so far will be helpful. First, to treat FRANKFURT CLOCK on a par with FAKE BARN, Pritchard needs a different scenario featuring an *actually* unfriendly environment. I have offered such scenario in FAKE CLOCK. The upshot is that FRANKFURT CLOCK and FAKE BARN are not analogous. Second, I have argued that FRANKFURT CLOCK is more similar to FAKE BARN\* due to a common instance of higher-order environmental luck, which is compatible with knowledge. *Pace* Pritchard, if the agent lacks knowledge in FRANKFURT CLOCK, it's not because of knowledge-undermining environmental luck. Pritchard's diagnosis is inadequate.

At this point, it's also worth pausing to set the record straight on the dialectical situation. Merely appreciating the inadequacy of Pritchard's diagnosis doesn't quite let the safety theorist off the hook yet. One fundamental question remains: does the difference in the two environments help the safety theorist to capture the verdicts of knowledge and ignorance in FRANKFURT CLOCK and FAKE BARN? Once we incorporate this difference into the general formulation of safety, the answer is 'yes'. In fact, the two vignettes display two crucially different environments. Forming a true belief on the basis of a good method (reading from a clock) in a *potentially-unfriendly-but-actually-friendly* environment (FRANKFURT CLOCK) is not the same as forming a true belief on the basis of a good method (visual perception) in an *actually unfriendly* environment (FAKE BARN). In the next section, I reformulate the safety condition in light of this difference and I show how it successfully meets Kelp's challenge.

#### 4 Environment-relative Safety

This is my proposed modification to the safety condition on knowledge:

**ENVIRONMENT-RELATIVE SAFETY.** In most or all close possible worlds in which S believes that  $p$  via the same method of belief formation  $M$  that S uses in the actual world (**sub-condition M**) *and* S occupies the same environment  $E$  that S occupies in the actual world (**sub-condition E**),  $p$  is true.<sup>17,18</sup>

Several ideas in this refined formulation of safety deserve unpacking. In what follows, I will:

- (a) Explain and motivate the *conjunctive* formulation of **Environment-relative Safety**.

<sup>17</sup> Readers might have qualms about my use of 'same'. In response, I offer the following remarks. First, methods and environments are not *literally identical*, but only *relevantly similar* (compare: different tokens of the same type are relevantly similar but not literally identical to each other). Second, modal epistemologists notoriously struggle to identify satisfactory criteria of individuation of methods and environments across relevantly close world. Some authors have even suggested that said criteria ultimately rest on a modal fallacy (Hetherington 2013, pp. 166–167). While I'm aware of the problems with the notion of similarity and with how to individuate methods and environments in nearby worlds, I also notice that they affect *any* modal condition on knowledge and do not bear directly on the formulation of safety on offer here.

<sup>18</sup> This safety condition is importantly different from environment-relative *ability* conditions that further entail a safety condition (e.g., Greco 2020a: Sect. 3). The focus here is exclusively on safety and environments, not on abilities. See also Mortini (2022: Sect. 3) for more discussion on **Environment-relative Safety**.

- (b) Clarify how **Environment-relative Safety** provides a satisfactory solution to the safety dilemma and also deals with further variations of FAKE BARN.
- (c) Show that, despite some important limitations, **Environment-relative Safety** is preferable to the conciliatory and revisionary strategies discussed above.

#### 4.1 Safety as a conjunctive condition

To bring out the distinctiveness of my formulation of safety, it might be useful to start by comparing it with other versions of the principle. For instance, Williamson's safety condition reads as follows: "if one knows, one could not have easily been wrong in a similar case" (Williamson, 2000, p. 147). Pritchard (2007, p. 281) writes: "S's belief is safe iff in most nearby 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". Finally, Sosa (1999, p. 146): "If S were to believe that  $p$ ,  $p$  would be true". Two points to note here. First, all these formulations are rather simple: instead of a *conjunction* of conditions, they feature only *one* condition. Second, they are relativized, more or less explicitly, to methods or bases *only* (e.g., Williamson, 2000, p. 128; Pritchard, 2008, p. 446; Sosa, 2007, p. 26).<sup>19</sup> Differently, we can think of **Environment-relative Safety** as the *conjunction* of two sub-conditions: just as a conjunction is true if and only if both conjuncts are true, so **Environment-relative Safety** is satisfied if and only if *both* sub-conditions are met. The first, **sub-condition M**, is familiar: safety theorists have already noted the importance of keeping the belief-forming method fixed. The second, **sub-condition E**, is new: it comprises environments, and it has received comparatively less attention. Crucially, **sub-condition E** expands the set of factors that determine which worlds count as relevantly close. As we have seen, these factors standardly include the subject S, the proposition P, the time T, and the method M: {S, P, T, M}. In the revised version, we keep also the environment E fixed, and enrich the set accordingly: {S, P, T, M, E}. This addition distinguishes our two cases of interest, and, far from being ad hoc, it also better captures the spirit of the safety condition on knowledge. Safety is supposed to explicate the sense in which knowledge equals non-accidentally true belief. Such 'non-accidentality' has to do with the modal profile of *both* the method employed *and* the environment in which it is employed. To fully appreciate this point, consider a non-epistemic case of safety. For example, if we aim to assess whether a car is safe, we will make sure to drive it in a suitably specified environment (say, in appropriate driving conditions). In general, judgements on safety tacitly assume environmental factors, so it's worth making these factors explicit in the formulation of the safety condition on knowledge. To count as safe, a true belief must meet each of the two sub-conditions: if it meets just one, then it fails to constitute knowledge. This is the key to solve the safety dilemma, as I shall explain next.

<sup>19</sup> Here, I am following Ranalli (2014, p. 230, fn. 12) for a similar (and helpful) list of formulations of the safety principle. See also Blome-Tillman (2020, p. 36) for an appendix with another list of formulations of safety.

## 4.2 A new solution to the safety dilemma

**Environment-relative Safety** not only accommodates presence of knowledge in FRANKFURT CLOCK, but it also captures the ignorance intuition in FAKE BARN. Let's take each in turn. Presence of knowledge in FRANKFURT CLOCK: if we keep the method (competently reading from a clock) and also the *potentially* unfriendly but *actually* friendly environment fixed, then Russell's belief is *safe*. In the close worlds where Russell competently reads from a clock and the demon doesn't intervene on the environment, the belief is true and satisfies **Environment-relative Safety**. Both sub-conditions E and M are met: knowledge is present. On to ignorance in FAKE BARN: if we keep the method (visual perception) and also the *actually* unfriendly environment fixed, then Barney's belief is *unsafe*. In the close worlds where Barney continues to occupy an environment filled with fake barns, the belief is false and fails to satisfy **Environment-relative Safety**. Sub-condition E isn't met: the environment is and remains unfriendly, so knowledge is absent.<sup>20</sup>

This is the relevant difference between the two cases, and now safety can detect it. In FRANKFURT CLOCK, the error possibility requires a change in the environment: for Russell to form a false belief, the demon has to intervene. In FAKE BARN, no such change is needed. The error possibility is *already* realised in the environment: because of the *actual* papier-mâché copies of the real barn, Barney's belief could too easily be false. So construed, safety rises to Kelp's challenge.

However, this formulation of safety incurs further difficulties when it comes to individuate the relevant environments more precisely.<sup>21</sup> For example, indexing safety to the agent's environment *and* the agent's actual position in that environment may yield the result that the agent knows in fake-barn cases: if we restrict the safety condition to all the worlds where Barney's environment is exactly as it and where he's located where he's actually located, these are all worlds where Barney is looking at the only real barn. Barney seems to have knowledge on such safety principle, and this is bad news for the safety condition I am offering here.

One moral to draw from this type of case is that, just like belief-forming methods, environments should *not* be individuated too narrowly. Barney's actual environment

<sup>20</sup> This point brings out a key difference between this formulation of safety and Sosa's aptness condition, which also stresses the importance of environmental factors. A belief is apt if and only if it's accurate because adroit (competent) and for the adroitness to explain the accuracy, the believer's exercise must succeed in a way that overcomes the influence of certain environmental interferences (for instance, Gettier-style deviant causal chains). However, Sosa's aptness condition remains compatible with *some* environmental interferences: in FAKE BARN, the agent's belief comes out as apt and constitutes (animal) knowledge (Sosa 2010, pp. 473, 2015, p. 81). *Despite* the close possibility of an environmental interference, their belief is accurate (true) because it's adroit (competent). Differently, and unlike Sosa's aptness condition, **Environment-relative Safety** yields a no-knowledge verdict in FAKE BARN: *precisely* in virtue of the close possibility of an environmental interference, the agent's belief comes out as unsafe and thus fails to constitute knowledge. When it comes to environmental interferences, the safety condition defended here is more demanding than Sosa's aptness condition, which at best entails a *weaker* type of safety relative only to the agent's abilities (see Ranalli 2014, p. 1232 and Greco 2020b, pp. 5152–5154 for discussion). Crucially, my formulation of safety focusses on environments directly, bypasses abilities entirely and thus differs from Sosa's aptness condition. Thanks to an anonymous reviewer for pushing me to explain this difference.

<sup>21</sup> Special thanks to an anonymous reviewer for raising and pressing me on these cases.

is *broadly* misleading: while it does include one real barn, it is more generally constituted by one real barn *amidst many fake barns*. If we keep such broadly misleading environment fixed, in these close worlds he ends up forming a false belief because in addition to the real barn there's many more fake barns. A knowledge verdict trades on a too narrow individuation of belief-forming methods (Barney forms his belief on the basis of seeing the *real* barn) but also on an equally narrow individuation of the relevant environment (Barney's location in the *only* portion of fake barn country where the real barn is visible). As I've argued already, neither of these approaches is ultimately correct.

It's important to be clear on the limits of this proposed safety condition, and testing it against an even harder case does so in an especially vivid way. Imagine a variation of FAKE BARN in which Barney travels through different regions: some are genuine barn regions and some are fake barn regions. Does Barney have knowledge in genuine barn region? Does he lack knowledge in fake barn region? And what about borderline cases where he's still in genuine barn region and yet very close to fake barn region? This variation of FAKE BARN raises a difficult question: which specific concept of environment can suit Environment-relative Safety? Crucially, providing an informative answer to this question turns out to be very challenging.

In order to make progress on this question, we might initially try to equate environments with the spatio-temporal regions of the actual world where the agent is located and then keep these regions fixed across close possible worlds.<sup>22</sup> With this conception of environments at hand, we could argue as follows: given Environment-relative Safety, Barney has knowledge as long as he's located in an actually friendly environment even if he's close and could easily have ended up in an epistemically unfriendly and misleading environment. But this way of identifying environments is deeply problematic because it fails to capture important intuitions concerning the modal closeness of relevant error possibilities. In fact, when Barney merely gets too physically close to the unfriendly environment it looks like he automatically counts as being in the unfriendly environment even if he hasn't crossed the physical border and he's not yet in the physical region filled with fakes. As such, this concept of environments is ill-suited for the safety condition on offer here. The crucial question pertaining to the relevant concept of environment still remains, but what more can be said in answer to this question? While there seems to be at least two available options, neither is fully adequate. Let's briefly review each option in turn.

The first option takes the relevant concept of environment as primitive: perhaps we have a pre-theoretic, context-sensitive concept of environment that we tacitly appeal to when making judgements about cases of knowledge in various environments. However, since it doesn't say much on the key features of knowledge-conducive environments, this line of response is not very informative, and, accordingly, not very satisfactory.

<sup>22</sup> This concept of environment draws inspiration from the closely related notion of *local reliability* of belief-forming methods: roughly, a belief-forming method is more or less locally reliable depending on the specific spatio-temporal regions in which the agent is located while employing such method. See Goldman (1986, p. 45, 2008, p. 91), Goldberg (2010, p. 52), Graham (2014) and Bernecker (2020: Sect. 2) for a thorough discussion of the notion of local reliability.

Building on Williamson's circular version of safety, the second line of response consists in providing a closely related circular characterisation of environments: just like we use intuitions about knowledge to guide determinations of safe belief (Williamson, 2009, p. 305), we use our intuitions about knowledge also to guide determinations of what counts as an environment suitable for knowledge. The circularity involved here is not obviously vicious: it just takes the concept of environment suitable for knowledge as being dependent on the concept of knowledge. Perhaps in some key cases we can give only a somewhat circular answer to the question of whether the agent's environment is suitable for knowledge: this is because we first have to answer the question of whether the agent knows, and then use it to that to say something more informative about the relevant features of their environment.<sup>23</sup> While it still leaves important questions open, this circular characterisation of environments can initially suit the safety condition offered here.

However, neither the primitive nor the circular option seem fully satisfactory: more work needs to be done to see whether each option can deal with further and potentially more challenging cases.<sup>24</sup> Taken together, these considerations suggest that **Environment-relative Safety** inevitably runs into some version of the generality problem raised against the individuation of environments instead of belief-forming methods: it is difficult to provide a sufficiently specific and ultimately informative account of environments that suits this version of safety. This variation of the generality problem is certainly pressing, but it doesn't affect the solution to the safety dilemma on offer: even without a fully worked out conception of suitable environments for knowledge, we can still register an important environmental difference between the two cases discussed here and use this difference to address Kelp's challenge. Giving a more precise account of what exact features an environment must display in order to be compatible with knowledge (perceptual or otherwise) is a task best left for another occasion. For now, we can more modestly focus on how **Environment-relative Safety** fares when compared to the other solutions offered in response to the safety dilemma. I do so in the remainder of the section.

### 4.3 Theoretical Virtues

Importantly, **Environment-relative Safety** does better than the conciliatory and revisionary strategies discussed above. First, unlike the conciliatory solution, **Environment-relative Safety** does not have to specify the belief-forming method in a too specific fine-grained way. Instead of focussing on a *properly functioning* clock and collapsing the two cases of interest, the relevant method is simply Russell's *general* ability to read a clock. Second, unlike the revisionary solution, **Environment-relative Safety** is not committed to a dubious conception of environmental luck, and it straightforwardly captures the correct intuitions of knowledge and ignorance in each case.

<sup>23</sup> This circular account is not necessarily a concession to the knowledge-first programme: knowledge can still be a type of safe belief formed in a suitable environment even if we ultimately need to consult the concept of knowledge to say something more informative on the features of such environment.

<sup>24</sup> Special thanks to an anonymous reviewer for very helpfully suggesting these two options.

A further index to both methods and environments makes great progress on the safety dilemma. While I acknowledge that such additional index incurs substantial difficulties in individuating the relevant environments, I also maintain that it's the most promising strategy to rescue the safety condition from Kelp's troubling conjunction of cases.<sup>25</sup>

## 5 Meeting the challenge

By conjoining FRANKFURT CLOCK with FAKE BARN, Kelp raises a thorny dilemma for safety-theoretic accounts of knowledge. He draws a pessimistic conclusion: no formulation of safety can be weak enough to grant knowledge in the former and strong enough to predict ignorance in the latter. The shortcomings of the conciliatory and revisionary solutions further confirm Kelp's conclusion: since they fail to capture the right verdict in one of the two cases, it seems that the safety theorist has no choice but to sacrifice one of these intuitively correct verdicts.

Yet, appearances are misleading. There's a crucial difference between a *potentially* unfriendly environment (FRANKFURT CLOCK) and an *actually* unfriendly environment (FAKE BARN). This difference requires to index the safety condition not only to belief-forming methods, but also to environments, and **Environment-relative Safety** does just that. True enough, this new version of safety is not without problems and leaves many important questions open: crucially, **Environment-relative Safety** fails to individuate environments in a precise manner and it doesn't fully clarify the range of environments in which agents can fall prey to epistemic Frankfurt cases. However, **Environment-relative Safety** also provides a principled way out of Kelp's dilemma: in fact, it can predict knowledge in FRANKFURT CLOCK and ignorance in FAKE BARN. To do away with the safety condition on knowledge we may have to wait another day.<sup>26</sup>

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

<sup>25</sup> I also flag that **Environment-Relative Safety** is not the only available solution to Kelp's dilemma. Beddor and Pavese's (2020) and Wedgwood's (2020) normality approaches also provide a principled distinction between the environment in FRANKFURT CLOCK (absent any demon intervention, quite normal) and the environment in FAKE BARN (since it's filled with mere barn façades, quite abnormal). However, since it's not wedded to the admittedly murky concept of normality, **Environment-relative Safety** is theoretically simpler, and thus preferable. Moreover, as Kelp (Manuscript) points out, these normality accounts still face a slightly different version of the safety dilemma.

<sup>26</sup> I would like to thank two very conscientious reviewers for having the patience to provide helpful comments to previous versions of the manuscript. Thanks to Adam Carter, Jack Lyons, Domingos Faria, Lilith Newton and Charlotte Connell for useful discussion. Special thanks to Chris Kelp for a great deal of support and advice. This project has received generous funding from the Scottish Graduate School of Arts and Humanities.

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Baumann, P. (2008). Is knowledge safe? *American Philosophical Quarterly*, 45(1), 19–30.
- Beddor, B., & Pavese, C. (2020). Modal virtue epistemology. *Philosophy and Phenomenological Research*, 101(1), 61–79.
- Bernecker, S. (2020). Against global method safety. *Synthese*, 197(12), 5101–5116.
- Blome-Tillmann, M. (2020). Non-reductive safety. *Belgrade Philosophical Annual*, 33, 25–38.
- Bogardus, T. (2014). Knowledge under threat. *Philosophy and Phenomenological Research*, 88(2), 289–313.
- Brown, J. (2000). VI-reliabilism, knowledge, and mental content. *Proceedings of the Aristotelian Society*, 100(2), 115–135.
- Carter, J. A. (2016). Robust virtue epistemology as anti-luck epistemology: a new solution. *Pacific Philosophical Quarterly*, 97(1), 140–155.
- Colaço, D., Buckwalter, W., Stich, S., & Machery, E. (2014). Epistemic intuitions in fake barn thought experiments. *Episteme*, 11(2), 199–212.
- Goldberg, S. (2010). *Relying on others: An essay in epistemology*. Oxford University Press.
- Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge: Harvard University Press.
- Goldman, A. I. (2008). Sosa on reflective knowledge and virtue perspectivism. In J. Greco (Ed.), *Ernest Sosa and his Critics* (pp. 86–95). Oxford: Blackwell.
- Goldman, A. (2009). Williamson on Knowledge and Evidence. In P. Greenough & D. Pritchard (eds.), *Williamson on Knowledge*. Oxford University Press. pp. 73–91.
- Goldman, A. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy*, 73(November), 771–791.
- Graham, P. (2014). Against transglobal reliabilism. *Philosophical Studies*, 169, 523–535.
- Greco, J. (2020a). Knowledge producing abilities. *Virtue-Theoretic Epistemology: New Methods and Approaches*. Cambridge: Cambridge University Press
- Greco, J. (2020). Safety in sosa. *Synthese*, 197(12), 5147–5157.
- Williamson, T. (2009). Reply to Goldman. In D. Pritchard & P. Greenough (eds.), *Williamson on Knowledge*. Oxford: Oxford University Press. pp. 305–312.
- Grundmann, T. (2018). Saving safety from counterexamples. *Synthese*, 197(12), 5161–5185.
- Hawthorne, J. (2004). *Knowledge and Lotteries*. Oxford University Press.
- Hetherington, S. (2013). Knowledge can be lucky. In M. Steup & J. Turri (eds.), *Contemporary Debates in Epistemology*. Blackwell.
- Hiller, A., & Neta, R. (2007). Safety and epistemic luck. *Synthese*, 158(3), 303–313.
- Hirvelä, J. (2019a). Knowing without having the competence to do so. *Thought A Journal of Philosophy*, 8(2), 110–118.
- Hirvelä, J. (2019b). Global safety: How to deal with necessary truths. *Synthese*, 196(3), 1167–1186.
- Kelp, C. (2017). Knowledge first virtue epistemology. In Adam Carter, Emma Gordon & Benjamin Jarvis (eds.), *Knowledge First: Approaches in Epistemology and Mind*. Oxford University Press.
- Kelp, C. (2018). *Good Thinking. A Knowledge First Virtue Epistemology*. London, Routledge.
- Kelp, C. (2009). Knowledge and Safety. *Journal of Philosophical Research*, 34, 21–31.
- Kelp, C. (2016). Epistemic frankfurt cases revisited. *American Philosophical Quarterly*, 53(1), 27–37.
- Kvanvig, J. L. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press.
- Lycan, W. G. (2006). On the Gettier problem problem. In S. Hetherington (Ed.), *Epistemology Futures* (pp. 148–168). Oxford University Press.
- Miracchi, L. (2015). Competence to know. *Philosophical Studies*, 172(1), 29–56.
- Mortini, D. (2022). The explanationist and the modalist. *Episteme* 1–16.
- Neta, R., & Rohrbaugh, G. (2004). Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly*, 85(4), 396–406.
- Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.
- Pritchard, D. (2005). *Epistemic Luck*. Oxford University Press.
- Pritchard, D. (2007). Anti-luck epistemology. *Synthese*, 158, 277–297.



- Pritchard, D. (2009). Safety-Based Epistemology: Wither Now?. *Journal of Philosophical Research*, 34, 33–45.
- Pritchard, D. (2008). Sensitivity, safety, and anti-luck epistemology. In J. Greco (ed.), *The Oxford Handbook of Skepticism*. Oxford University Press.
- Pritchard, D. (2012). Anti-luck virtue epistemology. *Journal of Philosophy*, 109(3), 247–279.
- Hetherington, S. (2013). Knowledge cannot be lucky. In M. Steup & J. Turri (eds.), *Contemporary Debates in Epistemology*. Blackwell.
- Pryor, J. (2004). Comments on Sosa's "relevant alternatives, contextualism included." *Philosophical Studies*, 119(1–2), 67–72.
- Rabinowitz, D. (2011). The safety condition for knowledge. *Internet Encyclopedia of Philosophy*.
- Ranalli, C. (2014). Luck, propositional perception, and the entailment thesis. *Synthese*, 191(6), 1223–1247.
- Sainsbury, R. M. (1997). Easy possibilities. *Philosophy and Phenomenological Research*, 57(4), 907–919.
- Siegel, S. (2021). The Contents of Perception. *The Stanford Encyclopedia of Philosophy (Fall 2021 Edition)*, In E. N. Zalta (ed.). <https://plato.stanford.edu/archives/fall2021/entries/perception-contents/>.
- Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives*, 13, 137–149.
- Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge*. Oxford University Press.
- Sosa, E. (2010). How competence matters in epistemology. *Philosophical Perspectives*, 24(1), 465–475.
- Sosa, E. (2015). *Judgment & Agency*. Oxford University Press UK.
- Sosa, E. (2017). *Epistemology*. Princeton University Press.
- The Safety Dilemma Revisited. Manuscript.
- Wedgwood, R. (2020). The internalist virtue theory of knowledge. *Synthese*, 12, 1–22.
- Whitcomb, D. (2008). Factivity without safety. *Pacific Philosophical Quarterly*, 89(1), 143–149.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.