CCiL
ciència cognitiva llenguatge

MASTER IN COGNITIVE SCIENCE AND LANGUAGE

MASTER THESIS

June 2023

Modalized Robust Virtue Epistemology Defended

by Marc Lara Crosas

Under the supervision of:

Fernando Broncano Berrocal

UNIVERSITAT DE BARCELONA

UAB Universitat Autònoma de Barcelona

Universitat de Girona

upf. Universitat Pompeu Fabra Barcelona

UNIVERSITAT ROVIRA i VIRGILI

Abstract: This essay addresses one of the most central and longstanding problems in epistemology: What is the nature of knowledge? According to robust virtue epistemology, which is a popular answer to this problem, knowledge is explained solely by appealing to cognitive abilities, epistemic competences, or intellectual virtues. In this essay, I defend a reliabilist and modalized version of robust virtue epistemology. More specifically, the main thesis of this essay is that Mortini's (2022) reformulation of the safety principle as "environment-relative safety", which is the best response to Kelp's (2009; 2016; 2018) safety dilemma, makes robust virtue epistemology even more plausible. In section 2, after the introduction, I will present Pritchard's defense of a modest (non-robust) virtue epistemology based on the independence thesis. In section 3, Kelp's (2009; 2016; 2018) safety dilemma, which is an objection to the necessity of safety for knowledge, is presented. However, I will argue that Mortini's (2022) reformulation of safety satisfactorily answers this objection and that it is the best answer to this dilemma thus far in the literature. Finally, in section 4, I argue that the manifestation of cognitive abilities ought to be relativized to actual features of the environment. And I will argue that the satisfaction of the ability condition relativized to those features entails the satisfaction of environment-relative safety.

Keywords: Virtue Epistemology, Environment-Relative Safety, Fake Barn Case, Cognitive Achievement.

Acknowledgments

This master's thesis has been supported by "Beca màster + UB", and additional funding from Banco Santander: "Ajuts complementaris programa màster + UB". I would like to thank Fernando Broncano Berrocal for his helpful comments and suggestions, and for accepting to be my supervisor so I could ask for funding. I would also like to thank Dario Mortini for his helpful comments and suggestions. And thanks to Julianna Farkas for telling me the Hungarian version of "Jennifer".

Even after quite a few years studying philosophy, my parents still do not understand what all this is about, but still support me unconditionally. So, special thanks to them. Many thanks to the Skittle squad: Àlex, Felipe, Lucía, and Kristina for including me in their community, listening to my crazy theories, and pretending that my jokes are funny. To my old friends, Marcel, Aina, Clara, Arnau, and Mireia, who will always be my first-class friends, many thanks are due.

Table of contents

1. Introduction

This essay addresses one of the most fundamental questions in contemporary epistemology: what is the nature of knowledge? I will argue that the best answer to this question is given by robust virtue epistemology (henceforth "RVE"), which explains knowledge solely by appealing to the exercise of intellectual virtues, epistemic competences, or cognitive abilities. More specifically, I will defend a virtue reliabilist and modalized version of RVE. The main thesis of this essay is that Mortini's (2022) reformulation of the safety principle as "environment-relative safety", which is the best answer to Kelp's (2009; 2016; 2018) safety dilemma, makes RVE even more plausible.

First of all, I will introduce the three accounts of knowledge that I consider in this essay: (i) safety-based accounts, (ii) RVE, and (iii) impure virtue epistemology. The moral that can be drawn from Gettier (1963) cases is that a belief being luckily true is incompatible with knowledge, and this knowledge-undermining luck is labeled as "veritic luck" by Pritchard (2007: 286). Besides, he distinguishes two kinds of "veritic luck": "intervening luck", as in typical Gettier-style cases, where something interferes between the subject and the fact that makes true her belief, and "environmental luck", such as in the fake barn case (see section 2), where there is no actual interference between the method of belief-formation and the relevant fact. Thus, it is tempting to add an anti-luck condition such as safety (see below) in our analysis of knowledge to exclude cases of veritic luck as cases of knowledge, and thereby solve the Gettier problem (i.e., the problem to explain why justified true beliefs in Gettier-style cases do not qualify as knowledge).

The safety principle has been notoriously defended as a necessary condition for knowledge by Sosa (1999), Williamson (2000), and Pritchard (e.g., 2009b), with significant differences. However, I will focus on Pritchard's formulation and defense of the safety principle as an anti-luck modal condition because I will argue against Pritchard's account. A belief is safe if and only if it could not have been easily false. This expression, "could not have been easily false", is further expanded in terms of possible worlds because safety is a modal condition. The method of belief-formation is usually held fixed across possible worlds and distance among possible worlds is based on similarity (Lewis 1986). More precisely, a belief is safe if and only if in most close possible worlds (if not all) in which the subject forms the same belief as in the actual world employing the same method of belief-formation as in the actual world, the belief is true (Pritchard 2007: 280).

Another possible anti-luck modal condition is the sensitivity principle: A subject, S, knows that p only if, were p false, S would not believe that p. Nonetheless, I will not consider sensitivity for it suffers from various problems: (i) it is incompatible with the closure principle, (ii) the denials of the skeptical scenarios are insensitive, therefore they cannot be known, and (iii) it struggles to accommodate inductive knowledge (see Pritchard 2013: 154-5).

The basic thesis of virtue epistemology is that knowledge is the result of intellectual virtues, epistemic competences, or cognitive abilities. Typically, two kinds of virtue epistemologies are distinguished: virtue responsibilism and virtue reliabilism. While the former takes intellectual virtues as stable character traits that contribute to the "flourishing" of the agent (Zagzebski 1996), the latter takes them as dispositions that reliably produce true beliefs. I will focus on virtue reliabilism, according to which, knowledge is a kind of cognitive achievement in which the cognitive success (i.e., the true belief)[1] is *sufficiently due to* cognitive ability. This relation, "sufficiently due to", can be interpreted either as the manifestation of competence following Sosa (e.g., 2015), as it will be endorsed in section 4, or as partial or full creditability following Pritchard (2012) and Greco (2003) respectively, as considered in section 2. In short, virtue epistemology endorses the ability condition as a necessary condition for knowledge. RVE explains knowledge solely in terms of the ability condition: the ability condition is not only necessary but also sufficient for knowledge (e.g., Sosa 2015; Greco 2010; Turri 2011).

> **Virtue epistemology:** S knows that p *only if* the cognitive success is sufficiently due to cognitive ability.
>
> **Robust virtue epistemology:** S knows that p *if and only if* the cognitive success is sufficiently due to cognitive ability.

RVE is a simple and appealing account of knowledge because, for instance, it plausibly solves the "value problem": why knowledge is more valuable than mere true belief? Roughly, since knowledge is a kind of achievement according to RVE and achievements are themselves valuable, knowledge is valuable (e.g., see Greco 2010). RVE cannot include modal principles such as safety or sensitivity, for they are conditions about the modal profile of beliefs. By contrast, virtue-theoretic conditions available to RVE are about the cognitive abilities or virtuous profile of epistemic agents. Nonetheless, there is a third way: impure virtue epistemology, which may include both modal conditions such as safety or sensitivity, and virtue-theoretic conditions.

[1] I assume that cognitive success is equivalent to true belief. By contrast, Kelp (2018) defends that knowledge is the relevant kind of cognitive success, for he defends a knowledge-first version of virtue reliabilism.

In section 2, I will present Pritchard's "anti-luck virtue epistemology" (henceforth "ALVE"), which is an impure virtue epistemology and the best vindication found in the literature of the independence thesis. According to the independence thesis, there are two necessary conditions for knowledge, the safety principle, and the ability condition; and the satisfaction of either condition does not entail the satisfaction of the other. In section 3, I will introduce Kelp's (2009; 2016; 2018) safety dilemma which is an objection to the necessity of the safety principle. After that, I will argue that Mortini's (2022) reformulation of safety as "environment-relative safety" is the best answer to this dilemma thus far in the literature. Besides, this reformulation of safety can accommodate other problematic cases for this modal condition.

In section 4, (i) I argue that the manifestation of cognitive abilities ought to be relativized to actual features of the environment in a way that has not been appreciated so far. In particular, actual features of the environment that do not actually intervene may preclude the manifestation of the cognitive ability, even if the subject forms a true belief. For instance, it will be argued that the mere presence of deceiving objects (e.g., fake barns) is sufficient to preclude the subject's cognitive success from being sufficiently due to cognitive ability. (ii) There is a kind of cases that can be thought of as counterexamples to this idea: cases in which, intuitively, an achievement is accomplished despite the actual presence of misleading or deceiving objects. Even if this intuition is compelling and plausible, I will argue that these cases do not constitute genuine achievements. (iii) Finally, contra Pritchard's independence thesis, I will argue that the ability condition relativized to actual features of the environment entails the satisfaction of environment-relative safety, which is the best response to Kelp's safety dilemma. In short, the aim is to show that neither ALVE, safety-based accounts, nor RVE, solve Kelp's safety dilemma. As a result, I will strengthen the environmental component of RVE to deal with these cases.

2. The independence thesis

In this section, I will present Pritchard's (2009a; 2012) defense of his "anti-luck virtue epistemology", which is the best vindication or motivation for the "independence thesis" (Carter 2013) offered in the literature. According to this thesis, two conditions are necessary for knowledge, an anti-luck condition and an ability condition; yet the satisfaction of neither of them entails the satisfaction of the other.

> **Independence thesis:** the satisfaction of the anti-luck condition does not entail the satisfaction of the ability condition, and the satisfaction of the ability condition does not entail the satisfaction of the anti-luck condition.

Pritchard (2012: 247-248) claims that our understanding of the concept of knowledge is governed by two intuitions: the anti-luck intuition and the ability intuition. For the sake of the argument, I will assume these two intuitions without arguing for their existence. On the one hand, the former (the anti-luck intuition) is the platitude that coming to believe a true proposition by luck is incompatible with knowledge, as Gettier-style cases show; on the other hand, the latter consists of the intuition that knowledge is the result of cognitive ability. Both are taken to be platitudes about knowledge that every account of knowledge ought to "accommodate" or "satisfy": no account should predict that luckily true beliefs (e.g., beliefs formed by wishful thinking) qualify as knowledge, and no account should predict that true beliefs which are not the result of cognitive ability *in the right way* (see Temp case below) are cases of knowledge. Pritchard defends that safety and the ability condition are both necessary for knowledge and the satisfaction of neither of them entails the satisfaction of the other (i.e., the independence thesis):

> **ALVE:** "S knows that p if and only if S's safe true belief that p is the product of her relevant cognitive abilities (such that her safe cognitive success is to a significant degree creditable to her cognitive agency)." (Pritchard 2012: 273)

To avoid the charge that ALVE is an ad hoc proposal (i.e., a proposal construed merely to avoid certain counterexamples), Pritchard (2012: 274-8) independently motivates that knowledge has a two-part structure by appealing to Craig's (2004) genealogical story about the arising of the concept of knowledge. Craig invites us to imagine a hypothetical situation in which the concept of knowledge has not originated (yet). Still, in this situation, it would be useful to identify good informants: people that, for instance, hold information about where to find important resources such as food or water. In short, a precursor of the concept of knowledge would emerge, and it would evolve into our current concept of knowledge. Most certainly, good informants need to be reliable, but Pritchard argues that reliability is ambiguous in this occasion: it can either mean that the subject's cognitive abilities are reliable for that domain or that it is a subject that we can "rely on" regarding that domain. Pritchard argues that both senses of reliability come apart and these two aspects explain that we have two central independent epistemic intuitions which impose different epistemic conditions (i.e., this ambiguity supports the independence thesis).

2.1 Safety-based accounts

In this section, I will present how Pritchard dismisses safety-based accounts of knowledge by showing that the modal profile required by safety does not secure that the belief is the result of cognitive ability (i.e., safety is not sufficient to satisfy the ability intuition). Thus, while Pritchard accepts safety as necessary for knowledge, he argues that it is not sufficient. First, let's see how safety can deal with a very special Gettier case:

> **Fake barn case:** Robert has robbed a bank and he is looking for a place to hide from the police. He is driving through the fake barn county when he sees a barn, which seems an appropriate place to hide, and forms the true belief that a (real) barn is before him. However, unbeknownst to him, he is in a region full of fake barns that are only façades. Had he seen a fake barn instead, he would have formed the false belief that a (real) barn is before him, given that the fake barns are visually indistinguishable from the real one.[2]

Robert has a true belief perceptually justified that does not amount to knowledge[3]. Safety, as a necessary epistemic condition, gives the right verdict in this case: given that, in most close possible worlds Robert sees a fake barn and forms the false belief that a (real) barn is before him, his belief is unsafe and, for this reason, he lacks knowledge. However, is safety, jointly with truth and belief, sufficient for knowledge? And does safety appropriately accommodate the ability intuition?

> **Temp case:** Temp is a philosophy student worried about the appropriate room temperature for philosophical theorizing. Currently, he is in a room with a malfunctioning thermometer that reads random degrees. Unbeknownst to Temp, his friend is controlling the temperature of the room, and every time Temp looks at the thermometer, his friend changes the temperature of the room to match the reading of the thermometer. Temp looks at the thermometer and forms a true belief about the temperature of the room. (Pritchard 2012: 260)

[2] This is a version of the fake barn case first published by Goldman (1976: 772-3) but attributed to Carl Ginet.

[3] I assume an ignorance verdict in the fake barn case. However, Colaço et al. (2014) show that laypeople tend to attribute knowledge to fake barn cases. And, for instance, Sosa (2010) defends that Robert's success manifests his competence, therefore he accepts that there is knowledge, of the animal kind.

First of all, Temp does not know the temperature of the room, for he is consulting a broken thermometer. Besides, the "direction of fit" of Temp's belief is not appropriate for knowledge: for our beliefs to be knowledge, they need a world-to-mind direction of fit (i.e., the belief should align or be responsive to the world facts) instead of the mind-to-world direction of fit of Temp's belief (i.e., the temperature, the fact, aligns or responds to Temp's beliefs). The cognitive success is not the result of cognitive ability, but rather, the result of the friend's intervention, so the ability intuition is not satisfied. However, his true belief is safe since in most nearby possible worlds, his friend also ensures that Temp's beliefs are true. As long as in most relevant nearby possible worlds the belief is true, safety will be satisfied irrespective of the direction of fit of the belief. Thus, Pritchard concludes that this case reveals that safety, truth, and belief are not sufficient for knowledge and that we need a further epistemic condition to accommodate the ability intuition.

2.2 Robust virtue epistemology

Despite the attractiveness of RVE, Pritchard thinks that it also fails to deal with both epistemic intuitions. According to Pritchard (2012: 267-8), the fake barn case is a difficult case to accommodate by RVE: after all, Robert's cognitive success (i.e., his true belief that a real barn is before him) seems to be sufficiently due to his perceptual abilities. Moreover, by hypothesis, his perceptual abilities are exercised in a seemingly appropriate environment: good lighting conditions, the distance between him and the barn is appropriate, etc. Thus, RVE *prima facie* predicts that Robert knows that a barn is before him, and this is an undesired consequence by Pritchard's lights. Furthermore, Pritchard (2012: 268-71) argues that RVE faces a problem even more pressing: the so-called "creditability dilemma". To see this point, consider the following case:

> **Dzsenifer case:** Dzsenifer is from a small town in Hungary, and she is visiting for the first time the capital, Budapest. She wants to visit their beautiful parliament, so she asks for directions from a stranger because her smartphone has run out of battery. The stranger gives precise and correct directions to Dzsenifer, who finally arrives at the parliament following the directions provided by the stranger.[4]

[4] This is a version of a case formulated by Jennifer Lackey (2007).

In this case, Dzsenifer acquires knowledge based on testimony, but her cognitive success is not primarily due to *her* cognitive abilities. Her cognitive success is primarily explained by the cognitive abilities of the stranger, therefore RVE seems to give the wrong verdict here: RVE would predict ignorance instead of knowledge. Following a weaker reading of the "sufficiently due to" relation, and by appealing to Dzsenifer's cognitive abilities to spot reliable informants, RVE might answer this objection satisfactorily. However, the Dzsenifer case taken together with the fake barn case form Pritchard's "credibility dilemma": (i) if RVE theorists *strengthen* the ability condition to give an ignorance verdict in the fake barn case, then RVE more clearly predicts ignorance in the Dzenifer case and (ii) if RVE theorists *weaken* the ability condition to deal with the Dzsenifer case, then RVE is committed to a verdict of knowledge in the fake barn case.

Since ALVE accepts safety as a necessary condition for knowledge, it gives the correct verdict, an ignorance verdict, in the fake barn case, for Robert's belief is unsafe. Moreover, Pritchard endorses a reading of the "sufficiently due to" relation as partial credibility, so ALVE correctly predicts knowledge in the Dzsenifer case, for her cognitive abilities at spotting good informants partially explain her cognitive success. Hence, ALVE, according to Pritchard, answers satisfactorily his own "creditability dilemma"[5].

3. Kelp's safety dilemma

In this section, firstly, I will present Kelp's (2009; 2016; 2018) safety dilemma which is an objection to safety as a necessary condition for knowledge. ALVE, or any theory that accepts safety as a necessary epistemic condition, is the victim of this dilemma. Secondly, I will argue that Pritchard's (2009b), Grundmann's (2020), and Faria's (2020) responses fail to answer Kelp's safety dilemma satisfactorily. And, finally, I will show that Mortini's (2022) reformulation of safety as "environment-relative safety" answers satisfactorily this objection.

[5] For a solution to this dilemma in terms of the *manifestation* of cognitive ability, see Broncano-Berrocal (2018: 407).

> **Kelp's Frankfurt case[6]:** Chris is staying at his grandparents' house for a few days. One day, he wakes up, goes downstairs, and looks at his grandparent's clock that reads 8:22. Accordingly, Chris forms the true belief that it is 8:22 AM. However, an evil demon who hates Chris wanted him to form the belief that it is 8:22 AM regardless of what the actual time is. Nevertheless, since Chris looked at the clock at the right time, the demon did not intervene because she is lazy. Even if lazy, had Chris looked at the clock one minute later or earlier, the evil demon would have manipulated the clock to induce the belief that it is 8:22 AM. (Kelp 2009: 27-8)

Kelp (2009; 2016; 2018) defends that this is a case of unsafe knowledge. Intuitively, Chris knows that it is 8:22 AM for it has formed this belief by looking at a functioning clock, and the evil demon has not actually intervened. Moreover, Chris' belief is unsafe for had he looked at the clock one minute earlier or later, he would have formed the false belief that it is 8:22 AM by the same belief-forming method: by looking at the clock. This case alone constitutes an objection to safety since Chris' unsafe belief qualifies as knowledge. Besides, taken together with the fake barn case, Kelp's safety dilemma arises (see below).

Kelp defends that in both cases the same kind of environmental luck is present, therefore safety must give the same verdict in both cases: (i) any formulation that is in any way *weakened* to predict knowledge in Kelp's Frankfurt case, then it would also predict knowledge in the fake barn case, and (ii) any formulation that is *strong enough* to predict ignorance in the fake barn case, it will also predict ignorance in Kelp's Frankfurt case. (i) goes against the initial purpose of safety: to avoid that luckily true beliefs, such as those in fake barn cases, qualify as knowledge, and (ii) counterintuitively entails that Chris lacks knowledge.

3.1 Responses

Pritchard (2009: 37-40) dismisses the knowledge verdict in Kelp's Frankfurt case, and he offers an error theory to accommodate the knowledge intuition for that case. According to ALVE, the satisfaction of the ability condition (i.e., a cognitive achievement) is necessary, but not sufficient for knowledge. Following Pritchard's reading of the "sufficiently due to" relation, Chris exhibits a

[6] Kelp labels his example as a "Frankfurt case" because it is an epistemic case analogous to Frankfurt's (1969) moral cases.

cognitive achievement, for his true belief is partially explained by his ability to read clocks. However, it falls short of knowledge since it is unsafe. The fact that a cognitive achievement is present in Kelp's Frankfurt case produces a misleading knowledge intuition. Furthermore, according to Pritchard, since the fake barn case and Kelp's Frankfurt case are perfectly analogous and an ignorance verdict applies to the fake barn case, the same ignorance verdict applies to Kelp's case.

Following Kelp (2018: 108-9), I reject this answer for it is, at best, incomplete: Pritchard does not explain why we have different intuitions in both cases. After all, in both cases, the agent exhibits a cognitive achievement that falls short of knowledge, yet in one case (i.e., the fake barn case), we have the intuition that there is no knowledge, and, in the other case (i.e., Kelp's Frankfurt case), we have a knowledge intuition. If both cases are perfectly analogous and we accept the knowledge intuition in Kelp's case, Pritchard's error theory would imply that we have a knowledge intuition in the fake barn case, which is an undesirable result for him. Insofar as Pritchard does not explain the disanalogy between both cases, his response remains, at best, incomplete.

Faria's (2020) response is dangerously question-begging: he assumes that knowledge being incompatible with a certain kind of luck is a platitude (fair enough) and he also assumes that safety precisely excludes the kind of luck incompatible with knowledge. From this, he concludes that Chris lacks knowledge because his belief is unsafe. More charitably interpreted, even if he does not explicitly say it, Faria (2020) is providing an error theory for the knowledge intuition in Kelp's Frankfurt case: the knowledge intuition is explained away by the fact that Chris' belief has some positive epistemic features (e.g., the belief is justified). Nevertheless, since he accepts that both cases are analogous, he is also affected by Kelp's (2018: 108-9) response to Pritchard: Faria would have to explain why the fake barn case elicits an ignorance intuition whilst Kelp's Frankfurt case produces a knowledge intuition.

Grundmann (2020) argues that his "method-relativized safety", in which the method is externally individuated in a more fine-grained manner, solves Kelp's safety dilemma by giving the correct verdict both in Kelp's Frankfurt case and in the fake barn case. In relevantly close possible worlds where Chris reads the time from a stopped clock, Chris is not using the same method: the externally individuated method includes that the reading is from a properly functioning clock. In contrast, in the fake barn case, Robert's "factive perception" is not a method, for it would trivially entail that his belief is safe; therefore, in relevant close possible worlds, Robert sees a fake barn and believes falsely that a real barn is before him.

Mortini's (2022: 4-7) objection to this strategy is that individuating the belief-forming method in such a fine-grained manner implies that Robert's belief is safe. He argues that Robert's method, externally individuated, ought to include the fact that he is looking at *that* barn which is a real barn. Thus, in close relevant possible worlds, Robert looks at a real barn, since the method is fixed across possible worlds, and forms a true belief. I agree that it is *ad hoc* to say that the properly functioning clock is part of Chris' method and at the same time that the real barn is not part of Robert's method. However, Mortini's objection is dialectally inadequate, for Grundmann (2020: 5177) anticipates this objection and argues, perhaps wrongly, that the "factive perception" of the barn is not part of Robert's externally individuated method. Nonetheless, Grundmann's response suffers from a more straightforward problem:

> **Dario's clocks case:** Dario wakes up and comes downstairs to look at the time since he does not want to be late for an important meeting. The wall is full of clocks that are stopped, yet he happens to look at the only properly functioning clock. He forms the true belief that it is 8:22 AM.[7]

Since Grundmann says that the properly functioning clock is part of Chris' method, he is committed to saying that it is also part of Dario's externally individuated method. For this reason, following Grundmann's "method-relativized safety", the properly functioning clock should be fixed across relevant possible worlds and Dario's belief turns out to be safe. However, Dario's belief is clearly not safe: he could very easily have looked at a stopped clock and formed a false belief. Besides, this is a variation of the fake barn case; thus, once we accept an ignorance verdict for the fake barn case due to unsafety, the same applies to Dario's clocks case. Even if this particular counterexample fails for some reason, Grundman's proposal has the implausible consequence that it is not possible to form unsafe beliefs about the time by looking at a functioning clock: Grundman claims that the functioning clock is part of the relevant method for safety so it must be fixed across possible worlds, and if, by assumption, it is a reliable enough functioning clock, then it will give the correct time in all relevant close possible worlds.

[7] This example is from Mortini (2022: 9), but he uses it to show a completely different point.

3.2 Environment-relative safety

Mortini (2022) reformulates the safety principle as "environment-relative safety" to solve Kelp's safety dilemma. The key idea of this dilemma is that both cases suffer from the same kind of environmental luck. Mortini (2022: 10) disputes this idea by distinguishing "actually unfriendly environments", such as the fake barn case given the actual presence of deceiving fake barns, and "potentially unfriendly environments" which are not actually unfriendly, such as Kelp's Frankfurt case given that no intervention from the evil demon actually takes place. Whereas the former is incompatible with knowledge, the latter is compatible with it. Mortini proposes to accommodate this epistemically relevant difference between both cases with another safety principle, which holds fixed across relevant possible worlds not only the belief-forming method but also the actual environment:

> **Environment-relative safety:** A belief is safe if and only if "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." (Mortini 2022: 10)

This reformulation of safety gives the correct verdict in both cases. On the one hand, regarding the fake barn case, we hold fixed the actual environment which includes fake barns, and, given that Robert looks at fake barns in close possible worlds, he forms false beliefs in those possible worlds. For this reason, his belief is unsafe and does not classify as knowledge. On the other hand, in Kelp's Frankfurt case, we hold fixed the actual environment which includes a perfectly functioning clock and an evil demon that does not actually intervene. Since the clock is, by hypothesis, reliable, it gives the correct time in close possible worlds. And, despite being *potentially* misleading, the evil demon is not misleading in the *actual* environment, so we hold fixed a *potentially-misleading-but-not-actually-misleading* evil demon in close possible worlds. Thus, Chris forms true beliefs in most relevant close possible worlds; for this reason, his belief is safe and a candidate for knowledge.

This proposal, however, is not without any problems: Mortini (2022: 13-4) recognizes that the individuation of the relevant environment is problematic, and it suffers from the generality problem. And, even if Mortini's proposal makes sense of the epistemically relevant difference between "actually unfriendly environments" and "potentially unfriendly environments", it should be further motivated to avoid being *ad hoc*. On a positive note, "environment-relative safety" gives intuitive verdicts in other problematic cases for safety:

> **Dachshund case:** Alvin is hiking in the forest when he sees a dachshund. Accordingly, he forms the true belief that a dachshund is before him. However, Alvin systematically tends to confuse wolves for dachshunds, thus, had he seen a wolf instead of a dachshund, he would have formed the false belief that a dachshund is before him (Goldman 1976: 779).

This case might be problematic for safety since Alvin's belief intuitively amounts to knowledge, yet his belief seems unsafe: in relevant close possible worlds, he sees a wolve and mistakenly believes it to be a dachshund. However, "environment-relative safety" sheds light on this example: (i) If the forest is full of wolves that Alvin would mistake for dachshunds, Alvin does not know after all, and his belief is not environment-relative safe; in contrast, (ii) if no wolves are present in the forest, then his belief is environment-relative safe and a candidate for knowledge. The epistemic risk of actually forming a false belief is higher in (i) compared to (ii), and "environment-relative safety" explains that.

In short, Mortini's (2022) reformulation of safety as "environment-relative safety" is the best response to Kelp's safety dilemma. In addition to giving the correct verdict in both the fake barn case and Kelp's Frankfurt case, it gives an intuitive result of the dachshund case and, potentially, it could deal with other problematic cases for safety. Nonetheless, for environment-relative safety to be considered the best formulation of safety, it needs to be further independently motivated. Still, I will argue for a conditional: if "environment-relative safety" *was* the best formulation of the safety principle, then RVE is even more plausible.

4. A new defense of a modalized robust virtue epistemology

4.1 Environment-relative abilities

Cognitive abilities are dispositions to cognitively succeed (i.e., to attain true beliefs) likely enough under certain triggering conditions. Consequently, abilities can be expressed by "triggering-manifestation conditionals": if one were to exercise cognitive ability A under triggering conditions T, one will likely enough cognitively succeed (Kelp 2018: 20). And, following Broncano-Berrocal (2014: 73), cognitive abilities are globally reliable when they tend likely enough to yield true beliefs regarding a certain domain of propositions in a certain set of circumstances. I will argue that the actual presence of deceiving objects[8] (e.g., fake barns) in the environment is sufficient to rule out that situation from the set of circumstances in which the cognitive ability is reliable. In those situations, even if the subject cognitively succeeds by exercising the relevant globally reliable ability, her success does not manifest the ability (i.e., her success is not sufficiently due to the ability exercised).

I will start by presenting a truism about abilities: abilities are sensitive, to one extent or another, to actual features of the environment. For instance, one might have the ability to play table tennis, but one cannot manifest this ability underwater. Instead of appealing to the actual environment, virtue epistemologists typically appeal to the *kind* of situation appropriate for the manifestation of the ability[9]. That is the set of circumstances in which the exercise of the ability is likely to succeed. Even so, they need to appeal to actual features of the environment: for instance, Robert is presumably in appropriate circumstances, for the *actual* lighting conditions are sufficiently good, the *actual* distance from the barn is adequate to identify it as a barn, etc. In short, all virtue epistemologists are committed to taking into account actual features of the environment, even if they do so only to determine whether or not the subject is in the relevant set of circumstances appropriate for the manifestation of the ability.

[8] For simplicity, I will talk about "deceiving objects", but the same idea applies to actual features of the environment, even if they are not objects, that would cause the subject to cognitively fail (i.e., to form a false belief).

[9] For instance, Sosa (2010) defends that complete competence has a triple-S structure, where the third S refers to the kind of situation. Also, see Greco's (2010) modalized cognitive abilities.

The crucial question arises: which actual features of the environment are relevant for the manifestation of an ability? Most certainly, actual features of the environment that *actually intervene* in the exercise of the ability might be relevant, such as the lighting conditions in the fake barn case. I will motivate the idea that actual features of the environment that *do not actually intervene* may also be relevant for the manifestation of ability.

> **Ishida Case:** Aiko is a student in the history of art at the University of Tokyo. One day, she visits an exposition of Tetsuya Ishida's work, an artist she has studied extensively in class, therefore she is capable of identifying his paintings. She enters the exposition, admires a painting with Ishida's usual protagonist, and forms the true belief that she has an original painting before her. Unbeknownst to her, a trickster has changed all the original paintings for visually indistinguishable ones before Aiko arrived, except for the one that Aiko looked at.[10]

Intuitively, we would say that Aiko does not have the ability to identify original Ishida paintings *in that environment*. For clarification, this is an intuitive response to the case which needs to be explained, but I am not defending that Aiko does not possess the relevant ability. Aiko *does possess* the relevant ability, she *exercises* it, and she *cognitively succeeds*, however, I will argue that *her success does not manifest her ability*. Given the actual presence of fake paintings, if Aiko were to exercise her perceptual-discriminatory ability to discern Fs from non-Fs (in this case, original Ishida paintings from fake ones), she would not likely enough succeed. Although she actually cognitively succeeds and she actually exercises this discriminatory ability which is globally reliable in appropriate circumstances, the actual presence of misleading objects precludes her cognitive success to be the manifestation of her ability. In other words, his cognitive success is not sufficiently due to her ability, but rather, it is due to the coincidence of stumbling upon the only original painting.

> **Cognitive achievements are non-coincidental:** A cognitive success must be more due to ability than coincidence/accident for it to be a cognitive achievement (Sosa 2015; Carter 2016).

This principle has been defended by Sosa (2015) and Carter (2016) in terms of "luck" instead of "coincidence" or "accident". Roughly, I understand that an event occurs by "coincidence" or "accidentally" when it is unexpected or unlikely given the circumstances, and without prejudging its relationship to "luck". Why should we accept this principle? Cognitive achievements are accomplishments for which the subject gets credit: "Manifestation determines credit and discredit"

[10] This case is a version of García's (2018: 37) "Fake Velázquez" case.

(Sosa 2015: 29). If the success is more due to coincidence than ability, the subject deserves less credit for it; as opposed to when the success is more due to ability than coincidence. For instance, Aiko's success is more due to the coincidence of stumbling upon the only original painting than her ability to identify original paintings. Even if she actually succeeds, her success is unlikely or unexpected because her ability is unreliable *in that environment*. Again, given the environment, were she to exercise her perceptual discriminatory abilities to discern originals from fakes, she would not likely enough succeed.

Someone could object that Aiko's success is the manifestation of her ability, for the success is non-deviantly connected by a causal chain to the exercise of a globally reliable ability; and the subject does not even pay attention to the fake paintings. However, even if her ability is globally reliable, the relevant ability is unreliable *in that environment*. That is, in her circumstances, the ability does not tend to yield true beliefs likely enough. Thus, her cognitive success is indeed causally connected to an ability that, despite being globally reliable, is unreliable *in that environment*. A cognitive success due to an unreliable ability is unexpected; therefore, if we accept that achievements are non-coincidental, a cognitive success from an unreliable ability is not a cognitive achievement.

This conclusion should be applied to the fake barn case: given the actual presence of fake barns in the environment, if Robert were to exercise his perceptual-discriminatory ability to discern Fs from non-Fs (in this case, the relevant non-Fs include fake barns), he would not likely enough succeed. The actual presence of fake barns in the environment precludes the situation to be in the set of circumstances appropriate for the manifestation of the ability. This diagnosis permits RVE to give the desired ignorance verdict in the fake barn case: according to RVE, knowledge is a kind of cognitive achievement in which the cognitive success is sufficiently due to cognitive ability, Robert's cognitive success is not sufficiently due to his cognitive ability, thus, Robert does not know that a real barn is before him.

Note that this approach also gives the desired verdict of Kelp's Frankfurt case, a knowledge verdict: given the absence of stopped clocks in the actual environment, Chris' true belief is sufficiently due to his cognitive ability; thus, Chris knows that it is 8:22 AM. Previously, I have granted that the evil demon is part of the actual environment. Thus, someone could object that the evil demon is a factor of the environment that would cause false beliefs just as fake barns or fake paintings do. As noted by Mortini (2022: 8-10), the key difference is that fake barns are actually misleading objects, even if the subject does not pay attention to them. By contrast, the evil demon is potentially-but-not-actually misleading. Hence, while Robert's relevant ability is actually unreliable in his environment, Chris' relevant ability is actually reliable in his environment, even if potentially unreliable.

One could object that it is controversial whether or not fake barns are part of the relevant environment in the fake barn case and, for this reason, merely claiming they are is question-begging. However, if fake barns are not considered part of the relevant features of the environment to determine whether the situation is within the set of circumstances appropriate for the manifestation of the ability, then the objector is committed to saying that Robert's circumstances are equally appropriate as in an epistemically friendly environment. The following example will illustrate this idea:

> **Real barn case:** Roberta has also robbed a bank and she is looking for a place to hide the money from the police. She is driving through the real barn county when she spots a barn and forms the true belief that a real barn is before her. Unbeknownst to her, the region is full of real barns. Had she seen another barn, she would have formed the true belief that a real barn is before her.

Given the absence of fake barns, if Roberta were to exercise her perceptual-discriminatory ability to discern Fs from non-Fs, she would likely enough cognitively succeed. In this case, the relevant non-Fs that are to be distinguished from real barns are objects present in the environment: cows, traffic signs, other cars, etc. If, in the fake barn case, the objector does not take fake barns as relevant features of the environment to determine whether or not the situation is part of the set of circumstances appropriate for the manifestation of cognitive ability, then the objector is committed to treating both cases alike: both Robert and Roberta, if they were to exercise their perceptual-discriminatory ability to discern Fs from non-Fs, they would both likely enough succeed in their respective circumstances. This is a hard bullet to bite, for it does not explain the relevant difference between both environments: while Robert would not cognitively succeed if he continued looking around, Roberta is in a friendly region for the exercise of her ability. And, crucially, Robert deserves less credit than Roberta for his cognitive success: Roberta's true belief is expected since she exercises a cognitive ability that is reliable *in that environment*; by contrast, Robert's success is more due to coincidence than ability.

4.2 Counterexamples

There is a kind of case that easily comes to mind against what I have defended in section 4.1: cases in which an ability is exercised, the subject succeeds, and she could have very easily failed due to environmental conditions, but these environmental conditions do not actually intervene in the exercise of the ability. I will take Pritchard's "Archie case" as a paradigmatic instance of this kind of case.

> **Archie case:** Archie, who is a skillful archer, selects a target among many to throw her arrow. She exercises a perfect technique and, as a result, the arrow hits the target. However, unbeknownst to her, all the other targets were protected by a force field. Had she thrown the arrow at another target, she would have failed in hitting the target (Pritchard 2008: 30).

This is an alleged counterexample to the idea that deceiving or misleading objects in the environment affect the satisfaction of the ability condition. After all, it seems that the mere presence of protected targets does not prevent Archie from manifesting her ability. In other words, Archie's success seems sufficiently due to her archery abilities. Even if this intuition is compelling, I will argue that it does not bear to scrutiny.

Archie's ability to hit targets is not reliable *in that environment*: were she to throw arrows in these circumstances, she would not hit the target likely enough. For this reason, her success is not sufficiently due to her archery ability, but rather, it is due to the coincidence of deciding to throw the arrow at the only unprotected target. The compelling intuition that Archie is manifesting her competence is explained away because Archie does indeed possess the ability to hit targets likely enough *in normal circumstances*. In normal circumstances, there are no force fields or other actual features that reduce her reliability in hitting targets, were she to throw arrows. Still, it seems that Archie's performance achieves the highest normative status since, by hypothesis, she exercised a perfect technique. And the success of her performance is the result of her throwing in a causally non-deviant way. Nevertheless, from a third-person perspective, it is clear that she is not in appropriate circumstances to exercise her archery abilities reliably.

Someone could argue that there are two abilities involved: the ability to choose the target and the ability to hit the target. Whereas Archie's success manifests the ability to hit the target, she fails to manifest her ability to choose appropriate targets. This objection simply moves the issue one step further: Archie does succeed to choose a suitable target, and this success is causally connected to her globally reliable ability to choose appropriate targets. She succeeds both in choosing a suitable target and in hitting the target, but both successes are not achievements, since both abilities are unreliable *in that environment*.

4.3 Entailment thesis

Finally, I will argue that the satisfaction of the ability condition as defended here entails the satisfaction of environment-relative safety. Thus, this concludes the defense of a modalized RVE by negating the independence thesis posited by Pritchard's ALVE.

> **Entailment thesis[11]:** the satisfaction of the ability condition entails the satisfaction of environment-relative safety.

I will take the fake barn case and Kelp's Frankfurt case as examples, but the conclusion can easily generalize. Let's assume that in the fake barn case, the only real barn is in the spatial location s_0, and the fake barns are at s_1, s_2, s_3,..., s_n; where n is the number of fake barns in the environment. Robert's belief is not environment-relative safe for in most nearby possible worlds, holding fixed the environment, he forms the false belief that a real barn is before him. In the actual world, w_0, Robert looks at s_0 and forms a true belief. However, in possible world $w_{1,}$ he looks at s_1 and forms a false belief; in possible world w_2, he looks at s_2 and forms a false belief; etc.

By contraposition of the entailment thesis, if a belief is not environment-relative safe, then it does not satisfy the ability condition. If there are nearby possible worlds like w_1, w_2, etc. in which, holding fixed the environment, Robert cognitively fails, this is due to the actual presence of deceiving objects at spatial locations s_1, s_2, s_3,..., s_n. And the presence of deceiving objects entails that Robert's actual cognitive success is not due to the manifestation of cognitive abilities. Necessarily, the contraposition holds because: if there are nearby possible worlds holding fixed the environment in which the subject cognitively fails due to encountering deceiving objects, then that environment does not pertain to the set of circumstances appropriate for the manifestation of the ability. In other words, the presence of deceiving objects will cause environment-relative unsafety

[11] This terminology comes from Carter (2016).

and this entails that the ability condition is not met because the presence of deceiving objects precludes the manifestation of the ability.

In Kelp's Frankfurt case, assuming that Chris is in good shape and has the skill to read clocks, the ability condition is satisfied in part due to the absence of stopped clocks in the environment. Since there are no deceiving objects in the environment, there are no nearby possible worlds holding fixed the environment in which Chris looks at a stopped clock and forms a false belief via the same belief-forming method. In general, if the subject is in good shape, possesses the relevant skill, and there are no deceiving objects in the environment, then the ability condition will be met. And, given the absence of deceiving objects, there will be no nearby possible worlds in which the subject cognitively fails, assuming the environment is fixed.

García (2018), Greco (e.g., 2009), and Littlejohn (2014) have also argued that the ability condition entails safety. However, there is a crucial difference between these proposals and mine: they define abilities in terms of possible worlds and, for this reason, safety is trivially entailed. By contrast, I have argued that the actual presence of deceiving objects in the environment precludes the manifestation of ability without appealing to possible worlds, so safety is not trivially implied. Besides, the ability condition as understood here entails environment-relative safety in particular, which is a reformulation of safety that avoids Kelp's safety dilemma.

5. Conclusion

In this essay, I have argued that Mortini's (2022) reformulation of the safety principle as "environment-relative safety", which is the best response to Kelp's (2009; 2016; 2018) safety dilemma, makes RVE even more plausible. First, I have presented Pritchard's ALVE, which is a modest or impure virtue epistemology based on the independence thesis. Second, I have presented Kelp's (2009; 2016; 2018) safety dilemma which is an objection to the necessity of safety, I have shown that Mortini's (2022) reformulation of safety as "environment-relative safety" satisfactorily answers this dilemma, and I have argued that this is the best answer thus far in the literature. In section 4, I have defended that the satisfaction of the ability condition ought to be relativized to actual features of the environments. More specifically, in cases like the fake barn case, I have argued that the mere presence of deceiving objects in the environment is sufficient to preclude the success from being the manifestation of cognitive ability. And, finally, I have defended a modalized version of RVE by negating the independence thesis and endorsing the entailment thesis: the satisfaction of the ability condition, where the manifestation of ability is relativized to actual features of the environment, entails the satisfaction of environment-relative safety.

Some questions remain open for further research. On the one hand, environment-relative safety suffers from a version of the generality problem based on environments and, arguably, the same problem applies to the relativization of the manifestation of cognitive abilities to actual features of the environment. However, since all virtue epistemologies need to appeal to actual features of the environment, even if it is just to determine whether or not the subject is in the set of circumstances in which the ability is globally reliable, this is not a specific problem for the version of virtue epistemology defended here. On the other hand, even if environment-relative safety is the best response to Kelp's safety dilemma and it makes sense of epistemically relevant differences in environments, it should be further independently motivated to be considered the best formulation of the safety principle.

6. References

Broncano-Berrocal, Fernando. 2014. «Is Safety In Danger?» *Philosophia* 42 (1): 63-81.

———. 2018. «Purifying Impure Virtue Epistemology». *Philosophical Studies* 175 (2): 385-410. https://doi.org/10.1007/s11098-017-0873-x.

Carter, J. Adam. 2013. «A Problem for Pritchard's Anti-Luck Virtue Epistemology». *Erkenntnis* 78 (2): 253-75. https://doi.org/10.1007/s10670-011-9315-x.

———. 2016. «Robust Virtue Epistemology As Anti-Luck Epistemology: A New Solution: Robust Virtue Epistemology As Anti-Luck Epistemology». *Pacific Philosophical Quarterly* 97 (1): 140-55.

Colaço, David, Wesley Buckwalter, Stephen Stich, y Edouard Machery. 2014. «Epistemic Intuitions in Fake-BarnThought Experiments». *Episteme* 11 (2): 199-212.

Craig, Edward. 2004. Knowledge and the State of Nature: An Essay in Conceptual Synthesis. Book, Whole. Oxford: *Oxford University Press*. https://doi.org/10.1093/0198238797.001.0001.

Faria, Domingos. 2020. «Is Epistemic Safety Threatened by Frankfurt Cases? A Reply to Kelp». *Diametros* 66: 1-6. https://doi.org/10.33392/diam.1448.

Frankfurt, Harry G. 1969. «Alternate Possibilities and Moral Responsibility». *The Journal of Philosophy* 66 (23): 829-39.

García Rodriguez, Ángel. 2018. «Fake Barns and our Epistemological Theorizing». Crítica; *Revista Hispanoamericana de Filosofía* 50 (148): 29-54.

Gettier, Edmund L. 1963. «Is Justified True Belief Knowledge?» *Analysis* 23 (6): 121-23. https://doi.org/10.1093/analys/23.6.121.

Goldman, Alvin I. 1976. «Discrimination and Perceptual Knowledge». *The Journal of Philosophy* 73 (20): 771-91. https://doi.org/10.2307/2025679.

Greco, John. 2003. «Knowledge as Credit for True Belief». In Intellectual Virtue: Perspectives from Ethics and Epistemology, Michael DePaul and Linda Zagzebski. Oxford: *Oxford University Press*.

———. 2009. «Knowledge and Success from Ability». *Philosophical Studies* 142 (1): 17-26. https://doi.org/10.1007/s11098-008-9307-0.

———. 2010. Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity. *Cambridge University Press*.

Grundmann, Thomas. 2020. «Saving Safety from Counterexamples». *Synthese* 197 (12): 5161-85. https://doi.org/10.1007/s11229-018-1677-z.

Kelp, Christoph. 2009. «Knowledge and Safety». *Journal of Philosophical Research* 34 (July): 21-31. https://doi.org/10.5840/jpr_2009_1.

———. 2016. «Epistemic Frankfurt Cases Revisited». *American Philosophical Quarterly* 53 (1): 27-37.

———. 2019. Good Thinking: A Knowledge First Virtue Epistemology. Vol. 114; Book, Whole. New York, NY: *Routledge*. https://doi.org/10.4324/9780429455063.

Lackey, Jennifer. 2007. «Why We Don't Deserve Credit for Everything We Know». *Synthese* 158 (3): 345-61. https://doi.org/10.1007/s11229-006-9044-x.

Lewis, David K. 2001. On the Plurality of Worlds. Malden, Mass: *Blackwell Publishers*.

Littlejohn, Clayton. 2014. «Fake Barns and False Dilemmas». *Episteme* 11 (4): 369-89. https://doi.org/10.1017/epi.2014.24.

Mortini, Dario. 2022. «A New Solution to the Safety Dilemma». *Synthese* 200 (2). https://doi.org/10.1007/s11229-022-03626-w.

Pritchard, Duncan. 2007. «Anti-Luck Epistemology». *Synthese* 158 (3): 277-97. https://doi.org/10.1007/s11229-006-9039-7.

———. 2008. «I—Duncan Pritchard: Radical Scepticism, Epistemic Luck, and Epistemic Value». *Supplementary Volume - Aristotelian Society* 82 (1): 19-41. https://doi.org/10.1111/j.1467-8349.2008.00160.x.

———. 2009a. Knowledge. Book, Whole. Basingstoke: *Palgrave Macmillan*.

———. 2009b. «Safety-Based Epistemology: Whither Now?» *Journal of Philosophical Research* 34 (July): 33-45. https://doi.org/10.5840/jpr_2009_2.

———. 2013. «Can Knowledge be Lucky?» In *Contemporary Debates in Epistemology*, Steup, M., Turri, J., and Sosa, E., 152-64. John Wiley & Sons.

———. 2012. «Anti-luck Virtue Epistemology». *The Journal of Philosophy* 109 (3): 247-79. https://doi.org/10.5840/jphil201210939.

Sosa, Ernest. 1999. «How to Defeat Opposition to Moore». *Noûs* 33: 141-53.

———. 2015. Judgment and Agency. First. Book, Whole. Oxford: *Oxford University Press*. https://doi.org/10.1093/acprof:oso/9780198719694.001.0001.

Williamson, Timothy. 2000. Knowledge and Its Limits. Oxford; *Oxford University Press*.

Zagzebski, Linda Trinkaus. 1996. Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge. Cambridge: *Cambridge University Press*.