Synthese (2023) 202:203
https://doi.org/10.1007/s11229-023-04412-y

ORIGINAL RESEARCH

Etiological proper function and the safety condition

Dario Mortini[1]

Received: 17 April 2023 / Accepted: 26 October 2023

Abstract
In this paper, I develop and motivate a novel formulation of the safety condition in terms of etiological proper function. After testing this condition against the most pressing objections to safety-theoretic accounts of knowledge in the literature, my conclusion will be the following: once safety is suitably understood in terms of etiological proper function, it stands a better chance as the right anti-Gettier condition on knowledge.

Keywords Safety · Proper function · Etiological theory of function · Epistemic luck

Introduction

Since the publication of Gettier (1963), epistemologists have widely agreed that even if belief, truth and (perhaps) justification may be individually necessary, they are nonetheless jointly insufficient conditions for knowledge. While this point should be uncontroversial, what still remains controversial is how to answer the crucial question raised by Gettier's paper: in addition to belief, truth and justification, which fourth 'anti-Gettier' condition can be necessary and, together with the other conditions, also sufficient for knowledge? Many prominent anti-Gettier conditions have been offered, and yet, to this day, this crucial question remains unanswered.

In this paper, my aim is to revive an old answer to this question. I propose a novel formulation of safety and offer it as a satisfactory anti-Gettier condition. I do so by developing a hitherto unexplored hybrid: I shall elucidate safety by appeal to etiological proper function. In keeping with safety-theoretic approaches, the account of knowledge proposed in this paper is modal in that it views knowledge as a matter of modal robustness across relevant possible worlds; however, modal robustness is in turn cashed out in terms of proper function of belief-forming methods in appropriate cognitive environments. This novel hybrid comes with important advantages. First, in virtue of being indexed to both belief-forming methods and environments, it advances

✉ Dario Mortini
damo@sdu.dk

[1] University of Southern Denmark, Odense, Denmark

Published online: 12 December 2023

our understanding of the safety condition. Second, not only can it withstand the most pressing objections moved against standard versions of safety, but it also passes the necessity and sufficiency tests for knowledge.

My plan is as follows. Section 1 gives a brief survey of the present state of play: I outline extant versions of safety and I situate them in the context of other prominent modal and non-modal anti-Gettier conditions. Section 2 diagnoses the reasons why safety fails as anti-Gettier condition, and it also identifies the key desiderata that a suitably understood safety condition ought to meet. In Sect. 3 I put forth and motivate my novel formulation of safety in terms of etiological proper function, and I show that it meets these key desiderata. Next, in Sect. 4, I check my proposed formulation against the most pressing counterexamples to both the necessity and the sufficiency of safety for knowledge, and I detail how it withstands each of them. Section 5 brings out the distinctiveness of my proposed formulation by comparing it with other prominent modal and non-modal anti-Gettier conditions. I conclude with a methodological afterthought, explaining to what extent the version of safety offered in this paper successfully yields a necessary and sufficient condition for knowledge.

1 Extant versions of safety

Because the literature on safety is extensive, at first pass we can introduce it by appeal to the following oversimplified but intuitive gloss: a belief is safe if and only if, given the way it was formed, it couldn't easily have been false.[1] Crucially, the 'couldn't' is unpacked *modally* – that is, in terms of relevantly similar close possible worlds. Early advocates of safety include (Williamson, 1994; Sainsbury, 1997; Sosa, 1999; Pritchard, 2005). Each of these authors has defended a different version of safety with different motivations, but two general points are worth emphasizing. First, safety has a valuable anti-sceptical import, as it permits to know the denial of sceptical hypotheses (Pritchard, 2008). Secondly, safety also explains absence of knowledge in several key cases routinely discussed in epistemology - most importantly Gettier cases and lottery cases.[2] But despite these *prima facie* advantages, the safety condition faces substantial objections, as well as pressing counterexamples. For this reason, many authors prefer to pair safety with a different condition, while a minority even rejects it altogether. In what follows, I provide a quick overview of the current state of play of safety, focusing

[1] See Rabinowitz (2011) for an overview of the safety condition *for* knowledge.

[2] Williamson (1994) firstly appeals to a safety-style argument in the context of the problem of vagueness, and then returns to the safety condition for knowledge in later work (Williamson, 2000, 2009). Sainsbury (1997) follows up on Williamson's early work and expands on the notion of easy possibility. Sosa (1999) focuses on a counterfactual formulation of safety which is meant to improve on both Nozick's sensitivity condition (Nozick, 1981) and Goldman relevant alternatives theory of knowledge (Goldman, 1976). Sosa's counterfactual formulation is not without problems: see Comesaña (2007) and McGlynn (2012) for criticism; see Holliday (2015) for a formal discussion of counterfactual theories of knowledge. Pritchard (2005, 2007) develops a safety condition in order to deal with cases of *veritic* luck - the type of luck which makes a belief only accidentally true and prevents it from constituting knowledge. Pritchard's safety condition also purports to explain absence of knowledge in Gettier and lottery cases. As should be clear, each of these authors develops the safety condition for different reasons; some of these authors have also revised their view during the years (e.g., Sosa and Pritchard). For present purposes, it's important to note the diversity of the main formulations of safety and the key reasons that make it attractive.

especially on the role of safety in the most prominent recent accounts of knowledge. I will begin with the accounts that are most sympathetic to safety and end with those that reject it altogether. The point of this quick and inevitably one-sided overview is to highlight the fact that a combination of safety and proper function so far has no takers in the literature.

At the most sympathetic end of the spectrum, we find what is sometimes called 'impure virtue epistemology' (Pritchard, 2012, 2016; Kelp, 2013). Impure virtue epistemologists accept some version of the safety condition, but then they supplement it with a *separate* ability condition. While safety does contribute to knowledge, it does so only *in conjunction* with a different ability condition.[3] As a result, according to impure virtue epistemologists, safety can't stand on its own and needs to be paired with a distinct ability condition.

On to pure or 'robust' virtue epistemologists (Sosa, 2007, 2015, 2021; Greco, 2010, 2020a; Carter, 2016; Turri, 2016). According to robust virtue epistemologists, a restricted safety condition can be derived from a more fine-grained and central ability condition. Sosa no longer accepts his original safety condition (Sosa, 2007), and now opts for an ability condition elucidated in terms of the key concept of aptness (Sosa, 2015). Sosa's ability condition may be interpreted as entailing a weaker *restricted* safety condition.[4] Similarly, Greco (2020a), Greco (2010) develops an environment-relative conception of abilities that entail a weaker, version of safety. Carter (2016) also proposes a more graded (as opposed to rigid) notion of abilities to capture the central intuitions standardly accommodated by the safety condition.[5] Crucially, according to pure robust virtue epistemologists, knowledge is primarily an achievement which may *at best* entail a weaker and restricted safety condition: accordingly, pure robust virtue epistemologists elucidate the safety condition in terms of an ability condition.

A significantly different and thus stand-alone version of virtue epistemology is Beddor and Pavese modal virtue epistemology (Beddor and Pavese, 2018). While they are sympathetic to safety, just like robust virtue epistemologists Beddor and Pavese also hold that some version of safety is entailed by their ability condition restricted to normal worlds (Beddor and Pavese, 2018, p. 68). As a result, also this version of virtue epistemology derives a safety condition from an ability condition.

Since I mentioned normal worlds, it's also worth mentioning recent versions of normalcy epistemologies (Goodman and Salow, 2018; Littlejohn and Dutant, 2020;

[3] Impure virtue epistemologists pair safety with an ability condition for reasons having to do with counterexamples to the sufficiency of safety for knowledge (e.g., Temp-style cases; see Pritchard, 2012). After outlining my novel safety condition, I return to these cases later on in Sect. 4.

[4] According to Greco (2020a), Sosa defends a SSS-Safety condition modelled after his account of abilities and restricted to Seat, Shape and Situation. Hirvelä and Paterson (2021) and Carter (2022, Chap. 2) also adopt Greco's interpretation.

[5] Robust virtue epistemologists notoriously struggle with cases of environmental luck exemplified by Goldman's (Ginet's) famous fake-barn scenarios. While the agents in these scenarios do seem to successfully exercise their cognitive abilities, they nevertheless intuitively lack knowledge because of the misleading nature of the environment (Goldman, 1976; Pritchard, 2012). Robust virtue epistemologists have either bitten the bullet and conceded knowledge in fake-barn cases (Sosa, 2015) or appealed to a more fine-grained conception of abilities to capture the no knowledge intuition (Carter, 2016; Littlejohn, 2014). I also note that fake-barn cases are especially controversial: the no knowledge intuition has been challenged on both philosophical (Lycan, 2006; Schellenberg, 2018)) and experimental grounds (Colaço et al., 2014). I return to these issues in Sect. 4, and discuss what my proposed safety condition predicts about these cases.

Horvath and Nado, 2021).[6] These views retain the spirit of modal conditions on knowledge, but the explanatory focus is on the key notion of normality (which is in turn unpacked differently depending on the version of normalcy epistemology under consideration). Accordingly, normalcy epistemologies do accept some version of the safety condition but construe safety in terms of normality.

We have now reached the opposite end of the spectrum—views that are unsympathetic to safety and reject it altogether. According to more or less recent explanationist approaches, it is the (non-modal) notion of explanation that holds the key to deliver a successful analysis of knowledge (Jenkins, 2006; Neta, 2002; Bogardus and Perrin, 2022). Explanationists are well aware that since their central notion of explanation is hyperintensional (Nolan, 2014, p. 157), it follows that knowledge "cannot be captured in purely modal terms" (Bogardus and Perrin, 2022, p. 7). Advocates of explanationism have adduced a variety of motivations for their approach,[7] But given their discontent with modal conditions on knowledge in general, they also reject *a fortiori* any relevance of safety for knowledge.

Before ending this section, I want to briefly focus on what we may call 'circular' accounts of safety. These accounts are prominently defended by champions of the knowledge-first approach to epistemology (Williamson, 2000; Lasonen-Aarnio, 2010). While these authors accept the necessity (Williamson, 2000) or the sufficiency of safety for knowledge (Lasonen-Aarnio, 2010), they elucidate safety in terms of knowledge rather than the other way around (Williamson, 2000; Kelp, 2017). As I explained above, I am primarily interested in a version of the safety condition that can serve as successful anti-Gettier condition and make predictions on whether a true belief constitutes knowledge: since these authors are not interested in developing safety as an anti-Gettier condition, their version of safety does not play the theoretical role that I am focusing on in this paper.

Let's take stock. As made clear by this admittedly quick overview, a combination of safety and etiological proper function is still lacking. Moreover, given some pressing problems faced the safety condition, many prominent contemporary accounts of knowledge either pair it with a distinct condition or reject it altogether. In the next section, I take a closer look at these problems: I provide a new diagnosis of the shortcomings of extant versions of safety and outline the key desiderata that a better safety condition ought to meet. With an eye on such desiderata, I will then be able to develop such better safety condition in the remainder of the paper.

[6] My focus is on knowledge, but see Smith's normic theory of (propositional) justification for a detailed discussion of a modal conception of normality (Smith, 2016). See also Peet and Pitcovski (2018) for an account of knowledge as safe true belief in which the notion of normal explanation plays a key role.

[7] Some explanationists buld on debunking arguments in metaethics (Korman and Locke, 2020) others emphasise the shortcomings of modal accounts of defeat (e.g., Bogardus and Perrin, 2022), and others emphasise the difficulties faced by modal accounts of knowledge in dealing with necessary truths (Faraci, 2019).

2 Diagnosis and desiderata

Despite the advantages and the motivations that once made safety attractive, we can identify at least two main difficulties currently faced by safety-theoretic accounts of knowledge.

Firstly, there's a serious worry about *extensional adequacy*: the safety condition suffers from very strong counterexamples to its necessity and sufficiency for knowledge. To cast doubt on the necessity of safety for knowledge, epistemologists have put forth several cases of unsafe knowledge. In these cases, agents intuitively seem to know in the actual world despite forming a false belief in relevantly similar close possible worlds (Neta and Rohrbaugh, 2004; Comesaña, 2005; Baumann, 2008; Kelp, 2009, 2016; Bogardus, 2014).[8] To put pressure on the sufficiency of safety for knowledge, epistemologists have also offered cases where agents form a trivially safe true belief and yet they lack knowledge: for instance, new versions of meta-incoherence cases à la Mr. Truetemp (Bonjour, 1980; Lehrer, 1990; Pritchard, 2012) and the long-standing problem with trivially safe necessary truths (Roland and Cogburn, 2011). Moreover, other authors have questioned the success of safety as anti-Gettier condition. *De re* versions of Goldman's fake-barn cases (Pryor, 2004), the conjunction of fake barn cases with epistemic Frankfurt cases (Kelp, 2016) and more refined Gettier-style cases (Miracchi, 2015) strongly suggest that safety fails as a satisfactory anti-Gettier condition. All these counterexamples led epistemologists to either implement safety with a separate condition or to abandon it entirely.[9]

So much for extensional adequacy. Importantly, problems with extensional adequacy are decisive only insofar as the project of providing a reductive analysis of knowledge is taken seriously. But on the top of this first worry, there's a second (and perhaps deeper) worry: the safety condition is too vague and uninformative. In the previous section, safety was glossed as follows: a belief is safe if and only if, given the way it was formed, it couldn't easily have been false (in most or all relevantly close worlds where the subject continues to form beliefs in the same/sufficiently similar way as in the actual world, the subject's belief continues to be true).[10] While this gloss captures both the core idea that modal robustness is essential to knowledge and that safety is best understood as relative to the way the belief is formed (bases, or belief-forming method), this basic formulation also leaves open far too many questions. Here's a few: how do we exactly individuate the relevant possible worlds? What are the appropriate belief-forming methods? Is safety relativized to belief forming methods only, or, just like some conceptions of abilities, should it be relative also to environments?

[8] See also Sosa (2010, p. 471) for the contention that the nearness of a dream scenario puts pressure on the necessity of safety for knowledge.

[9] This is not say that safety theorists have not tried to overcome these difficulties. Along the way, I will have a little more to say about these proposed solutions, and clarify how they differ from mine.

[10] This gloss captures the spirit of the most prominent formulations of safety. 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". See Ranalli (2014, p. 230, fn.12) and Blome-Tillmann (2020, Appendix) for a nice taxonomy of formulations of safety.

Some advocates of safety such as Williamson (2009) and Pritchard (2012) hold that safe belief-forming methods yield true beliefs not in one single proposition p but in a *range* of propositions sufficiently similar to p.[11] If so, how can we precisely demarcate the relevant range of propositions? The success of safety crucially depends on these questions, and yet safety theorists have either answered too vaguely or not answered at all.[12]

Taken together, these worries highlight the two main shortcomings of the safety condition. Firstly, it faces too many counterexamples to its necessity and sufficiency for knowledge: it's not extensionally adequate. Secondly, it is also too vague and underspecified: overall, it's not informative enough to serve as a satisfactory necessary and sufficient anti-Gettier condition. This diagnosis suggests two main desiderata for a better safety condition, which ought to be:

- *Extensionally adequate*, and withstand the most pressing counterexamples to its necessity/sufficiency for knowledge and its tenability as an anti-Gettier condition.
- *More informative*, and provide a set of exhaustive criteria to individuate *(i)* the relevant possible worlds, *(ii)* the appropriate belief-forming methods, *(iii)* the appropriate environments and *(iv)* range of propositions.

It's worth pausing to set the record straight on the main upshot of the present discussion. The safety condition remains too underdeveloped: it just offers a useful *template* with different parameters that must be filled in a more informative and precise manner.[13] Such parameters include (but are not limited to): the proposition(s) believed, the relevantly similar close worlds, the appropriate belief-forming methods and perhaps also the appropriate environments where said belief-forming methods can be employed. More crudely put: to stand a chance as an extensionally adequate, informative and successful anti-Gettier condition which is necessary and sufficient for knowledge, safety must be made more precise. Importantly, a hitherto unexplored way to make the safety condition more precise is to appeal to proper functionalism: such a framework provides the necessary explanatory resources to fill all the parameters that the safety template leaves blank. In the next section, I rely on the proper functionalist framework and I proceed to fill such parameters. The resulting safety condition will meet the two desiderata outlined above, and thus stand a much better chance as extensionally adequate, informative and successful anti-Gettier condition.

[11] For discussion on how to identify the relevant range of propositions, see Hirvelä (2019) and Bernecker (2020).

[12] See Pritchard (2008, pp. 444–446); Pritchard (2013, p. 158) and Rabinowitz (2011, Sect. 3) for a brief discussion of individuation of belief-forming methods, and the exchange between Goldman (2009) and Williamson (2009). See also Alfano (2009), Grundmann (2018) and Hirvelä (2019). With the exception of Hirvelä's virtue-theoretic criterion, these discussions have focused on fine-grained/coarse-grained or internal/external individuation of belief-forming methods; what is absent is a precise criterion that specifies belief-forming methods by appeal to more than some relevant features of the actual world. I provide such a criterion in the later section.

[13] For sake of vividness, take a moment to compare safety with clearly more complex virtue-theoretic ability conditions (e.g., Sosa's triple A performance normativity framework and triple S account of abilities, or Greco's environment-relative/contextualist account). There's a clear asymmetry here: the vagueness of the safety condition contrasts unfavorably with these more developed and informative virtue-theoretic ability conditions.

3 Etiological proper function safety outlined

Given that safety stands in need of further clarification, it is natural to look for the right explanatory framework that can achieve such a goal. My suggestion is to look at proper functionalism, and I do so for two main reasons. Firstly, proper functionalism has a proven track record of success in elucidating key epistemic concepts like *knowledge* (Millikan, 1984; Plantinga, 1993), *entitlement* and *justification* (Bergmann, 2006; Simion, 2019). Secondly, as I explain in this section, the safety condition is compatible with proper functionalism and perhaps even naturally lends itself to a proper functionalist reading.

Let's begin with the central notion of proper function. This notion is meant to explicate not only the essentially teleological sense in which objects like organisms and artifacts have a certain *function* (they are *for* something; they have a *purpose* or they display a *design*), but also how said objects are *supposed* to function properly in suitably specified circumstances (appropriate environments). For example, a human heart is supposed to pump blood by beating at approximately 70 beats per minute in a healthy enough human body. When it does so, it's functioning *properly* and it fulfils its *purpose*. But how does an item acquire a purpose? When does it function properly, and how is an environment appropriate? According to the influential etiological theory of functions, an item's purpose is equated with the selected effect that explains why the item was replicated through biological reproduction or otherwise,[14] When the item does what it was selected for, it's functioning *normally* and *properly*. And to function normally and properly, the item needs to be located in a suitable normal environment, similar enough to the one where the item originally came to have its function. This normal environment is *appropriate* for the item's function. Setting several complications aside, in order to develop a new version of safety we can focus on these core concepts: purpose as etiological function, proper (normal) function and appropriate (normal) environments.[15]

By appeal to these core concepts, we can fill more precisely and systematically the parameters left empty by the too vague safety template. Recall that a properly understood safety condition ought to be informative and specify all the following: a criterion of individuation for belief forming methods, the environments where the belief-forming methods are employed, the relevant swath of possible worlds and also the relevant range of propositions. To fill all these parameters, let's plug in the notion of proper function and the etiological theory of functions:

- *Belief-forming methods (bases)*. Properly (normally) functioning belief-forming methods individuated with reference to their etiological function.

[14] While I favour the naturalistic etiological theory of functions defended by Wright (1973), Millikan (1984), Buller (1998), Graham (2012), and Simion (2019), the proposed safety condition can be developed by appeal to different theories of functions, such as Plantinga's. See Plantinga (1993) for a *locus classicus* and Graham (2011; 2012, p. 476, fn. 5) for a detailed comparison of these two theories of functions. See also Graham (forthcoming) for an updated discussion of the role of naturalistic theories of functions in epistemology.

[15] A note on terminology. I will use 'proper function' as 'normal function' and 'appropriate environment' as 'normal environment'. After all, for the etiological theory of functions proper function just is normal function. (Graham 2011, p. 75) also identifies the two.

- *Environments*. Appropriate cognitive environments for the proper function of the belief-forming methods.
- *Modal robustness*. All the relevant possible worlds where the proper functionalist conditions for both belief-forming methods and environments are met.
- *Range of propositions*. All the propositions specified in accordance with the etiological function of the belief-forming methods.

Some comments on each of these parameters are in order. As far as concerns belief-forming methods, three points are worth noting. Firstly, the method under consideration ought to be functioning properly. For example, safe perceptual beliefs will result from properly functioning perceptual capacities, and in the case of inferential beliefs formed on the basis of an instrument, said instrument also ought to be functioning properly (it should not display any type of malfunction). Secondly, the belief-forming method must be designed to aim at truth in line with the etiological theory of functions.[16] To wit, a belief-forming method like reading from a horoscope won't count as such: its etiological function has nothing to do with truth. The same applies also to only accidentally reliable belief-forming methods: for example, forming beliefs on the basis of a fully accurate brain lesion will not count as a properly functioning belief forming method because, despite the accuracy, such brain lesion fundamentally lacks any function (etiological or otherwise). Thirdly, the etiological function of belief-forming methods plausibly includes a wider range of propositions instead of one single proposition. This applies to both non-inferential belief-forming methods constituted by biological kinds (perceptual systems) and inferential belief-forming methods based on artifacts (instruments). For example, a clock's etiological function is to reliably yield true beliefs in a range of propositions about the current time; a barometer does just the same with a range of propositions about the air pressure. *Mutatis mutandis*, similar considerations also hold for biological kinds: eyesight's etiological function is to reliably yield true beliefs in a set of propositions about the features of one's surroundings. Once again, it is the etiological function of the belief-forming method under consideration that determines the relevant range of propositions.

[16] On the issue of design, I want to tread carefully. For a satisfactory individuation of properly functioning belief forming methods, all that matters is only that they are designed to aim at truth. What is the exact source of the design is an interesting question, but it doesn't affect the proper functionalist criterion of individuation of belief forming methods offered here. That said, there are various proposals in the literature, and they all seem prima facie compatible with a proper functionalist criterion of individuation of belief-forming methods. (Plantinga 1993, Chap. 11) favours a theistic conception of design and opts for conscious intentional design. Following the influential work in philosophy of biology of Wright (1973), Millikan (1984) and Buller (1998), Graham (2011) offers a different naturalistic evolutionary conception of design, and adopts an etiological theory of functions according to which functions are equated with biological benefit historically selected across a large enough period of time. Simion (2016) also offers an etiological theory of functions, but, unlike Graham, she relaxes the historical condition and demands only a more broadly epistemic (rather than specifically biological) type of benefit. As I said, my preferences lie with Graham and Simion's accounts. However, I hasten to flag that questions about the source of the design are orthogonal to the more pressing question for the safety condition, which is how to best individuate belief-forming methods. All that is needed for the individuation of belief-forming method is some design and the resulting proper function, regardless of the ultimate source of each.

On to environments. Importantly, proper function of belief-forming methods is not sufficient to elucidate the safety condition. A belief-forming method may be functioning normally, properly, and even perfectly, but if the environment is not appropriate the belief-forming method in question will not yield neither true nor safe beliefs. Beliefs formed on the basis of properly functioning perceptual capacities employed in a broadly misleading environment will not be accurate, nor safe. Accordingly, the safety condition must be indexed not only to properly functioning belief-forming methods, but also to the appropriate cognitive environment for the proper function of such belief-forming methods. Given the etiological theory of functions, these will be suitably normal environments for the normal functioning of the belief-forming method.

Finally, modal robustness. In keeping with safety-theoretic accounts of knowledge, the safety condition developed here requires that the belief under consideration is true in a set of relevant possible worlds. However, modal robustness is cashed out in terms of (etiological) proper function: the relevant possible worlds are those where both the (properly/normally functioning) belief forming-method and the (appropriate/normal) cognitive environment of the actual world are kept fixed. Small changes in the belief-forming methods or environments will make the possible worlds under consideration irrelevant for the proper functionalist safety condition developed here. Putting this all together, we get:

> **Etiological Proper Function Safety**. S's belief that *p* is safe if and only if in *most* of the close similar possible worlds in which S continues to form her belief in the target propositions *p* via the same properly functioning belief-forming method M employed in the actual world (sub-condition **M**) and in the same appropriate cognitive environment E occupied in the actual world (sub-condition **E**), the belief continues to be true.

Unlike other formulations, this version of the safety condition incorporates the central insights of proper functionalism and the etiological theory of functions. As such, it is indexed to *both* properly (normally) functioning belief-forming methods and appropriate (normal) cognitive environments. Notice finally how this is a *conjunctive* formulation: just like in order for a conjunction to be true *both* conjuncts need to be true, in the same way both sub-conditions **M** and **E** have to obtain for this safety condition to be met. This reflects the observation made above about the insufficiency of proper function of belief-forming methods: what must be added and specified is the proper function of belief-forming methods in cognitive environments that are appropriate (normal) for the method under consideration.

While this formulation makes some progress, it is still not good enough: absent a satisfactory explanation of what it means for a belief-forming method to be functioning properly and what it is for an environment to be appropriate and normal, this formulation of safety might fall prey to the same problems discussed in Sect. 2 (in particular, the lack of informativeness). Appealing to the relevant theory of functions will address this worry. For example, on the *etiological* theory of functions, a belief-forming method meets the proper function condition insofar as it (reliably) yields true belief, and (reliably) yielding true beliefs *explains* why the belief-forming method

was selected and reproduced. Appropriate cognitive environments are those (normal) environments in which the belief-forming method acquired the function of (reliably) yielding true beliefs. This is but one way of giving criteria for the proper function of belief-forming methods and appropriate environments: different theories of functions will make different predictions.

Having clarified this, I want to conclude this section by offering some remarks in support of a proper functionalist reading of the safety condition. I shall do so by highlighting a few relevant conceptual connections between safety and proper function within and outwith epistemology. These connections clarify the hitherto unappreciated compatibility between these two approaches, and also bring out the distinctiveness of the novel formulation of safety I am offering.

Firstly, starting with safety, it's important to note that safety theorists have excellent reasons to appeal to proper function in order to make their modal condition on knowledge more precise. Recall the key idea behind safety: given the way they are formed, safe beliefs are meant to be modally robust, so that they continue to be true across a relevant swath of possible worlds. Crucially, once we understand 'the way they are formed' by reference to properly functioning belief-forming methods employed in appropriate cognitive environments, such beliefs are indeed modally robust: proper function offers the type of (modal) protection from error which is essential to safety. Yet, despite the clear sense in which safety and proper function are compatible and even complementary, safety theorists have never developed their proposed condition by specifying belief-forming methods and environments in terms of proper function. As a result, the safety condition incurred the major difficulties pointed out above.

Secondly, as far as proper functionalism is concerned, there's no clear tension between safety and proper function, or between safety and the etiological theory of functions. Nothing prevents advocates of proper functionalism from incorporating a modal reading of their core notion of proper function,[17] The two approaches are not incompatible. Rather, they are complementary and this hitherto unexplored hybrid formulation can be very beneficial to the safety theorists. Finally, and zooming out from epistemology, the combination of safety and proper functionalism appears to be also pre-theoretically plausible because it gains additional support from non-epistemic considerations. Interestingly, judgements about safety seem to align with judgements about proper function. Take for instance the case of a car that we deem to be safe. As such, the car is likely to be functioning properly and to be well designed; conversely, a malfunctioning or poorly designed car will be rather unsafe. Much like the safety condition on offer here, all these judgments also assume some implicit relativization to environments: a perfectly functioning and exceptionally designed car will not be safe if driven of a very icy road—environments always play a key role. Overall, it looks

[17] The chancy nature of natural selection that contributes to determining the etiological function of belief-forming methods might seem in tension with the anti-luck requirement on knowledge that motivates modal epistemologies. But even granting that evolution is driven by a type of luck, such luck will ultimately be of a benign kind. Borrowing from Pritchard's taxonomy (Pritchard, 2005) we can think of the chancy elements of natural selection as a type of *capacity* luck. It is indeed a matter of good luck that belief-forming methods acquired their etiological function and thereby made the agent 'capable of Pritchard (2005, p. 135), but this doesn't prevent these belief-forming methods from successfully excluding key instances of veritic luck - the type of luck which is incompatible with knowledge. I offer these remarks in support of the compatibility between safety, proper functionalism and the etiological theory of functions.

like there are rather compelling reasons to combine safety and proper functionalism both within and outwith epistemology.[18]

Let's take stock. After arguing that the safety condition stands in crucial need of clarification, I've proceeded to make it more precise by appeal to proper function and the etiological theory of functions. For the resulting safety condition, properly functioning belief-forming methods individuated by the etiological theory of functions and employed in appropriate cognitive environments will yield safe beliefs—beliefs that continue to be true in the relevant possible worlds where the standards of proper function of methods and environments are suitably met. So understood, the safety condition offered here is more informative than standard versions of safety: as such, this formulation meets the first *desideratum* outlined in Sect. 3. With this novel safety condition at hand, I now proceed to defuse the most pressing counterexamples moved to the necessity and sufficiency of safety for knowledge and the main objections against the success of safety as anti-Gettier condition. Doing so will show how this novel formulation also meets the remaining *desideratum*: in fact, it can also serve as satisfactory necessary and sufficient condition for knowledge.

4 Etiological proper function safety in action

The main objections against safety's extensional adequacy are best sorted into three categories. In the first category we find objections to the success of safety as anti-Gettier condition. Next, we find further counterexamples against its necessity (second category) and sufficiency (third category) for knowledge. I now take each category of objections in turn; I shall start with checking my new safety condition against the two most discussed Gettier-style cases.

4.1 Gettier cases: clocks and barns

Here's two familiar Gettier cases featuring stopped clocks and fake barns:

> **Stopped Clock**. You take a competent reading from a clock that you know to be usually reliable and have no reason to think is currently not working. Based on this reading you form a belief that it is noon. What's more, your belief is true: it is indeed noon. Crucially, however, the clock is broken and the reason your belief is true is that *it happened to stop working exactly twelve hours ago*. (Russell, 1948)

[18] While these are *prima facie* good reasons to combine safety and proper function, a closer scrutiny reveals that there are also reasons against the pre-theoretical appeal of this combination. A proper functionalist account of safety fails to capture the intuitive idea that a belief is safe insofar as it enjoys some kind of robust protection from error: as long as there cases of trivially safe beliefs which are not the result of a properly functioning belief-forming method, this account conflicts with the ordinary use of the term 'safe'. I will consider some of these cases below. For now, it's important to observe that this novel formulation of safety does not perfectly align with the ordinary use of the term. Thanks to an anonymous reviewer for pushing me to clarify these issues.

> **Fake Barns**. 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*. (Goldman, 1976).[19]

According to the version of safety defended here, the beliefs under consideration may be justified and true, but they don't constitute knowledge because, as emphasized in italics in the text, they fail to meet the key proper functionalist conditions on belief-forming methods and environments in the actual world. Accordingly, the beliefs under consideration lack the modal robustness necessary for knowledge. Take for instance **Stopped Clock**. Because the clock is broken, the belief is not based on a properly functioning instrument. More exactly, it does not meet its etiological function, namely (reliably) indicating the correct time. As such, the belief under consideration doesn't continue to be true in the relevant possible worlds where the belief-forming method meets a proper functionalist condition. What's more, the belief under consideration could very easily be false: if we keep the (malfunctioning) belief-forming method fixed, the subject forms a false belief in nearby worlds. The lack of proper function explains why the belief under consideration could easily be false: the instrument is malfunctioning, hence the possibility of error. Standard formulations of safety capture absence of knowledge in Gettier cases by emphasizing that the belief under consideration could easily be false, but they don't explain why exactly it could easily be false.[20] By appeal to proper function, we are now better positioned to offer such explanation: in Gettier-style cases agents could easily form a false belief *because of* a cognitive malfunction in their belief-forming method, some inappropriate feature of their cognitive environment, or both.

What about **Fake Barns**? The safety condition defended here predicts absence of knowledge in this much debated Gettier-style case, but instead of relying on some vaguely described intuitions it does so for a principled reason: the no knowledge verdict is explained by the further index of safety to appropriate (normal) environments. Even if the agent's perceptual capacities are functioning properly, the environment is clearly inappropriate. Recall that, according to the etiological theory of functions, a cognitive environment will be normal and appropriate only if it is sufficiently similar to the environment in which the item acquired its function. But since it is filled with mere replicas of real barns, the environment in **Fake Barns** is too different from the suitable normal environment in which the belief-forming method under consideration (perception) acquired its etiological function. Accordingly, given the inappropriate and abnormal environment where the belief under consideration is formed, it could easily be false: if we keep the inappropriate and abnormal environment fixed, the subject forms a false belief in nearby worlds. The modal proximity of error is again explained

[19] These cases are originally due to Bertrand Russell and Carl Ginet respectively. Here, I borrow the presentation from Kelp (2018).

[20] Pritchard (2013, p. 156): "Safety can also deal with Gettier-style cases, for these are characteristically cases in which the agent forms (on the same basis as in the actual world) a false belief in the target proposition in a close possible world." But Pritchard does not say more: we are never told *why exactly* the agent forms a false belief in the target proposition in a close world.

by appeal to the further index of safety to appropriate normal environments: the belief-forming method is functioning properly, but the environment is not appropriate for the (etiological) proper function of belief-forming method.[21]

This proper functionalist safety condition predicts absence of knowledge in the two most discussed Gettier-style cases (**Stopped Clock** and **Fake Barn**). Each of these cases features a cognitive malfunction or an inappropriate cognitive environment: this impacts on the modal profile of the beliefs under consideration, and it explains why they could easily be false or why they lack the relevant modal robustness which is necessary for knowledge.[22]

4.2 Necessity

But does a true belief even need to be safe to constitute knowledge? Some epistemologists think that the answer is 'no', and have provided several cases of unsafe knowledge. Kelp (2009); Kelp (2016); Kelp (2018) has pursued this line of argument most consistently. Leaning on Frankfurt's famous case (Frankfurt 1969), he begins by raising this type of epistemic Frankfurt case in isolation:

> **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)

The case is supposed to show that Russell knows the correct time even if he forms a false belief in nearby worlds - he could after all have checked the clock a few minutes later and looked at a stopped clock. Then, Kelp proceeds to pair this case with the already mentioned **Fake Barns**: once these two cases are raised in conjunction, it looks like the safety condition will be either too weak to predict absence of knowledge in **Fake Barns** or too strong to accommodate the knowledge intuition in **Frankfurt Clock**. Borrowing from Kelp's useful terminology (Kelp, 2016), this is the 'safety dilemma'. Safety doesn't deliver the right verdict, and given the second horn of the dilemma, it doesn't even seem to be necessary for knowledge. To back up the second

[21] Notice that neither the belief-forming method nor the environment is so specific as to include the only real barn: if it did so, safety would be trivially satisfied and thereby fail to capture the no knowledge verdict in **Fake Barns** See Pryor (2004), Ranalli (2014) Bernecker (2020) for discussion.

[22] I should notice one important limitation of this approach: appeal to the etiological theory of functions does not provide an airtight method to distinguish between appropriate and inappropriate cognitive environments. An anonymous reviewer raises a case that brings out this point clearly. Suppose that an agent A is very skilled at performing mathematical calculations, but right after receiving a piece of good news, she gets distracted and makes a mistake. No matter how the case is described, it seems hard to explain absence of knowledge only by reference to etiological proper function of belief-forming methods or appropriateness of cognitive environments.

horn of Kelp's dilemma, we can list more cases of unsafe knowledge. For instance, consider the following similar vignette:

> **Atomic Clock**. The world's most accurate clock hangs in Mia's office. The clock's accuracy is due to a clever radiation sensor. However, this radiation sensor is very sensitive and could easily malfunction if a radioactive isotope were to decay in the vicinity. This morning, against the odds, someone did in fact leave a small amount of a radioactive isotope near the world's most accurate clock in Mia's office. This alien isotope has a relatively short half-life, but - quite improbably - it has not yet decayed at all. It is 8:20 am. The alien isotope will decay at any moment, but it is indeterminate when exactly it will decay. Whenever it does, it will disrupt the clock's sensor, and freeze the clock on the reading "8:22." Therefore, though it is currently functioning properly, the clock's sensor is not safe. The clock is in danger of stopping at any moment, even while it currently continues to be the world's most accurate clock. (Bogardus, 2014, p. 12; Bogardus and Perrin, (2022, p. 4)

We notice a common feature: much like **Frankfurt Clock**, the agent could very easily form a false belief—indeed, they almost *do* form a false belief but at the very last moment they don't. Taken together, these two cases suggest that safety isn't necessary for knowledge: even if agents form false beliefs in relevantly close worlds, they nevertheless know. This is bad news for safety in general and, *a fortiori*, also for the safety condition on offer here.

For these cases to go through, it needs to be showed that the beliefs under consideration lack the relevant modal robustness necessary for knowledge. Recall that once safety is elucidated in terms of proper function, all the relevant possible worlds that matter for modal robustness are only those where both the properly functioning belief-forming method and the appropriate environments are kept fixed. As I hinted in previous work (Mortini, 2022), for the possibility of error to obtain, all these vignettes assume a change in the environment that then impacts on the proper function of the belief-forming method. In **Frankfurt Clock**, the demon has to *actually* intervene to manipulate the clock. Similarly, in **Atomic Clock**, the isotope has to *actually* decay. However, both the belief-forming method and the environment meet the proper function condition in the actual world, and once both belief-forming methods and environments are kept fixed, the subjects do continue to form true beliefs. After all, according to the etiological theory of function, the subjects are employing a properly functioning belief-forming method in an appropriate environment: because the clock is not *actually* manipulated, it does meet its etiological proper function of (reliably) indicating the correct time. And because the isotope does not *actually* decay, the cognitive environment is appropriate and normal (that is, sufficiently similar to the environment in which clocks acquired their etiological function and can function properly). In general, the beliefs under consideration display the modal robustness that is necessary for knowledge: the error possibilities trade on a change in the environment and the belief-forming methods, and in turn such changes place the worlds where the agents form false beliefs further away. As a result, the allegedly close worlds where the agents form false beliefs are irrelevant, and, since they fully meet the proper function condition, the beliefs under consideration display the modal robustness which

is necessary for knowledge. Accordingly, these counterexamples to the necessity of safety for knowledge do not land against the safety condition on offer here.[23]

However, given the central appeal to etiological proper function, it is possible to develop *novel* counterexamples to the necessity of proper function safety for knowledge, or to the necessity of proper function for safety. In the remainder of the subsection, I will consider each type of counterexample in turn. Here's a purported case of a safe true belief which amounts to knowledge in the absence of etiological proper function:

> **Expert Musicologist**. Anna is fond of classical music. She listens to two superficially similar compositions, and after some reflection, she detects a subtle difference between the two pieces. She forms the belief that while the former was composed by Schönberg, the latter was composed by Berg. Anna is exactly right: the first is by Schönberg, and the second by Berg.

This is a case of knowledge: Anna's cognitive abilities yield the true knowledgeable belief that the two pieces of music are by Schönberg and Berg respectively. However, it seems implausible to say that her belief-forming method was selected to reliably distinguish between Schönberg and Berg composition: it is not clear which is the relevant etiological proper function, and absent such a relevant proper function, the belief under consideration is unsafe and therefore not in the market for knowledge. This is a bad result, and it suggests that proper function safety is not necessary for knowledge.

To handle the counterexample, the advocate of proper function safety may pursue one of these two strategies. First, they could interpret the case as an instance of the *generality problem*, namely the problem of specifying the belief-forming method with sufficient precision. Since the generality problem is a problem for every theory of knowledge, the problem is not specific to proper function safety. This response may be plausible, though it's not fully satisfactory. A second (more promising) line of answer consists in expanding the sources of etiological proper functions and include *learning*: as Anna becomes more skilled in classical music, her belief-forming methods acquire the etiological proper function of discriminating between subtly different pieces. Advocates of the etiological theories of functions are happy to count learning as a source of functions ((Graham, 2014), Sect. 6). Once developed further, each of these strategies can address the worry that proper function safety is not necessary for knowledge.

Here's a purported case of a safe true belief which intuitively amounts to knowledge and yet it is based on a malfunctioning object:

> **Sophisticated Height Gauge**. Anna takes a reading from a sophisticated instrument that measures heights – a height gauge. The instrument works as follows: while it can provide extremely precise measurements down to the smallest nanometer, it only shows remarkably coarse-grained measurements (small,

[23] This strategy also applies to Neta and Rohrbaugh (2004), Comesaña (2005) and Baumann's (2008) further alleged cases of unsafe knowledge. Notice finally that the safety condition on offer here predicts knowledge in **Frankfurt Clock** and ignorance in **Fake Barns**; accordingly, it also rises to Kelp's dilemma outlined at the beginning of the section.

medium, high). This is because the height gauge has a first nodule that takes an initial measurement in nano-meters, a second nodule that converts the first measurement to the nearest millimetre, and a third nodule that reads big, medium or small depending on the output of the second nodule. As it happens, the second nodule is malfunctioning: it randomly varies within a full meter given the measurement taken by the first nodule. Anna points the height gauge at a mountain: the second nodule is still malfunctioning, but the instruments correctly reads "high".

To repeat, the case shows that proper function is not necessary for safety, since belief can be safe despite what looks like a major malfunction. Can the safety condition canvassed here capture this intuitive case of knowledge despite the malfunctioning object?

The case is ingenious, and it does cast doubt on the connections between proper function and safety from error.[24] However, given the different purposes of the three nodules, it seems fair to attribute a different function to each: while the second nodule's function is to convert the initial measurement, the third nodule's function is to provide a coarse-grained output. If Anna is forming beliefs on the *precise* heights of objects (to the nearest millimeter), then her beliefs will be unsafe *exactly* because of the malfunction of the second nodule. But if she is forming beliefs about whether a certain object is small, medium or high then her beliefs will be safe, as the third nodule if functioning properly. While there may be further problems with the posited connection between safety and proper function, the proposed safety condition is able to capture the correct verdict in a number of cases: counterexamples to the necessity of safety for knowledge or to the necessity of proper function for safety can be addressed.

4.3 Sufficiency

Even granting that safety is necessary, is it also sufficient for knowledge? There are two main influential objections that suggest otherwise. First, variations of meta-incoherence cases show that safe true beliefs do not amount to knowledge: even if the belief under consideration is true in many close possible worlds, subjects may still not know. Secondly, the problem of trivially safe necessary truths also shows that even if the belief under consideration is true in virtually every possible world, it still doesn't amount to knowledge. Here are these two types of cases:

Temp. Temp forms beliefs about the temperature by consulting a broken thermometer. However, Temp has a guardian angel in the room who controls the thermostat, ensuring that the room's temperature matches the reading displayed on the thermometer. Assuming the angel manipulates the thermostat in all nearby worlds, any belief Temp forms about the temperature in such worlds will be true. (Pritchard, 2012)

Necessary truths. Emma is using a calculator to count the sum of 12 x 13. As a result, she forms a true belief that 12 x 13 = 156. However, the calculator is

[24] Many thanks to an anonymous reviewer for raising this case.

actually broken and it's generating answers at random. Emma's belief is true in (every) close possible worlds. (Roland and Cogburn, 2011; Pritchard, 2012).

Safety theorists have responded to these cases in the following way. To deal with **Temp**, Pritchard (2012), has added a distinct ability condition in his analysis of knowledge. Such condition explains why Temp doesn't know (after all, his belief is not true in virtue of the exercise of a relevant cognitive ability). As far as concerns **Necessary Truths**, safety theorists have globalised their safety condition (Williamson, 2009; Pritchard, 2012). For a belief-forming method to count as safe, it has to yield true beliefs in a range of sufficiently similar propositions. Accordingly, while Emma's broken calculator yields a true belief in one mathematical proposition, given that it's broken, it fails to do so with *other* sufficiently similar but distinct mathematical propositions. Neither strategy is particularly satisfactory. On the one hand, adding a distinct ability condition concedes too much to virtue epistemology and indirectly suggests that safety fails as stand-alone modal condition on knowledge. On the other, advocates of globalized versions of safety face the difficult task to demarcate with precision the range of sufficiently similar propositions, and some globalized versions of safety may be too strong and thus not necessary for knowledge.[25] Taken together, these cases are particularly troubling, and have independently contributed to the overall decrease in popularity of safety. As a result, epistemologists have either resorted to distinct conditions (e.g., Pritchard's impure virtue epistemology) or abandoned modal conditions on knowledge altogether (e.g., explanationist epistemologies and their claimed advantage in solving the problem of necessary truths).

What does the safety condition on offer here predict about these troubling cases? Fortunately, a proper functionalist reading of safety correctly predicts absence of knowledge in each. To see why, notice that the vignettes feature clear instances of *malfunctions*: the thermometer and the calculator are broken. Moreover, in at least one case, the environment isn't appropriate for the belief-forming method: thermometers are not supposed to function in environments featuring interventions of guardian angels. Recall that, according to the etiological theory of functions, an item's appropriate (normal) environment must be sufficiently similar to the one where the item came to acquire its function. This crucial environmental condition is clearly not met given the obviously abnormal and inappropriate features of Temp's environment.[26]

With these points in play, here's how the proposed proper functionalist version of safety captures absence of knowledge in each counterexample. Start with **Necessary Truths**. Since the calculator is broken, it fails to fulfil its etiological function. As a result, it also fails to yield true beliefs in the range of propositions that it was designed to yield true beliefs in. Plausibly, a calculator's etiological function is to reliably yield true beliefs in a range of basic mathematical propositions: it will not have extraordinary computational powers, but, when functioning properly, it will be

[25] Hirvelä (2019) proposes a promising virtue-theoretic criterion of demarcation; Bernecker (2020, p. 5108) argues that globalized versions of safety are not necessary for knowledge.

[26] According to a different interpretation, the belief-forming method should be individuated with reference to the broken thermometer *and* the angel's helping hand (broken thermometer *cum* angel). In this case too, the environment is neither normal nor appropriate: helping angels are not part of the environment in which thermometers acquire their function. Thanks to an anonymous reviewer for suggesting this interpretation.

able to do many simple mathematical calculations. This key condition is not met: the broken calculator will yield false beliefs in a range of propositions specified with reference to its etiological function. As such, the belief under consideration is not safe and doesn't amount to knowledge.

Let's move on to **Temp**. The belief under consideration is formed in inappropriate circumstances: thermometers acquire their etiological function in normal environments that do not feature interventions of guardian angels. As a result, Temp's belief lacks the relevant modal robustness: in the relevant possible worlds where the environment is appropriate for the thermometer's etiological function (that is, in absence of guardian angels), Temp's belief is false. Because it lacks the relevant modal robustness which is essential to etiological proper function safety, Temp's belief is not safe and as such it doesn't constitute knowledge.[27] Accordingly, these counterexamples to the sufficiency of safety for knowledge also fail to land.

5 Comparisons and open questions

I now proceed to compare this new formulation with more standard versions of safety and with other anti-Gettier conditions more generally. I do so to bring out the distinctiveness of this proper functionalist safety condition and to mark further differences with other prominent modal and non-modal conditions on knowledge.

Let's begin with the standard versions of safety. Unlike Sosa's, Pritchard's and Williamson's formulations, my proposed safety condition is indexed to both (properly functioning) belief-forming methods and (appropriate) cognitive environments. For this reason, it is importantly different from standard versions of safety.

These differences also distinguish the safety condition on offer from further prominent conditions on knowledge. Because it doesn't incorporate any ability condition, it diverges from both impure and pure/robust versions of virtue epistemology. Recall also that the condition on knowledge developed in this paper is essentially modal: it views knowledge as a matter of (suitably understood) modal robustness across relevant possible worlds. Accordingly, it clearly contrasts with explanationist epistemologies that reject modal conditions on knowledge altogether. Finally, this version of safety is

[27] This suggests that relevant modal robustness cannot be equated simply with the possible worlds similar to the actual world where the belief under consideration is true. Since modal robustness is cashed out in terms of (etiological) proper function, the relevant possible worlds for this safety condition are only those where the belief-forming method and the environment meet a proper functionalist condition specified with reference to the etiological theory of functions. These relevant possible worlds need not be the close similar worlds where the belief is true: the proper functionalist conditions must also be met in the worlds relevant for this version of safety. Beddor and Pavese (2018, p. 69) also divorce relevant possible worlds from close possible worlds. While their ability condition is restricted to possible worlds that are "normal for the task at hand" Beddor and Pavese (2018, p. 70) this safety condition is restricted to possible worlds that are normal given the etiological function of the belief-forming method under consideration. Their 'performative' account of normality is thus different from the etiological account of normality endorsed here. For example, facts about history and natural selection in the actual world contribute to determining the modal robustness relevant for this safety condition, but they don't matter for the modal robustness of Beddor and Pavese's modalised skill condition.

offered as robust anti-Gettier condition: as such, it parts ways also with knowledge-first circular accounts of safety. This version of safety thus differs from other extant anti-Gettier conditions.[28]

Before concluding, I should flag important open questions that I could not address in this paper. Firstly, it is unclear to what extent the proposed safety condition complies with a number of independently plausible closure principles for knowledge. One initial reason to think that safety preserves closure is that, unlike sensitivity, it permits to know the denial of sceptical hypotheses. However, as a some authors have argued (e.g., Murphy, 2005; Alspector-Kelly, 2011; Goldstein and Hawthorne, forthcoming), safety still yields closure failures in several cases.[29] Whether this proper function version of safety complies with closure is a thorny topic which is best left for another occasion. Secondly, this proposed safety condition does not address important questions pertaining to the social division of epistemic labour (Hardwig, 1985). Accordingly, it also remains unclear whether a suitable relativisation to properly functioning belief-forming methods in appropriate environments can explain *epistemic dependence* and capture important cases of *collective knowledge*. While certainly pressing, fortunately these issues need not be settled here.[30]

6 Concluding remarks

In this paper, I have developed a new version of safety in terms of etiological proper function, and I have explained how this version of the safety condition withstands the most pressing objections to the necessity and sufficiency of safety for knowledge. However, just like many previously proposed versions of the safety condition, this novel version of safety may still fall prey to pressing objections and counterexamples, especially given the difficulties in providing a satisfactory criterion of individuation of appropriate environments. But, not all is lost: even so, this proper functionalist version of safety remains worthy of serious consideration. This proper functionalist version of safety does better than standard versions of safety: it is more informative, it gains motivation from a number of epistemic and non-epistemic considerations, it avoids the major counterexamples to the necessity and sufficiency of safety for knowledge and it yields the right result in the most discussed Gettier-style vignettes. For these

[28] In pointing out these differences, I don't want to overlook also important similarities. While an etiological proper function elucidation of environments is new, a more general relativisation to environments is not. Robust virtue epistemologists who defend a safety condition restricted to seat, shape and situation (SSS-safety) also (more or less implicitly) relativise safety to both belief-forming methods and environments (Sosa, 2015; Greco, 2020b). Moreover, Hirvelä (2019) may also be interpreted as relativising safety to belief-forming methods to be employed in environments appropriate for the agent's inquiry. And, as noted already, Beddor and Pavese (2018) restriction's to normal worlds also features some relativisation to normal environments. Similarly, Sosa's most recent defense of robust virtue epistemology is centred around a type of *telic* normativity (Sosa, 2021), thus making the line between robust virtue epistemology and etiological proper functionalism more blurred. While these similarities are worth emphasising, I nevertheless hope that an etiological proper function reading of safety can still advance our understanding of the safety condition.

[29] Though see Schulz (2021) for a recent argument in support of the compatibility between safety and closure.

[30] Special thanks to an anonymous reviewer for pushing me to consider these issues.

reasons, despite some lingering difficulties, the safety condition canvassed in this paper deserves to be taken seriously.

Acknowledgements Thanks to two kind, patient and conscientious reviewers for many excellent comments which have greatly improved the manuscript. I have presented this paper at the COGITO (Glasgow) work in progress epistemology seminar, the Scottish Epistemology Early Career Researchers' seminar, the Brown Bag Meeting hosted by CONCEPT (University of Cologne) and the work in progress philosophy seminar at the University of Duisburg-Essen: thanks to the audiences for their very helpful comments. Special thanks to Chris Kelp, Mona Simion, Adam Carter, Ernie Sosa, Sven Bernecker, Paul Silva, Thomas Grundmann, Mikkel Gerken, Chris Wyllard-Kyle, Ross Patrizio, Giorgia Foti, Marc Lara, Lukas Schwengerer and Sabina Dominguez Parrado for helpful feedback and guidance on the topic of the paper.

Funding Open access funding provided by University Library of Southern Denmark. I wrote this paper while in receipt of a Royal Institute of Philosophy Jacobsen studentship. I could also benefit from financial support provided by the Danmarks Frie Forskningsfund (grant number: 8018-00053B).

Declarations

Conflict of interest There are no conflicts of interests directly or indirectly related to this work submitted for publication. I haven't been employed by any organisation that could be damaged by the publication of this manuscript. This research did not involve animals. There are no financial nor no interests associated with the publication of this manuscript. The topic of this paper does not need any study-specific approval by any appropriate ethics committee for research.

References

Alfano, M. (2009). Sensitivity theory and the individuation of belief-formation methods. *Erkenntnis, 70*(2), 271–281.

Alspector-Kelly, M. (2011). Why safety doesn't save closure. *Synthese, 183*(2), 127–142.

Baumann, P. (2008). Is knowledge safe? *American Philosophical Quarterly, 45*(1), 19–30.

Beddor, B., & Pavese, C. (2018). Modal virtue epistemology. *Philosophy and Phenomenological Research, 101*(1), 61–79.

Bergmann, M. (2006). *Justification without awareness: A defense of epistemic externalism*. Oxford University Press.

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.

Bogardus, T., & Perrin, W. (2022). Knowledge is believing something because it's true. *Episteme, 19*(2), 178–196.

Bonjour, L. (1980). Externalist theories of empirical knowledge. *Midwest Studies in Philosophy, 5*(1), 53–73.

Buller, D. J. (1998). Etiological theories of function: A geographical survey. *Biology and Philosophy, 13*(4), 505–527.

Carter, A. (2016). Robust virtue epistemology as anti-luck epistemology: A new solution. *Pacific Philosophical Quarterly, 97*(1), 140–155.

Carter, A. (2022). *Stratified virtue epistemology: A defence*. Cambridge University Press.
Colaço, D., Buckwalter, W., Stich, S., & Machery, E. (2014). Epistemic intuitions in fake-barn thought experiments. *Episteme, 11*(2), 199–212.
Comesaña, J. (2005). Unsafe knowledge. *Synthese, 146*(3), 395–404.
Comesaña, J. (2007). Knowledge and subjunctive conditionals. Philosophy. *Compass, 2*(6), 781–791.
Faraci, D. (2019). Groundwork for an explanationist account of epistemic coincidence. *Philosophers' Imprint*, 19.
Frankfurt, H. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy, 66*(23), 829.
Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis, 23*(6), 121–123.
Goldman, A. (2009). Williamson on knowledge and evidence. In P. Greenough, D. Pritchard, & T. Williamson (Eds.), *Williamson on knowledge* (pp. 73–91). Oxford University Press.
Goldman, A. I. (1976). Discrimination and perceptual knowledge. *Journal of Philosophy, 73*, 771–791.
Goldstein, S., Hawthorne, J. (forthcoming). Safety, closure, and extended methods. *Journal of Philosophy*.
Goodman, J., & Salow, B. (2018). Taking a chance on kk. *Philosophical Studies, 175*(1), 183–196.
Graham, P. J. (2011). Intelligent design and selective history: Two sources of purpose and plan. In J. L. Kvanvig (Ed.), *Oxford studies in philosophy of religion* (Vol. 3, pp. 67–88). Oxford University Press.
Graham, P. J. (2012). Epistemic entitlement. *Noûs, 46*(3), 449–482.
Graham, P. J. (2014). Functions, warrant, history. In A. Fairweather & O. Flanagan (Eds.), *Naturalizing epistemic virtue* (pp. 15–35). Cambridge University Press.
Graham, P. J. (forthcoming) Proper functionalism and the organizational theory of functions. In L. R. G. Oliveira, ed., *Externalism about knowledge*. Oxford University Press, forthcoming.
Greco, J. (2010). *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge University Press.
Greco, J. (2020). Safety in sosa. *Synthese, 197*(12), 5147–5157.
Greco, J. (2020). *Knowledge-producing abilities,* (pp. 124–146). Cambridge University Press.
Grundmann, T. (2018). Saving safety from counterexamples. *Synthese, 197*(12), 5161–5185.
Hardwig, J. (1985). Epistemic dependence. *Journal of Philosophy, 82*(7), 335–349.
Hirvelä, J. (2019). Global safety: How to deal with necessary truths. *Synthese, 196*(3), 1167–1186.
Hirvelä, J., & Paterson, N. (2021). Need knowing and acting be sss-safe? *Thought: A Journal of Philosophy, 10*(2), 127–134.
Holliday, W. H. (2015). Epistemic closure and epistemic logic i: Relevant alternatives and subjunctivism. *Journal of Philosophical Logic, 44*(1), 1–62.
Horvath, J., & Nado, J. (2021). Knowledge and normality. *Synthese, 198*(12), 11673–11694.
Jenkins, C. S. (2006). Knowledge and explanation. *Canadian Journal of Philosophy, 36*(2), 137–164.
Kelp, C. (2009). Knowledge and safety. *Journal of Philosophical Research, 34*, 21–31.
Kelp, C. (2013). Knowledge: The safe-apt view. *Australasian Journal of Philosophy, 91*(2), 265–278.
Kelp, C. (2016). Epistemic frankfurt cases revisited. *American Philosophical Quarterly, 53*(1), 27–37.
Kelp, C. (2017). Knowledge first virtue epistemology. In A. Carter, E. Gordon, & B. Jarvis (Eds.), *Knowledge first: Approaches in epistemology and mind.* Oxford University Press.
Kelp, C. (2018). *Good thinking: A knowledge first virtue epistemology*. Routledge.
Korman, D. Z., & Locke, D. (2020). Against minimalist responses to moral debunking arguments. *Oxford Studies in Metaethics, 15*, 309–332.
Lasonen-Aarnio, M. (2010). Unreasonable knowledge. *Philosophical Perspectives, 24*(1), 1–21.
Lehrer, K. (1990). *Theory of knowledge*. Westview Press.
Littlejohn, C. (2014). Fake barns and false dilemmas. *Episteme, 11*(4), 369–389.
Littlejohn, C., & Dutant, J. (2020). Justification, knowledge, and normality. *Philosophical Studies, 177*(6), 1593–1609.
Lycan, W. G. (2006). On the gettier problem problem. In S. Hetherington (Ed.), *Epistemology futures* (pp. 148–168). Oxford University Press.
McGlynn, A. (2012). The problem of true-true counterfactuals. *Analysis, 72*(2), 276–285.
Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. MIT Press.
Miracchi, L. (2015). Competence to know. *Philosophical Studies, 172*(1), 29–56.
Mortini, D. (2022). A new solution to the safety dilemma. *Synthese, 200*(2), 1–17.
Murphy, P. (2005). Closure failures for safety. *Philosophia, 33*(1–4), 331–334.
Neta, R. (2002). S knows that p. *Noûs, 36*(4), 663–681.

Neta, R., & Rohrbaugh, G. (2004). Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly, 85*(4), 396–406.
Nolan, D. (2014). Hyperintensional metaphysics. *Philosophical Studies, 171*(1), 149–160.
Nozick, R. (1981). *Philosophical explanations*. Harvard University Press.
Peet, A., & Pitcovski, E. (2018). Normal knowledge: Toward an explanation-based theory of knowledge. *Journal of Philosophy, 115*(3), 141–157.
Peter, J. (2012). Graham. Epistemic entitlement. *Noûs, 46*(3), 449–482.
Plantinga, A. (1993). *Warrant and proper function*. Oxford University Press.
Pritchard, D. (2005). *Epistemic luck*. Oxford University Press.
Pritchard, D. (2007). Anti-luck epistemology. *Synthese, 158*(3), 277–297.
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.
Pritchard, D. (2013). Knowledge cannot be lucky. In M. Steup & J. Turri (Eds.), *Contemporary debates in epistemology.* Blackwell.
Pritchard, D. (2016). Epistemic risk. *Journal of Philosophy, 113*(11), 550–571.
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. In: T. Poston, ed., *Internet Encyclopedia of Philosophy*.
Ranalli, C. (2014). Luck, propositional perception, and the entailment thesis. *Synthese, 191*(6), 1223–1247.
Roland, J., & Cogburn, J. (2011). Anti-luck epistemologies and necessary truths. *Philosophia, 39*(3), 547–561.
Russell, B. (1948). *Human knowledge: Its scope and limits*. Routledge.
Sainsbury, R. M. (1997). Easy possibilities. *Philosophy and Phenomenological Research, 57*(4), 907–919.
Schellenberg, S. (2018). *The unity of perception: Content, consciousness, evidence*. Oxford University Press.
Schulz, M. (2021). Finding closure for safety. *Episteme, 18*(4), 711–725.
Simion, M. (2016). Perception, history and benefit. *Episteme, 13*(1), 61–76.
Simion, M. (2019). Knowledge-first functionalism. *Philosophical Issues, 29*(1), 254–267.
Smith, M. (2016). *Between probability and certainty: What justifies belief*. Oxford University Press.
Sosa, E. (1999). How to defeat opposition to moore. *Philosophical Perspectives, 13*, 137–49.
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). *Judgement and agency*. Oxford University Press.
Sosa, E. (2021). *Epistemic explanations: A theory of telic normativity, and what it explains*. Oxford University Press.
Turri, J. (2016). A new paradigm for epistemology from reliabilism to abilism. *Ergo: An Open Access Journal of Philosophy, 3*, 1.
Williamson, T. (1994). *Vagueness*. Routledge.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.
Williamson, T. (2009). Reply to goldman. In D. Pritchard & P. Greenough (Eds.), *Williamson on knowledge* (pp. 305–312). Oxford University Press.
Wright, L. (1973). Functions. *Philosophical Review, 82*(2), 139–168.