UC Irvine
UC Irvine Previously Published Works

Title

In defence of modest anti-luck epistemology

Permalink

https://escholarship.org/uc/item/64p736bn

Author

Pritchard, Duncan

Publication Date

2012-08-02

DOI

10.1017/cbo9780511783630.014

Copyright Information

Peer reviewed

In favor of safety over sensitivity

In defence of modest anti-luck epistemology

Duncan Pritchard

I ANTI-LUCK EPISTEMOLOGY

Most epistemologists would accept that knowledge excludes luck in the specific sense that if one knows then it is not a matter of luck that one's belief is true.[1] Call this the *anti-luck intuition*. There is a certain kind of epistemological project – which I have christened *anti-luck epistemology* – which takes this intuition as central to our understanding of knowledge.[2] Essentially, the idea is that once we identify which epistemic condition can satisfy the anti-luck intuition (call this the *anti-luck condition*), then we will have thereby identified a key component in a theory of knowledge. Central to this enterprise, as I explain below, is to gain a proper understanding of the nature of luck itself.

We can distinguish between two forms of anti-luck epistemology. According to *robust anti-luck epistemology*, knowledge is nothing more than true belief that satisfies the anti-luck condition. According to *modest anti-luck epistemology*, in contrast, the anti-luck condition is merely a key necessary condition for knowledge, but it is not sufficient (with true belief) for knowledge. In what follows I will be offering a defence of modest anti-luck epistemology.

[1] Most, but not all. See note 8, below.
[2] See especially Pritchard (2005a, 2007a; cf. Pritchard 2004, 2005b, 2007b, 2008a, 2008b, 2009b, 2011). For a key precursor to this approach, see Unger (1968).

II SAFETY VERSUS SENSITIVITY

There are two competing ways of understanding the anti-luck condition in the contemporary literature. Call the *safety principle* the claim that knowledge entails *safe* belief, and call the *sensitivity principle* the claim that knowledge entails *sensitive* belief. In order for an agent's belief (formed on a certain basis) to be safe it needs to be a true belief which could not have easily been false (and yet formed on the same basis).[3] Safety is usually cashed out in modal terms as demanding that an agent has a true belief such that, in nearby possible worlds, insofar as the agent forms her belief on the same basis as in the actual world, then her belief continues to be true.[4] In contrast, in order for a true belief (formed on a certain basis) to be sensitive, it must be such that, had what the agent believed been false, she wouldn't have believed it (on the same basis).[5] Sensitivity is usually cashed out in modal terms as demanding that in the closest possible world in which what the agent actually believes is false, the agent no longer believes it on the same basis as in the actual world.[6]

For a wide range of cases involving knowledge-undermining epistemic luck, these two conditions perform equally well. Consider, for example, a standard Gettier-style case:

SHEEP: Roddy, in good epistemic conditions – in good light, at close range, and so on – sees what he takes to be a sheep, and so forms the belief that there is a sheep in the field. While this belief is true, in that there is a sheep in the field, Roddy is not looking at a sheep but rather a sheep-shaped object (such as a hairy dog). The genuine sheep is hidden from view behind the sheep-shaped object.[7]

Most epistemologists hold that knowledge is lacking in standard Gettier-style cases in virtue of the knowledge-undermining epistemic luck in play. It is, after all, just a matter of luck that Roddy's belief is true in this

---

[3] Versions of safety-type principles have been offered by a number of authors, including Luper-Foy (1984; cf. Luper 2003); Sainsbury (1997); Sosa (1999b); Williamson (2000) and Pritchard (2002, 2005a, 2007a).

[4] We will be looking in more detail at how best to unpack safety below.

[5] For the key texts in this regard, see Dretske (1970, 1971) and Nozick (1981). For some recent texts which sympathetically explore the sensitivity principle, see Roush (2005); Becker (2007); Black and Murphy (2007; and Black (2008).

[6] Both safety and sensitivity are expressed here in a basis-relative form, as is standard in the literature. Note that for reasons of space I will not be exploring the reasons why we need to opt for a basis-relative formulation of these notions, and neither will I be offering an elucidation of what is involved in being a 'basis' for belief. For a very helpful recent discussion of these issues, see Williamson (2000, ch. 7). Note too that I will be following common practice and understanding these principles as relativized to a certain time.

[7] This is a variation on a famous example given by Chisholm (1977, 105).

case, in that it is true in virtue of the happenstance that there is a genuine sheep hidden from view behind the sheep-shaped object.[8]

Roddy's true belief in SHEEP is both unsafe and insensitive, and hence both construals of the anti-luck condition can handle this case (and analogous cases). The belief is unsafe because it could very easily have been false. That is, there are close possible worlds where Roddy continues to believe (on the same basis) that there is a sheep in the field (because he is still looking at the sheep-shaped object), but where this belief is false because the sheep in question has moved to a different field. The belief is insensitive because in the closest possible world where there is no sheep in the field – that is, the world in which the sheep has moved into another field, but everything else, including the presence of the sheep-shaped object, remains the same – Roddy would have continued to believe (on the same basis) that there is a sheep in the field regardless.

But while safety and sensitivity fare equally well when it comes to a range of cases involving knowledge-undermining epistemic luck, there is at least one respect in which safety fares much better than sensitivity when it comes to being compatible with genuine cases of knowledge. That is, there is a wide range of beliefs which intuitively count as knowledge, and which involve safe belief, but where the belief in play is intuitively *in*sensitive.

This problem concerns inductive knowledge.[9] Consider the following case:

CHUTE: Ernie drops a bag of rubbish into the garbage chute next to his high-rise apartment, and a few moments later forms the true belief that the rubbish is now in the basement. The rubbish chute is in fact very reliable in this regard – indeed, it has never failed to deliver rubbish to the basement, over a long history – and it is well maintained and serviced. Ernie knows about all of this. Moreover, there is nothing amiss with the rubbish chute on this occasion, nor

---

[8] Although most epistemologists grant that agents lack knowledge in Gettier-style cases, there are some detractors. See, especially, Hetherington (1998, 2002, ch. 1). For a very useful recent discussion of Hetherington's view, see Madison (2011). For a recent exchange on this issue, see Hetherington (in press) and Pritchard (in press). Note that a complication in this regard is that some epistemologists, while granting that knowledge is in general lacking in standard Gettier-style cases, are inclined to argue that it is possessed in particular kinds of Gettier-style case. The main focus of attention in this regard has been the 'barn façade' example, put forward by Goldman (1976), but credited to Carl Ginet. See, especially, Sosa (2007, ch. 5; cf. Gendler and Hawthorne 2005). I critically discuss Sosa's reasons for ascribing knowledge in this case in Pritchard (2009a, 2011).

[9] There are actually quite a few problems facing sensitivity, but I focus on the problem of inductive knowledge because I think it is particularly pressing. For a more thorough discussion of some of the problems faced by the sensitivity principle, see Pritchard (2008b).

any reason for Ernie to worry about the reliability of the rubbish chute in this specific instance.[10]

Intuitively, Ernie has knowledge in this case. Indeed, if Ernie doesn't have knowledge, then it would appear that inductive knowledge is very hard to come by, since Ernie's inductive basis for his true belief is about as good as an inductive basis can be. And yet the belief is clearly insensitive, in that if the rubbish hadn't made it to the basement for some reason – had it somehow snagged on something on the way down, say – then Ernie would clearly continue to believe that his rubbish was in the basement regardless since his inductive basis for this belief would be unchanged.

In contrast, Ernie's belief is certainly safe. For given the general reliability of the rubbish chute to deliver rubbish to the basement, the fact that it is well maintained and regularly serviced, and given also that there is nothing amiss with the rubbish chute on this particular occasion, then it couldn't have easily been the case that the rubbish did not make it into the basement. Ernie's true belief thus could not have easily been false.

I think that this problem for sensitivity is fairly formidable, though, of course, it is not a lethal blow. Now, as we will see below, there are problems facing safety too. But as I will be explaining in a moment, once we understand safety properly within the context of a modest anti-luck epistemology, then it can deal with these problems. Furthermore, both the general idea behind safety, and the specific formulation of safety that we will settle upon, can be motivated in terms of the theory of luck that forms part of the methodology of anti-luck epistemology. All things considered, then, safety has considerable merits over sensitivity when it comes to offering the best rendering of the anti-luck condition.[11]
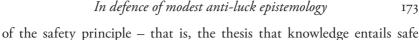
III SAFETY-BASED EPISTEMOLOGY

As noted above, the basic idea behind safety is that one has a true belief which could not have easily been false, where this is usually cashed out as the claim that one has a true belief such that, in close possible worlds, if one continues to form a belief on the same basis as in the actual world, then one's belief continues to be true. As we will see, the plausibility

---

[10] This counterexample to sensitivity is due to Sosa (1999b).
[11] There is a further dialectical option in this regard, one that I am quite sympathetic towards but which I have not the space to explore further here. This is that safety and sensitivity, at least when properly formulated, are basically extensionally equivalent, in that they both deliver the same verdicts with regard to specific cases.

of the safety principle – that is, the thesis that knowledge entails safe belief – very much depends on how we understand safety.

In order to see this point, consider the following case:

LOTTERY: Lottie and Luttie have each bought a ticket for a fair lottery with very long odds of a million to one. The lottery has been drawn, but neither agent has heard the result. Lottie reflects on the fact that the odds are massively stacked against her and so, solely on this basis, forms the (true) belief that her ticket has not won. Luttie, in contrast, doesn't even know what the odds for the lottery are, and certainly isn't the sort of person to reflect on how these odds are stacked against her. But she reads the result of the lottery in a reliable newspaper, and so on this basis forms the (true) belief that her ticket has not won.

Here's the thing. Intuitively, Lottie doesn't know that she has lost while Luttie does. But this is puzzling, since the odds in favour of Luttie being right are astronomically in her favour. Indeed, even though reliable newspapers are very careful when it comes to printing lottery numbers (for obvious reasons – think of the problems that they would face if they printed the wrong result), nonetheless the probability that these results have been misprinted is surely higher than the astronomical probability that one's ticket wins this lottery. So how then can it be that Luttie has knowledge in this case but Lottie doesn't?

Interestingly, the sensitivity principle offers us a very attractive way of dealing with the lottery problem, for notice that while Luttie's belief is sensitive, Lottie's belief is not. The closest possible world where what these agents actually believe is false is where they are in possession of a winning lottery ticket. Crucially, though, while this will be a world in which the reliable newspaper prints the winning result, it will continue to be a world in which the odds in question overwhelming suggest that one has lost. Thus, if one forms one's belief about whether one has lost on the basis of the odds concerned (as Lottie does), then one will form a false belief in this world; but if one forms one's belief by consulting a reliable newspaper (as Luttie does), then one will form a true belief. It is in this sense, claim sensitivity theorists, that Lottie's belief that she has lost the lottery is only luckily true, even though the odds are massively in her favour when compared with Luttie, who, by consulting the reliable newspaper, really does know that she has lost.

But while the sensitivity principle offers a very compelling way of dealing with the LOTTERY case, safety can also handle this example, though we need to be a little more precise in how we understand this notion in order to see this. The formulation above talks simply of the

agent continuing to form a true belief on the same basis as in the actual world across nearby possible worlds. But this naturally prompts the question of how extensive this range of nearby possible worlds should be, and to what extent, if any, safety is consistent with there being some false beliefs formed within this range of possible worlds. On the face of it, there is a dilemma in play here.[12]

On the one hand, the LOTTERY case would seem to suggest that safety ought to demand that the agent does not form a false belief in *any* (or at least hardly any) of the nearby possible worlds. A lottery win, while the kind of thing that does occur in nearby possible worlds (all that needs to be different, after all, is that a few coloured balls fall in a different configuration), is clearly not something that generally occurs in nearby worlds. But, on the other hand, most normal cases of knowledge do not seem to make such an austere modal demand, and hence a rendering of safety that was this strong might be in conflict with a range of cases which we intuitively regard as instances of knowledge. For example, couldn't we imagine a version of CHUTE where it is plausible that there might be a small class of nearby possible worlds where Ernie believes (on the same basis) falsely, but where intuitively he nonetheless has knowledge?

The dilemma is thus between a weak version of safety which accords with our ordinary judgements about when knowledge is possessed, but which does not deliver the right result in the LOTTERY case, and a strong version of safety that can handle the LOTTERY case, but which does not deliver the right result in a range of ordinary cases of knowledge.

In order to resolve this dilemma, we need to think about how the motivation for safety is rooted within a certain account of the nature of luck itself. Very roughly, lucky events are events which obtain in the actual world but which don't obtain in a wide class of nearby possible worlds where the initial conditions for that event remain (sufficiently) in play. A lottery win, for example, is a paradigm case of a lucky event in that while it obtains in the actual world, in most nearby possible worlds where the relevant initial conditions for that event remain the same (e.g., the lottery continues to be run fairly) it fails to obtain.[13]

This way of thinking about luck motivates the view that safety is the right way to think about the anti-luck condition. For the lucky event

---

[12] Which we might refer to as 'Greco's dilemma', since he was the first (so far as I know) to pose it for safety-based theories of knowledge. See Greco (2007).

[13] I develop this account of luck in a number of places. See, especially, Pritchard (2005a, ch. 5). See also Pritchard and Smith (2004). For some recent discussions of this proposal, see Coffman (2007); Riggs (2007, 2009); Lackey (2008); and Levy (2009).

that we are trying to eliminate is where it is a matter of luck that one's belief is true. That is, with this account of luck in mind, to say that one's belief is only luckily true is to say that while it is true in the actual world, in a wide range of nearby possible worlds where what gave rise to that belief – that is, the 'basis', as we have been calling it – is the same, the event of one having a true belief fails to obtain. (Instead, one believes falsely.) So, in a Gettier-style case like SHEEP, for example, Roddy happens to believe truly in the actual world, but in a wide range of nearby possible worlds where the basis for his belief is kept fixed (i.e., he continues to see the sheep-shaped object), he believes falsely (because the sheep is no longer in the field). But, of course, this way of spelling out the nature of luck in the epistemic case is directly leading us to a version of the safety principle.

Moreover, by locating safety within the context of a theory of luck, we are also in a position to motivate a specific rendering of this notion, one that can help us deal with the dilemma posed above. For notice that the extent to which an event is lucky is a function of how modally close the non-obtaining of that event is. For example, consider an agent who is very nearly shot by a sniper, with the bullet whizzing by just a few feet away from him. That he is not shot is lucky, because there are close possible worlds where he is shot. All other things being equal, had the bullet passed by him quite a few feet away, then the event would have been less lucky (since the world in which he is shot is now modally more remote), and had the bullet passed by him only inches away, then the event would have been more lucky (since the world in which he is shot is now modally closer). Put in terms of the notion of risk, the agent in the first case (where the bullet was some way off) was at less risk of being shot than in the second case (where the bullet was very close).

With this point in mind, we should not be surprised that the modal closeness of the relevant error (i.e., the agent forming a belief, on the same basis as in the actual world, which is false) has a direct bearing on how safe the belief is and thus on whether the belief amounts to knowledge. In cases where the possibility of error is very close, then the belief will be subject to a high degree of epistemic risk and hence very unsafe; while in cases where the possibility of error is further out the belief will be subject to a much lower degree of epistemic risk, and hence will be more likely to qualify as safe.

Safety is thus capturing an intuition about our tolerance of the risk of error. In the very closest nearby possible worlds we are extremely intolerant when it comes to such epistemic risk, and so would not want to be

forming any false beliefs on the target basis. In far-off possible worlds, however, we are extremely tolerant about such epistemic risk, on account of their modal remoteness. In-between we have a descending scale of epistemic intolerance, from extreme intolerance to epistemic risk to extreme tolerance. When we say that a belief is safe, we are saying that epistemic risk has been excluded to a sufficient degree that the belief is (on this score at least) in the market for knowledge. With this point in mind, we need to think of safety as completely excluding false belief in the very closest possible worlds, but becoming increasingly tolerant to such falsity as one moves further away from the actual world. That is, safety is compatible with there being some false belief in nearby possible worlds, just so long as those worlds are not especially close.

We can see this point in action in the lottery case. What is crucial to this case is that the modal remoteness of error is very different for Lottie and Luttie. All it takes for Lottie to form a false belief is for a few coloured balls to fall in a slightly different configuration. In contrast, what is required for Luttie to form a false belief is a range of mishaps at the newspaper office (e.g., the person inputting the results, despite taking lots of care in doing so, somehow makes a mistake, a mistake that is not spotted when the various copy-editors, hired and retained for their conscientiousness in such matters, somehow collectively fail to spot the error). This is why Lottie's belief is unsafe, and hence not knowledge, while Luttie's belief is safe and hence is in the market for knowledge.[14]

Moreover, once we think of safety in this way, then there is no tension with ordinary cases of knowledge, such as inductive knowledge. We noted earlier that on the face of it we could imagine a version of the CHUTE case where it was plausible that there be some nearby possible worlds where Ernie forms his belief on the same inductive basis and yet believes falsely. I think that's right, but notice that such error had better not be taking place in very close possible worlds, which is what happens in the LOTTERY case (i.e., when it comes to Lottie's belief).

For example, suppose the reason for the error is that there is an imperfection in the shaft of the chute such that the rubbish very nearly snags on it each time (but so far hasn't). In this case the error in question would be modally very close, and thus analogous to the degree of modal closeness of error when it comes to Lottie's belief. Crucially, however,

---

[14] For scepticism about the prospects of a safety-based account of knowledge dealing with the lottery problem, see Greco (2007) and McEvoy (2009). For responses, see (respectively) Pritchard (2007a, 2009b).

on this reading of the example I take it that there is no longer any reason to think that Ernie has knowledge, since his cognitive success is just too lucky. Safety thus delivers the same result both in this case and the Lottie case.

But that means that the error must be modally much further out if Ernie is to quality as having knowledge. So construed, however, we can allow that Ernie has knowledge without this causing problems for our diagnosis of the LOTTERY case. In particular, allowing that Ernie has knowledge under this reading of CHUTE is entirely compatible with claiming that Lottie lacks knowledge.

So once we understand safety correctly – and, in particular, once we set our understanding of safety within the context of a modal conception of luck – then it is not at a disadvantage relative to sensitivity when it comes to dealing with cases like LOTTERY.

There is a further kind of case which seems on the face of it to pose problems for safety, which concerns our beliefs in necessary propositions.[15] One can see how the objection would run. Such propositions are true in all nearby possible worlds, and hence all one needs to do is happen to form a true belief in a necessary proposition in the actual world and – hey presto! – one has a belief which is necessarily safe. This is because in such a case there cannot by definition be nearby possible worlds where one continues to form this belief (on the same basis) and yet believes falsely.

Moreover, notice that opting for sensitivity would not offer one a route out of this problem, since on the face of it this notion is subject to the very same difficulty – that is, that a belief in a necessary proposition is necessarily sensitive. This is because in such a case there cannot by definition be a closest possible world where what one believes is false but one believes it (on the same basis) regardless.

Since I am defending only the safety principle here, I will focus my attentions on this problem as it affects this thesis. What is crucial in this regard is that one should not evaluate the safety of a belief by focusing only on nearby possible worlds where the agent continues (on the same basis as in the actual world) to form a belief *in the very same proposition* as in the actual world. Rather, what one is interested in is the truth-value of the belief that is formed in nearby possible worlds on the same basis as in the actual world, even when the resulting belief is not of the same proposition.

---

[15] Or, indeed, our belief in any proposition which is true across all nearby possible worlds, even if not necessarily true.

In order to see this, imagine that an agent forms a true belief in a mathematical proposition – that 2 + 2 = 4, say – by flipping a coin. Since there is no possible world where the proposition believed is false, there is thus trivially no nearby possible world in which the agent believes *this specific proposition* and believes falsely. But that doesn't mean that the belief is thereby safe, and the reason for this is that there are lots of nearby possible worlds where the agent's actual way of forming her belief – that is, by flipping a coin in order to determine mathematical truths – leads to false belief, such as the possible world where the coin toss prompts her to believe that 2 + 2 = 5. The key point here is that in assessing whether a belief that *p* is safe, we are interested in whether the agent forms a belief in the same way in nearby possible worlds and believes falsely, but this is different from being interested in whether the agent forms a belief *that p* in nearby possible worlds and believes falsely.

Again, then, we see that the safety principle is highly plausible so long as we understand safety correctly.

IV MODEST VERSUS ROBUST ANTI-LUCK EPISTEMOLOGY

Earlier I distinguished between modest and robust anti-luck epistemology. The former merely endorses the safety principle and hence argues that safety is a key necessary condition for knowledge. The latter, by contrast, makes the much stronger claim that, provided we have formulated the anti-luck condition correctly, we ought to have an epistemic condition which is sufficient, with true belief, for knowledge. So, for example, if one agrees with me that safety (as I am conceiving of this notion, anyway) offers the best formulation of the anti-luck condition, then robust anti-luck epistemology would be the view that knowledge is safe true belief. I noted that I am interested only in defending modest anti-luck epistemology, and this is a good juncture to explain why.

I maintain that aside from the anti-luck intuition we also have a strong intuition that knowledge reflects ability, in the sense that when an agent has knowledge, then her cognitive success (i.e., her true belief) is to some significant degree creditable to her cognitive agency (i.e., her exercise of her cognitive abilities). Call this the *ability intuition*.[16] Now one might think that the ability intuition is just the other side of the coin of the anti-luck intuition, in that they are both tracking the same overarching

---

[16] I offer a more detailed defence of the ability intuition in Pritchard (2011), where I examine this intuition in play in the contemporary epistemological literature.

intuition. That is, one might think that the fact that one's cognitive success is significantly creditable to one's cognitive agency is just to say that it is not down to luck. And, conversely, one might think that the fact that one's cognitive success is not down to luck is just to say that it is significantly creditable to one's cognitive agency.

On closer inspection, however, it is clear that these two intuitions are not just two aspects of the same master intuition. In fact, they come apart in both directions, in that there are both cases of cognitive success which are significantly creditable to the agent's cognitive ability and yet lucky (unsafe) nonetheless, and cases of cognitive success which are non-lucky (safe) and yet not significantly creditable to the agent's cognitive ability. Our focus here, however, will be on the latter direction of fit, since it is these cases which demonstrate the falsity of robust anti-luck epistemology.[17]

Consider the following case:

COIN: René forms his beliefs about what the weather will be tomorrow purely on the basis of flipping a coin, since he has been assured by someone whom he trusts (but who was in fact lying to him) that this is the best way to form one's beliefs about this subject matter. As it happens, René's beliefs, so formed, are guaranteed to be true since there is a helpful demon watching René and who desires it to be the case that all his beliefs about tomorrow's weather are true. Accordingly, the demon ensures that tomorrow's weather always accords with what René believes it will be.

Clearly, René cannot gain knowledge by guesswork in this way. Note, however, that the problem here isn't that René's beliefs are only luckily true, since given the interference of the helpful demon they are in fact *guaranteed* to be true, and hence can't help but be safe (and, for that matter, sensitive, too). Instead, the problem is that the cognitive success on display in no way reflects René's cognitive abilities, but is rather simply the result of the interference of the helpful demon. Put simply, although there is a perfect match-up between belief and fact across the relevant possible worlds (such that René always believes truly in this regard), there is the wrong direction of fit in play for knowledge, in that the facts are changing to fit with what René believes rather than René's beliefs being responsive to the facts.

---

[17] For the argument for the claim that there are cases of cognitive success which are significantly creditable to an agent's cognitive ability, but which are non-lucky (i.e., unsafe) nonetheless, see Pritchard (2011; cf. Pritchard, Millar and Haddock 2010, ch. 3).

Given that the problem in play here does not concern knowledge-undermining epistemic luck, and given that the role of safety is to exclude such luck, it should be clear that it is not a failing of safety that it is unable deal with such cases.[18] Rather, what such cases remind us is that there is more to knowledge than safe true belief. In particular, at the very least what is also required is some sort of ability condition on knowledge, where this is an epistemic condition which captures the idea that one's cognitive success is significantly creditable to one's cognitive agency.[19] But that means that robust anti-luck epistemology is untenable as a theory of knowledge, and thus that at most we should be defending a modest version of anti-luck epistemology.[20]

Still, modest anti-luck epistemology is highly plausible. Moreover, as we will see below, making this distinction between modest and robust anti-luck epistemology is important for our purposes since we need to be sure that a putative counterexample to the necessity of safety for knowledge is not in fact trading on something other than the anti-luck intuition.

V IN DEFENCE OF MODEST ANTI-LUCK EPISTEMOLOGY

A range of putative counterexamples have been put forward to the idea that knowledge entails safety, and thus to the view that we are here characterizing as modest anti-luck epistemology. I will here consider a representative sample to explain why they fail to achieve their intended aim.[21]

The first is due to Ram Neta and Guy Rohrbaugh (2004):

---

[18] In any case, no formulation of safety could exclude such cases. The reason for this is that a modal principle like safety cannot capture the 'direction of fit' between belief and fact that is key to the ability intuition by virtue of how it simply specifies a match between belief and fact across a range of worlds. For more on this point, see Pritchard, Millar and Haddock (2010, ch. 3) and Pritchard (2011).

[19] I have argued elsewhere for a view I call *anti-luck virtue epistemology* according to which knowledge is essentially safe true belief plus a further epistemic condition (an 'ability' condition) which handles the ability intuition. See, for example, Pritchard, Millar and Haddock (2010, ch. 3) and Pritchard (2011).

[20] Although I did not explicitly endorse robust anti-luck epistemology in Pritchard (2005a, 2007a), I think it is implicit in those texts that I thought the view was at least viable as a theory of knowledge. As is clear from the foregoing, however, I now hold only that modest anti-luck epistemology is a viable position in this regard (which of course is not a complete theory of knowledge). The particular theory of knowledge that I now endorse is called *anti-luck virtue epistemology* (see note 19, above).

[21] Note that these cases are often put forward against particular formulations of the safety principle, but our interest will be whether they work against the specific formulation we offer here.

WATER: 'I am drinking a glass of water which I have just poured from the bottle. Standing next to me is a happy person who has just won the lottery. Had this person lost the lottery, she would have maliciously polluted my water with a tasteless, odorless, colorless toxin. But since she won the lottery, she does no such thing. Nonetheless, she *almost* lost the lottery. Now, I drink the pure, unadulterated water and judge, truly and knowingly, that I am drinking pure, unadulterated water. But the toxin would not have flavored the water, and so had the toxin gone in, I would still have believed falsely that I was drinking pure, unadulterated water … Despite the falsity of my belief in the nearby possibility, it seems that, in the actual case, I know that I am drinking pure, unadulterated water.' (Neta and Rohrbaugh 2004, 399–400)[22]

My initial reaction to such a case is to hold that it is simply not a case of knowledge. Is it really intuitive that the agent in WATER could gain knowledge that what she is drinking is water even despite the clear epistemic unfriendliness of her environment? After all, could so very easily have been drinking the toxin rather than the water, and *ex hypothesi* she wouldn't have been able to tell the difference. Indeed, were our agent to discover just how lucky it was that she formed a true belief in this case, then surely she wouldn't ascribe knowledge to herself. But of course, if this isn't a case of knowledge, then it can't be a counterexample to the necessity of safety for knowledge.

There is, however, an obvious dialectical drawback to dismissing such a case out of hand. After all, it is clear that others have found this example compelling, and so we are in danger of merely trading opposing intuitions here. Fortunately, I think we can diagnose why someone might hold that the protagonist in WATER has knowledge, even though (so say I anyway) she doesn't.

Consider the following case:

OVERHEAR: Purely by luck, Peter happens to be passing at just the right moment to clearly overhear a conversation that two of his senior colleagues are having. As a result, he gets to hear that the firm will be making 5 per cent budget cuts this year, and so believes this proposition on this basis.[23]

I take it that it is uncontroversial that Peter gains knowledge in this case. For although there is a sense in which the knowledge in question is lucky, the luck in play is not of the knowledge-undermining sort which concerns epistemologists (and which safety is designed to exclude). That is, while it is a matter of luck that Peter is in a position to acquire knowledge in this case,

---

[22] See also the very similar counterexample to the necessity of safety for knowledge offered in Hiller and Neta (2007, 310–11).

[23] This example is adapted from one offered by Unger (1968, 159) to make the same point.

it is not a matter of luck that what he believes is true (i.e., his true belief, so formed, is safe), and it is only this latter type of epistemic luck which is knowledge-undermining. The former kind of epistemic luck, in contrast, is entirely compatible with knowledge, as OVERHEAR illustrates.

If one thought that WATER was relevantly analogous to OVERHEAR, then one might be inclined to ascribe knowledge in this case. There is certainly a surface similarity between the two cases, in that one might think that just as it is a matter of luck that the agent happens to overhear the conversation in OVERHEAR, so it is a matter of luck that the agent happens to drink the uncorrupted water in WATER. Crucially, however, there is a key difference. For although it is a matter of luck that the agent in OVERHEAR overhears what he does, nonetheless he is in a great position to epistemically exploit this opportunity (since he hears what is said so clearly). But the same is not true of the agent in WATER. After all, what looks and tastes like water in her environment need not be water. This is why the agent's belief in WATER is unsafe, but the agent's belief in OVERHEAR is safe. I suggest that once we understand the difference between these two types of case, one can see why some epistemologists might be inclined to ascribe knowledge in a case like WATER, even though knowledge isn't in fact possessed by the agent concerned.[24]

A more interesting case is offered by Christoph Kelp (2009), which we can express as follows:

DEMON: A demon wants our hero – let's call him 'Chris' – to form the belief that the time is 8.22 a.m. when he comes down the stairs first thing in the morning (the demon doesn't care whether the belief is true). Since he is a demon, with lots of special powers, he is able to ensure that Chris believes this proposition (e.g., by manipulating the clock). Now suppose that Chris happens to come downstairs that morning at exactly 8.22 a.m., and so forms the belief that the time is 8.22 a.m. by looking at the accurate clock at the bottom of the stairs. Accordingly, the demon achieves what he wants without having to do anything.[25]

Kelp's claim is that insofar as the demon doesn't intervene then, given how Chris formed his belief, he gains knowledge. But since the demon will ensure that Chris continues to believe that the time is 8.22 a.m. in all

---

[24] The same diagnosis will apply to the structurally similar, though more complex, 'Halloween party' case offered by Comesaña (2005, 397), which is also meant to be a counterexample to the necessity of safety for knowledge. That said, as Kelp (2009) points out, it isn't at all obvious that Comesaña's example even involves an unsafe belief in the first place.
[25] This is essentially a type of 'Frankfurt-style' example in that what is significant is that the demon *would* have intervened rather than that he did intervene. See Frankfurt (1969).

nearby possible worlds, even when this is false, Kelp also claims that this belief is nonetheless unsafe.

While I think this example is ingenious, I don't think it works. In particular, I don't at all share the intuition that the agent in DEMON has knowledge. After all, given how Chris formed his belief it is pure luck that this belief happens to be true – had he come downstairs a minute earlier or a minute later, then he would have formed a false belief. Indeed, Chris is effectively finding out the time by looking at what is (for him anyway) a stopped clock, since whatever time he comes downstairs the clock will say '8.22 a.m.'. But one cannot gain knowledge about the time by consulting a stopped clock, even when one happens to form a true belief!

That said, I do think that there is something epistemically laudable about the agent's true belief, in that (given that the demon didn't in fact intervene) it is a cognitive success that is significantly attributable to his cognitive ability and thus to his cognitive agency. In this sense, it constitutes a cognitive achievement on the part of the subject, even though it isn't knowledge. Often knowledge and cognitive achievement go hand-in-hand (which I think may explain Kelp's inclination to ascribe knowledge to the agent in DEMON), but what I think cases like this illustrate quite neatly is that they can come apart. In particular, they will come apart in cases where the luckiness of the cognitive success is entirely due to some feature of the modal environment which is absent in the actual world.[26]

The third counterexample is due to Ian Church (2010):

VIRUS: Smith is ill and exhibits a unique set of symptoms, S. Given these symptoms, Dr Jones forms the belief that 'Smith has Virus X', which she deduces from the true proposition that 'Virus X is the only known virus to exhibit S.' What is more, Dr Jones does a blood test which verifies that Smith's body contains antibodies for Virus X, further justifying Jones' belief. Based on the evidence, it is extremely feasible that Smith has Virus X. As it happens, however, Smith's symptoms are in fact due to an unknown virus, Virus Y, which exhibits identical symptoms to Virus X; Smith only exhibits antibodies for Virus X due to an idiosyncratic feature of Smith's particular biochemistry which causes his immune system to maintain high levels of antibodies long past a given infection. Nevertheless, Dr Jones' belief turns out to be true divorced from Smith's symptoms or his blood work, because

---

[26] For more on this point, see Pritchard (2009b). I discuss the more general claim that knowledge and cognitive achievement come apart in a number of places. See, for example, Pritchard, Millar and Haddock (2010, ch. 2) and Pritchard (2011).

Smith was infected with Virus X just before meeting with Dr Jones – the infection being so recent that blood work cannot detect it and it is causing no symptoms. (Church 2010, 9)[27]

Notice that this case is different from the other two, in that rather than being a putative case of knowledge where the belief in question is unsafe, it is instead an example of a safe belief which doesn't amount to knowledge. The reason why Church thinks that VIRUS is nonetheless a counterexample to the necessity of safety for knowledge is that he holds that the reason knowledge is lacking in this case is due to the epistemic luck involved. But since it is the job of safety to exclude knowledge-undermining epistemic luck, it follows that this case presents a problem for proponents of safety.

Unfortunately, the case doesn't work. Even if we further add – as Church (2010, 10) himself suggests – that the circumstances of the case are such that Smith is virtually guaranteed to catch Virus X just before seeing Dr Jones, this is at best only a counterexample to a crude form of safety which focuses only on the subject's continued belief *that p* across the relevant possible worlds. Remember the point made above about how a belief in a necessary proposition can be unsafe, even though there is obviously no nearby possible world where the necessary proposition in question is believed falsely. The same applies here. Even if there can be no nearby possible world in which Dr Jones believes that Smith has Virus X and believes falsely – because this proposition is true across all nearby possible worlds – it doesn't follow that the belief is safe, since we also need to consider the other beliefs that Dr Jones forms in nearby possible worlds on the same basis as in the actual world. Once we remember this, it becomes clear that Dr Jones will form false beliefs in nearby possible worlds on the same basis as in the actual world. Consider, for example, the close possible world where all that is different is that Smith doesn't happen to maintain a high level of antibodies in his blood and doesn't clearly exhibit the symptoms for Virus X. In such a world Dr Jones would likely form the false belief that Smith *didn't* have Virus X, or the false belief that Smith had another virus which he didn't in fact have.

Of course, we can always set a case up in which the agent is guaranteed not just to form a true belief in the actual world, but also to form

---

[27] This example is a variation on a case originally proposed by Zagzebski (1994, 71), albeit to illustrate a different point.

a true belief on the same basis across all nearby possible worlds. Perhaps the case offered by Church could be reconstructed to do this. But even so, this need be nothing for the defender of the necessity of safety to knowledge to worry about. For as noted above, such an epistemologist does not claim that safe true belief is sufficient for knowledge, and so it is perfectly compatible with the story they tell that there may be cases of safe true belief which aren't thereby cases of knowledge. Moreover, given that the agent is now guaranteed to have a true belief on the relevant basis across all possible worlds, it seems that what is lacking in such a case is nothing to do with the anti-luck intuition. After all, if one has a true belief in these circumstances, then it is surely not a matter of *luck* that one's belief is true.

Indeed, the point made above about how we shouldn't expect the safety condition to fully capture our intuitions about the role of cognitive ability in the acquisition of knowledge is salient here. For it seems that with the case so construed what has gone awry, epistemically, is not that the agent is forming beliefs such that they could so very easily have been false, but rather that her cognitive success is not appropriately related to her cognitive ability. Put another way, it seems that what is epistemically problematic about such beliefs is not that they fail to satisfy the anti-luck intuition about knowledge, but rather that they fail to satisfy the ability intuition.

The fourth, and final, case that we will look at is adapted from one originally offered by Alvin Goldman (1976, 779):

CAR: Alexander is a young boy who is very good at spotting the particular type of car that his dad drives, which is a Vauxhall Zafira. In the environment that he is in, Alexander would not easily mistake a Vauxhall Zafira for another kind of vehicle. In general, however, Alexander is not very good at identifying cars in his environment, since he tends to classify most vehicles he sees as cars, including lorries, buses, and so forth. Since there are a lot of vehicles in his environment that he would falsely classify as cars, he would very easily falsely classify a vehicle as a car.

Here is the problem posed by this case. Suppose that Alexander sees a Vauxhall Zafira parked outside and so forms the belief that there is a car parked outside. Intuitively, this ought to be knowledge, given how we have described the case. And yet since Alexander has a tendency to misclassify vehicles in his environment as cars, there do seem to be close possible worlds where Alexander forms his belief in

the same perceptual manner as in the actual world and yet believes falsely (e.g., the close possible world where there is a lorry parked outside). What we have here, then, appears to be a case of unsafe perceptual knowledge.

I think that once we spell out the details of CAR, then it ceases to be a case of unsafe knowledge. In order to see this, we first need to notice that had Alexander formed the belief that the object before him was a Vauxhall Zafira, then this belief would have been safe, and hence in the market for knowledge. After all, given how CAR is described there is no close possible world where Alexander forms this belief on the same basis as in the actual world and yet believes falsely.

Now this might be thought to exacerbate the puzzle in play here, since how can it be that Alexander can form a safe belief that the object before him is a Vauxhall Zafira and yet be unable to form a safe belief that the object before him is a car? After all, I take it that we are meant to be supposing that Alexander knows full well that Vauxhall Zafiras are cars. Does this mean that Alexander can know that the object before him is a Vauxhall Zafira while being unable to know that it is car? I think the answer to this question is 'yes'.

In order to see this, imagine that you are the parent of Alexander and that you have a guest visiting who is waiting on a car to come and collect her. Consider now the following two scenarios. In the first scenario, Alexander comes into the room and announces to the guest that there is a car parked outside. Would one allow this assertion to go unqualified? I think not. Rather, one would feel compelled to inform the guest that Alexander tends to classify all kinds of things that are not cars as cars, and hence that the guest should be wary about forming the belief that there is a car outside on the basis of Alexander's testimony. But isn't that just to say that one is regarding Alexander as not having knowledge of this proposition?

If one needs further convincing on this score, imagine that one is not in the room when Alexander tells your guest that there is a car parked outside. Since no one qualifies Alexander's assertion, your guest will now form the belief that there is a car outside. Let's stipulate that this belief is true. Does your guest now have knowledge that there is a car outside? I suggest not. But if Alexander did know that there is a car outside, then your guest ought to be able to come to know this proposition, too, by receiving his testimony. This is thus another reason for thinking that Alexander lacks knowledge of this proposition.

Now consider scenario two. In this scenario Alexander comes into the room and announces that there is a Vauxhall Zafira outside. Would one feel the need to qualify this assertion? I don't see why, and this suggests that we are here treating Alexander as knowing this proposition. Moreover, I think it is pretty clear that those in the room who receive this testimony will come to know that there is a Vauxhall Zafira outside upon hearing this assertion. But that also suggests that Alexander knows what he asserts, since if he did not, then it would be puzzling how in this case one could gain testimonial knowledge from his assertion.

It is entirely possible, then, that Alexander can know that the object before him is a Vauxhall Zafira but not that it is car. More precisely, it is entirely possible that in the environment that he is in Alexander can know, just by looking, that an object is a Vauxhall Zafira, but that he cannot know, just by looking, that it is a car. And this is so even though Alexander knows that Vauxhall Zafiras are a type of car.

Now one might think that this claim, even if true, is irrelevant for our purposes since we are considering a case in which Alexander sees a Vauxhall Zafira and forms the belief that what he is looking at is a car. But what the foregoing illustrates is that there is a crucial ambiguity in this claim. For notice that if Alexander forms the belief that the object before him is a car on the basis that it's a Vauxhall Zafira, and Vauxhall Zafiras are cars, then *that* belief, so formed, *is* safe and it is a case of knowledge, too. Forming one's belief by deducing it from something that you know is, after all, a safe way of forming one's belief. But this claim is entirely compatible with the fact that Alexander cannot come to know that an object is a car just by looking. For that method of belief-formation, as we have noted, is unsafe.

So provided that we make explicit what the basis for the belief is, then Alexander can come to know that there is a car outside in the case where he in part bases this on inference from what he knows, even though he cannot in general come to know that there is a car outside just by looking.

VI CONCLUDING REMARKS

In this chapter I have argued for three main claims. First, that safety offers the best rendering of the anti-luck condition. Second, that safety is merely necessary, and not sufficient (with true belief) for knowledge. That is, that we should prefer modest anti-luck epistemology over robust

anti-luck epistemology. Third, that the main counterexamples offered to the necessity of safety – and thus to modest anti-luck epistemology – do not hit their target. Along the way we have seen the importance of understanding safety correctly and locating this notion within an anti-luck epistemology.