

*Forthcoming in The Oxford Handbook of Skepticism*, (ed.) J. Greco, (Oxford: Oxford University Press)

## SENSITIVITY, SAFETY, AND ANTI-LUCK EPISTEMOLOGY

*Duncan Pritchard*

*University of Stirling*

**ABSTRACT.** This paper surveys attempts in the recent literature to offer a modal condition on knowledge as a way of resolving the problem of scepticism. In particular, safety-based and sensitivity-based theories of knowledge are considered in detail, along with the anti-sceptical prospects of an explicitly anti-luck epistemology.

### 1. BACKGROUND

It is often noted by epistemologists that knowledge is, at root, non-lucky or non-accidental true belief, the point being that what we seek when we try to analyse knowledge is that condition (or set of conditions) which ensures that when our true belief is an instance of knowledge it is not subject in any substantive way to luck. On the face of it, a justification condition of some sort would seem to achieve this end, since the difference between a justified true belief and a mere true belief is, it seems, that genuine justification eliminates the possibility that one's true belief is only true by accident. Famously, however, the cases cited in Edmund Gettier's (1963) seminal article showed that justification, at least as it is usually understood, cannot achieve this end, and thereby opened up the discussion of a whole range of potential new ways of defining knowledge.

One sub-class of proposals in this regard involved offering a *modal* condition (or conditions) on knowledge which captured this anti-luck requirement—that is, a condition which was concerned with the way in which one's target belief is responsive to relevant counterfactual circumstances (we will look at some examples in a moment). As we will see, one key motivation for these proposals came from their apparent ability to deal with the sceptical problem, and it is the applicability, or otherwise, of these modal conditions on knowledge to the sceptical problem that will be our primary concern here. In particular, we will be considering the ability of accounts of knowledge of this sort to deal with the sceptical problem concerning our knowledge of contingent facts about the external world.<sup>1</sup>

## 2. SENSITIVITY-BASED THEORIES

Consider first the idea—prefigured in the work of Fred Dretske (1970), amongst others, but first clearly outlined by Robert Nozick (1981, ch. 3)—that we should understand knowledge primarily in terms of true belief that meets some sort of *sensitivity* condition.<sup>2</sup> Nozick’s insight was to recognise that the anti-luck intuition underlying much of our thinking about knowledge was in effect a modal intuition regarding how our beliefs in genuine cases of knowledge should be responsive to the facts not just in the actual world (i.e., be true), but also in a relevant range of possible worlds.

Here, in essence, is how sensitivity is usually understood:

*Sensitivity*

An agent *S* has a *sensitive* belief in a true contingent proposition  $p$  =<sub>df</sub> in the nearest possible worlds in which  $p$  is not true, *S* no longer believes  $p$ .

Note that the ordering of the possible worlds here is to be understood in the standard way in terms of a ‘similarity’ function. That is, the worlds closest to the actual world will be those worlds most similar to the actual world—i.e., where very little is different—while those worlds furthest from the actual world will be those worlds most dissimilar to the actual world—i.e., where a great deal is different.

To illustrate this principle, consider my belief right now that I am sitting at my desk. Provided that circumstances are normal, then this belief is not only true, but is also sensitive, in the sense that in the nearest possible worlds in which it is not true—those worlds in which I am standing next to my desk, for example—then I will no longer believe that I am seated at my desk (but will believe instead that I am standing).

This principle is able, it seems, to deal with a number of problems facing contemporary theories of knowledge. To begin with, note how it copes with the Gettier-style counterexamples. Consider, for example, the following case.<sup>3</sup> Imagine that our protagonist comes downstairs one morning and sees that according to her normally reliable clock in the hall the time is 8.20am. Moreover, suppose that this belief is true, in that it is indeed 8.20am. Here we have a case in which an agent forms a true belief in a way that ensures that her belief is justified, given any normal understanding of what justification might involve. The twist in the tale, however, is that the clock has, in fact, stopped twenty-four hours earlier, and so it is

just a matter of luck that the agent happens to look at it at the one time in the day in which it is ‘telling’ the right time. Since, intuitively, one cannot gain knowledge of what the time is by looking at a stopped clock, it follows that the agent in this example does not know what the time is, even though she has a justified true belief in this regard.

A common feature of all Gettier-style examples is that they involve an agent who would normally have formed a false belief given the circumstances in which she formed that belief, but which is justified nonetheless, and yet, through luck, happened to form a true belief after all. In the case just considered, for example, the agent would normally have formed a false belief by looking at the stopped clock, but by happening to look at the clock at the moment that she did, she ensured that her belief was true after all. As Linda Zagzebski (1999) has neatly put the point, what we have here is a case of bad epistemic luck (the bad epistemic luck of trying to find out the time by looking at a stopped clock) being cancelled out by good epistemic luck (the good epistemic luck of happening to look at the clock at the one moment at which it was ‘telling’ the right time).

Sensitivity-based theories of knowledge deal with Gettier-style cases by, in essence, eliminating the possibility that a putative knower could be subject to such an opposing mix of good and bad luck. Consider, for example, how sensitivity-based views respond to the stopped-clock case just described. As we noted, the problem with such examples is that one can form a justified true belief about what the time is by looking at a stopped clock, even though one cannot gain knowledge of what the time is in this way. Sensitivity-based views can account for the lack of knowledge here because forming one’s belief about the time by looking at a stopped clock is not a *sensitive* way of forming a true belief, and hence cannot on this view qualify as knowledge. After all, although, as in this case, one might happen to form a true belief in the target proposition in the actual world, in the nearest possible worlds in which the target proposition is false—the worlds in which, for example, the time is presently 8.19am or 8.21am, but everything else is the same—one will continue to form the same belief that one did in the actual world (i.e., that the time is 8.20am), and hence form a false belief as a result. In essence, what the sensitivity condition ensures (or at any rate is meant to ensure)—in contrast to the justification condition—is that one cannot acquire knowledge simply by accidentally gaining a true belief.

Another key motivation for sensitivity-based theories of knowledge is that they are able to explain why we lack knowledge in the so-called ‘lottery’ case. This concerns an agent

who forms her belief that she is in possession of a losing ticket for a fair lottery with long odds by considering the low probability involved of her winning. Suppose further that the draw has been announced and the agent's belief is true in that she is indeed in possession of a losing lottery ticket. Given the nature of the odds involved, one would surely regard this agent as being justified in believing what she does. Nevertheless, even though this belief is justified and true, we also have the strong intuition that it is not a case of knowledge.

Although superficially similar to the Gettier-style cases, in that it concerns a justified true belief which, intuitively, is not a case of knowledge, the problem posed by this example is rather different. This is because unlike Gettier-style cases this is not a situation in which the agent has apparently met the conditions for knowledge through a fortuitous combination of good epistemic luck counteracting bad epistemic luck (and hence has a belief which is only accidentally true). Indeed, the way in which the agent is forming her belief in the lottery example is in fact very reliable in at least one sense of the term, in that it will usually enable the agent to form a true belief. Thus, the scenario is very unlike, say, the stopped clock Gettier-style case, where the agent lacks knowledge because although she happens to form a true belief, this is despite the fact that the way in which the belief is formed is in fact very *unreliable*.

The difficulty posed by the lottery example is thus to explain why, intuitively, the agent lacks knowledge in this case even though she is forming a justified true belief via a highly reliable method. In order to bring this problem into sharp relief, notice that the agent could, again intuitively, gain knowledge of the target proposition by ignoring the odds involved and basing her belief solely on reading the results in a reliable newspaper. Interestingly, however, the probability that she forms a false belief by forming her belief in this way could well be a lot higher than the probability that she forms a false belief by considering only the odds involved in the lottery (note that one can set the odds involved in the lottery as high as one wishes in order to ensure that this is the case). The lottery case thus appears to indicate that knowledge is not—or at least not simply—a function of the strength of one's evidence, where strength here is measured probabilistically. Instead, it seems that evidence of a certain kind can sometimes suffice for knowledge even though, surprisingly, stronger evidence would not suffice.

Sensitivity-based views can accommodate our intuitions in this regard, and in doing so cast some light over why knowledge isn't (simply) a function of the strength of one's

evidence in this way. Take the agent's true belief that she owns a losing lottery ticket on the basis of her evaluation of the odds involved. Although this belief is true in the actual world, it is not sensitive (and hence on this view not a case of knowledge), in that in the nearest possible world in which the target proposition is false—i.e., the world in which she wins the lottery, but everything else consistent with this change stays the same—she will continue to form the same belief and so will as a result form a false belief. Thus, although the evidence that the agent has in support of her true belief is probabilistically very strong, it is not evidence which ensures the sensitivity of her belief, and this is what is essential, on this view, for knowledge.

Contrast this case with that of the agent who forms her true belief that she owns a losing lottery ticket by reading the results in a reliable newspaper. This true belief is sensitive since in the nearest possible worlds in which the target proposition is no longer true—again, where she wins the lottery, but everything else consistent with this change stays the same—she will no longer believe that she has lost but will instead believe that she has won (since this is what the reliable newspaper will tell her). On this view, then, there is no bar to the agent being counted as possessing knowledge of this proposition, just as intuition would predict. So while the evidence the agent has in this case may well be probabilistically weaker than the evidence the agent possesses in the previous case, this evidence does suffice to ensure that the agent has a sensitive belief, and this is what is important, it seems, when it comes to knowledge possession.

### 3. SENSITIVITY AND SCEPTICISM

The main motivation for sensitivity-based theories of knowledge, however, and the motivation that is of most concern to us here, is that they seem to offer a very neat resolution of at least one form of the sceptical problem. We can represent this form of scepticism in terms of two premises and a conclusion. The first premise states that we are unable to know the denials of sceptical hypotheses, such as the sceptical hypothesis that one is presently a brain-in-a-vat ('BIV') that is hooked-up to supercomputers and is being artificially 'fed' its apparently normal experiences. The second premise states that if one is unable to know that, then one does not know any 'everyday' proposition ('E') which is known to be inconsistent

with the relevant sceptical hypothesis, where by an ‘everyday’ proposition is meant the sort of proposition which we would all typically take ourselves to know in normal circumstances, such as (in my case) that one is seated. Finally, with these two premises in hand, the sceptic validly concludes that one does not know any everyday proposition.

Focussing on the BIV hypothesis, one can thus capture the basic form of this sceptical argument as follows:

*The Sceptical Argument*

- (S1) I do not know that I am not a BIV.
- (S2) If I do not know that I am not a BIV, then I do not know E.
- (SC) I do not know E.

Since this argument could be applied to just about any E-type proposition (one would just have to vary the sceptical hypothesis to suit), one is only a few steps away from concluding that just about E-type proposition is unknown. Moreover, since the two premises are held to be in some sense necessary truths (we will consider them further in a moment), and the argument is valid, it also follows that we are just a few steps away from the dramatic sceptical conclusion that knowledge of just about E-type proposition is *impossible* (at least for creatures like us).

Both the premises in this argument are surely intuitive. The first premise is intuitive since sceptical hypotheses are defined such that there is nothing in one’s immediate experience which could indicate that one is not the victim of such a sceptical scenario. No matter what the nature of my present experience, the scenario in which I am a BIV being artificially ‘fed’ these experiences and the scenario in which I am not a BIV and am experiencing my environment normally are, *ex hypothesi*, entirely indistinguishable.

The second premise is not immediately intuitive, but it can be made to be intuitive once one reflects that it rests on the principle that knowledge is closed under known entailments, or the ‘closure’ principle for short. This principle can be roughly stated as follows:

*Closure*

For all  $S$ ,  $\varphi$ ,  $\psi$ , if  $S$  knows  $\varphi$ , and  $S$  knows that  $\varphi$  entails  $\psi$ , then  $S$  also knows  $\psi$ .<sup>4</sup>

In words, closure demands that if one knows one proposition, and one knows that this proposition entails a second proposition, then one knows the second proposition. This principle is certainly intuitive, since it is hard to see how it could possibly fail. For example,

if I know that I'm seated, and I know that if I'm seated then I'm not standing, then surely I also know that I'm not standing. What could be more uncontroversial?

The trouble is, however, that one can feed sceptical hypotheses and everyday propositions into this principle with devastating effect. Suppose I know, for example, that I am seated, and I also know (as surely I do) that if I am seated then I am not a BIV (since BIVs don't *sit* anywhere), then it follows via closure that I know that I am not a BIV. Conversely, however, if I am unable to know that I am not a BIV, then I can't know that I am seated, and hence we get the inference from lack of anti-sceptical knowledge to lack of everyday knowledge at issue in (S2). Moreover, since the closure principle is highly intuitive, it follows that (S2) inherits this intuitiveness once the supporting role of closure is made explicit.<sup>5</sup>

Sensitivity-based theorists are quite happy with the first premise of the sceptical argument. Indeed, they claim that they have a very good explanation of why we are unable to know the denials of sceptical hypotheses, which is that the relevant belief in this regard will inevitably be insensitive. Take one's belief that one is not a BIV, for example. Even if this belief is true, in the nearest possible worlds in which what one believes is no longer true—i.e., the worlds in which one is a BIV, but everything else consistent with this change remains the same—one will by definition continue to believe that one is not a BIV even though this belief is now false. That is, since there is nothing in one's experience which could enable one to distinguish between BIV and non-BIV experiences, it follows that one's belief in this regard is inevitably insensitive and hence, since sensitivity is essential for knowledge, one cannot know these anti-sceptical propositions, just as the sceptic contends.

We've already noted, however, that our everyday beliefs, such as the belief that one is seated, can be entirely sensitive, and so candidates for knowledge. The problem premise in the sceptical argument for the sensitivity-based theorist is thus the closure-based second premise. Indeed, defenders of sensitivity, like Nozick, explicitly contend that closure must be rejected and that in its rejection lies the resolution of the sceptical problem. One can know everyday propositions, such that one is seated, in virtue of possessing a sensitive belief in these propositions, know that they entail the denials of sceptical hypotheses, like the BIV hypothesis, and yet fail to know the denials of sceptical hypotheses, in virtue of lacking a sensitive belief in these propositions.

One can get a sense of why closure fails on this view by noting that sensitivity is a

modal principle which will demand consideration of very different possible worlds depending on which proposition is at issue. Suppose that the actual world is pretty much as we take it to be. On this supposition, as regards one's knowledge of the everyday proposition that one is seated, the relevant possible worlds will be the relatively near-by worlds in which one is no longer seated but where everything else consistent with this change remains the same. In contrast, as regards one's knowledge that one is not a BIV, the relevant possible worlds will be the relatively far-off worlds in which one is BIV, worlds in which a great deal would need to be different from the actual world in order to effect the difference. Since on this view different classes of worlds are relevant to the determination of knowledge, it is hardly surprising that a principle like closure will fail, since there will be cases, such as the sceptic's own inference, in which the possible worlds at issue when one considers whether the agent knows the antecedent proposition will be very different from the possible worlds at issue when one considers whether the agent knows the consequent proposition, meaning that the possession of knowledge in the former case tells us very little about whether or not knowledge is possessed in the latter case, even if the entailment in question is known.

#### 4. PROBLEMS FOR SENSITIVITY-BASED THEORIES

The obvious problem with responding to radical scepticism in this way is that the closure principle is highly intuitive, so intuitive in fact that many feel that rejecting closure is too costly a price to pay for a resolution of the sceptical challenge. After all, can we really make sense of the idea that an agent might know one proposition, know that it entails a second proposition, and yet fails to know the second proposition? As Keith DeRose (1999, 201) has neatly expressed the point, sensitivity-based theorists who deny closure are committed to "the abominable conjunction that while you don't know that you're not a bodiless (and handless!) BIV, still, you know you have hands." But there is clearly something bizarre about endorsing a conjunction of this sort.<sup>6</sup>

Intriguingly, however, it may be that one could retain closure *and* endorse a sensitivity-based theory of knowledge which is able to resolve the sceptical problem outlined above. In order to see why this might be so, we need to consider an essential modification that Nozick makes to the principle of sensitivity which involves relativizing it to a method.

Consider, for example, the following ‘grandmother’ case (cf. Nozick 1981, 179ff.). This involves a grandmother who has a highly reliable ability to tell, by looking at him in good lighting and so forth, that her grandson is well. Crucially, however, in the nearest possible worlds in which her grandson is unwell, he would be kept away from his grandmother and she would be told by his relatives that he is well regardless, so that she wouldn’t worry about him. Accordingly, on the version of sensitivity outlined above, it follows that the grandmother’s belief that her grandson is well is insensitive, and thus not a case of knowledge, even despite the fact that it is formed in a highly reliable fashion, because in the nearest possible worlds in which the proposition believed is false, she continues to believe it regardless.

Explicitly relativizing the principle of sensitivity to a method solves this problem, since what is crucial to this example is that the method by which the belief is formed—i.e., having a good look at her grandson in good environmental conditions—is not the same method by which she forms her belief in the nearest worlds in which the proposition believed is false—which involves the testimony of her relatives. If we keep her method fixed, however, then we are able to retain the intuition that the grandmother has a sensitive belief and thus has knowledge. After all, in the nearest possible worlds in which she gets a good look at her grandson in good environmental conditions and he is unwell, she will recognise that this is the case and no longer believe that he is well.

With the sensitivity principle modified to deal with this problem, however, it is now no longer clear that one gets the counterexamples to closure in the sceptical case that Nozick envisages. After all, the manner in which one forms one’s belief that one is not a BIV in the actual (non-sceptical) world is surely, at least in substantial part, via one’s normal perceptual faculties, and yet one cannot be forming one’s belief that one is not a BIV in this way in the possible worlds in which one *is* a BIV since, *ex hypothesi*, those perceptual faculties are not available to one in these worlds.<sup>7</sup> If this is right, then one’s belief that one is not a BIV can, it seems, be sensitive (and thus, in principle at least, an instance of knowledge), since there cannot be a possible world in which what one believes is false—i.e., where one is a BIV—where one continues to form this belief *via the same method as in the actual world*.

The reason why Nozick does not realise that this option of retaining closure exists for his position is that he understands the notion of a method ‘internalistically’. That is, he insists that any two methods which are ‘experientially the same, the same “from the inside”, will

count as the same method' (1981, 184-5). So construed, if one grants that one's experiences are the same in the actual world in which one is not a BIV and the nearest BIV possible world, then it follows that one's method would be the same in both cases. This is an oddly restrictive way of understanding methods, however, especially if, like Nozick, one is in other respects an epistemic externalist.

As an epistemic externalist, Nozick holds that what is most important to knowledge possession is that one's belief stands in an *external* relationship to the world—that is, a relationship which is not reflectively accessible to the subject—where this is cashed-out in terms of the agent's belief meeting modal conditions like sensitivity (one cannot come to know by reflection alone that one's belief has met the sensitivity condition).

Consider, for example, the famous 'chicken sexer' example. This concerns an agent who has a highly reliable way of distinguishing between male and female chicks, but who is unaware of how reliable she is and, moreover, has false beliefs about how she is making this distinction (she believes that she is touching and seeing something distinctive, when in fact there is nothing distinctive to see or touch and it is her sense of smell that she is employing). Internalists characteristically deny knowledge to this agent on the grounds that she lacks good reflectively accessible grounds in support of her knowledge. In contrast, externalists typically allow that such agents have knowledge on account of how the agent concerned really does have the reliable ability in question. This is certainly the case with sensitivity-based accounts. Consider the chicken sexer's true belief that she is holding two chicks of a different sex. In the nearest possible worlds in which what she believes is no longer true (where she is holding two male chicks rather than a male and female, say), she will no longer believe the target proposition but will instead believe that she is holding chicks of the same sex. Her highly reliable chicken-sexing ability thus suffices to meet the sensitivity condition on knowledge even though she lacks the good reflectively accessible grounds that the internalist demands.

If one is happy to endorse externalism, however—and note that this is not an uncontroversial move to make—then it is not obvious why one should accept that the defining mark of a method should be something that is reflectively accessible—i.e., the nature of one's experiences. Why couldn't one's method be determined by facts which were not reflectively accessible to one as well? Isn't it plausible to suppose that a cognitive psychologist could empirically determine that there are two cognitive processes which have the same experiential upshot? Indeed, think again of the chicken sexer. It is not beyond the

bounds of plausibility to suppose that there need be no experiential difference between the chicken sexer who is exercising her genuine chicken sexing ability and someone who lacks this ability but merely thinks that she has it. On the Nozickean conception of methods, however, sameness of experience would entail sameness of method, and this seems counter to the guiding externalist intuition. The chicken sexer has knowledge on the externalist account because she is, as a matter of fact, exercising a highly reliable cognitive ability, even though she lacks good reflectively accessible grounds in favour of her beliefs so formed. In contrast, someone who merely thinks that they have a highly reliable ability, even if she has the same experiences as the genuine chicken sexer, is not a knower. Nozick's account of methods seems unable to allow him to make this central externalist distinction.<sup>8</sup>

There is also a further issue here, which is that while it is uncontroversial to suppose that one would not be able to distinguish between the experiences that one has when one is envatted and the corresponding normal non-envatted experiences, it *is* controversial to suppose that they are (thereby) the same experiences. Nozick seems to be implicitly supposing that indistinguishability of experiences entails sameness of experiences, but this is not an obvious entailment, especially if one factors in content externalist considerations. On the content externalist view, the content of one's experiences can be dependent upon 'worldly' facts obtaining of which one is entirely unaware, such as facts concerning the relationship between one's experiences and what those experiences are about. On such a picture, it is far from obvious that the content of one's experiences would be exactly the same in the BIV and corresponding non-BIV case, even though it would still be true that one could not discriminate between the two experiences.<sup>9</sup>

A very different problem for the sensitivity-based account of knowledge is raised by counterexamples which appear to show that sensitivity prevents us from having lots of everyday knowledge. Ernest Sosa (1999), for instance, offers the following example. Imagine someone dropping her rubbish down the rubbish chute in her high-rise condo. Does she know, moments later, that her rubbish is in the basement? According to the sensitivity-based account of knowledge, argues Sosa, the answer to this is 'no' because her belief in this regard is insensitive. After all, were the bag to have somehow snagged on the way down the chute, such that what she believes is false, she would still have continued to believe (via the same method she used in the actual world) that the rubbish is now in the basement regardless. The trouble is, claims Sosa, this is a fairly paradigm instance of everyday knowledge, and if

sensitivity-based theories cannot account for cases like this then the view is in serious trouble. That is, sensitivity-based views are meant to at least have the advantage of ensuring that our everyday knowledge is secure, even if they are not also able to ensure that we have anti-sceptical knowledge as well. If Sosa is right, however, then sensitivity-based views cannot accommodate all cases of everyday knowledge, and that would be fatal blow to the view.<sup>10</sup>

## 5. SAFETY-BASED THEORIES

Sosa offers a very different modal principle to replace sensitivity, what he refers to as *safety*. Here, in essence, is how safety is usually understood:

### *Safety*

An agent *S* has a *safe* belief in a true contingent proposition  $p$  =<sub>df</sub> in most near-by possible worlds in which *S* believes  $p$ ,  $p$  is true.

Note that, as with the sensitivity principle, some sort of relativisation to methods will also be required in order to enable safety to deal with potential counterexamples. Furthermore, note that, again like the sensitivity principle, safety aims to capture our intuition that knowledge is non-lucky true belief, the guiding interpretation of that intuition being that one's true beliefs should, if they are to count as knowledge, be safe in the sense that they could not have easily been false.

Safety can account for our putative knowledge in the 'rubbish chute' case because although there is a possible world in which what is believed is false and yet one believes it anyway (via the same method)—i.e., the world in which the rubbish snags on the way down the chute—in most near-by possible worlds one will believe this proposition and one's belief will be true, thereby ensuring that it is safe.

Safety can also be put into service to explain why we have knowledge of the denials of sceptical hypotheses, and thus there is not the same problem regarding closure on this view that sensitivity-based theories have. For example, provided the worlds in which sceptical hypotheses are true are indeed far-off, then one's true belief that one is not a BIV will be *guaranteed* to be safe, since there will be no near-by possible world in which what is believed is false. Indeed, proponents of safety-based theories of knowledge, like Sosa (1999) and

myself (Pritchard 2002*b*; 2005*a*, part one; *forthcominga*), have made the fact that safety can account for our knowledge of the denials of sceptical hypotheses an explicit part of the motivation for the principle.

The anti-sceptical view that results is known as ‘neo-Mooreanism’, in that, like the commonsense approach to scepticism offered by G. E. Moore (e.g., 1925; 1939), it offers a very straightforward response to the sceptic, one that treats us as having knowledge of both everyday propositions and the denials of sceptical hypotheses.

## 6. PROBLEMS FOR SAFETY-BASED THEORIES

There have been two main foci of attack on safety-based views. The first concerns its effectiveness as an anti-sceptical strategy on account of its treatment of our putative knowledge of the denials of sceptical hypotheses. The crux of the issue here is that the role of a modal condition on knowledge is usually thought to capture some counterfactual sense in which our beliefs are responsive. In the case of the sensitivity principle this is clear to see, since it demands that if the fact in question did not obtain, then one would no longer believe that it had obtained. Typically, one also gets a type of responsiveness to the facts by appeal to the safety principle. Take the chicken sexer, for example. There are near-by possible worlds in which what she believes in the actual world is false, and near-by possible worlds in which what she believes in the actual world is true. That she only tends to continue to believe what she does in those worlds where what she believes continues to be true indicates the counterfactual responsiveness of her beliefs.

This sort of picture of knowledge as belief which is suitably counterfactually responsive does not, however, seem to work when it comes to one’s putative knowledge of the denials of sceptical hypotheses on the safety-based view. Instead, it seems that one only has such knowledge according to the safety-based theorist in virtue of how one has a counterfactually *stubborn* belief in this regard and there is, as it happens, no wide class of near-by possible worlds in which the proposition believed is false. After all, if there were a wide class of near-by possible worlds in which the relevant sceptical hypothesis obtained, then it would follow that there was a wide class of near-by possible worlds in which one believed the target proposition and yet the belief was false, thereby ensuing that one’s belief

is unsafe and hence, on this view, that one lacked knowledge. Put simply, whereas in most cases where one has knowledge on the safety-based view—such as chicken sexer cases—there is a wide class of near-by possible worlds in which what one believes is false and one does not continue to believe the target proposition, this is not so for one’s putative safety-based knowledge of the denials of sceptical hypotheses. This makes the knowledge look rather lucky, in that it does not appear to be arising out of any cognitive ability that one has but merely reflecting a fortunate match across near-by possible worlds between what one believes and the relevant counterfactual circumstances.<sup>11</sup>

We will return to look at this problem again in the next section. First, however, I want to consider the second main prong of attack against safety-based views—which has been pressed by John Greco (2002; *forthcoming*)—to the effect that there is no stable construal of the principle which can accommodate both everyday cases of knowledge and the kind of knowledge that one has in the lottery case.

Recall that the sensitivity principle can deal with the lottery case since the difference between forming your belief that you have lost a lottery simply by reflecting on the odds involved rather than looking up the result in a reliable newspaper is that forming your belief via the former method would lead to an insensitive belief, whereas forming your belief via the latter method would produce a sensitive belief. The problem for safety-based theories of knowledge is that, seemingly, they cannot account for our lack of knowledge in the lottery case. This is because given the probabilities involved there are very few near-by possible worlds in which one’s belief that one has lost is false, and thus in most near-by possible worlds in which one believes that one has lost (on the same basis as in the actual world) one’s belief will be true. Thus, the belief will be safe and hence, on this score at least, a potential case of knowledge, contrary to intuition.

One way around this problem could be to insist on a further condition on knowledge over and above safety, but one might worry that such an amendment to the view simply to deal with this case would be *ad hoc*. Alternatively, one could deal with the problem by demanding that one should construe safety so that it requires that one’s belief matches the truth not just in most near-by possible worlds, but in all of them, a position which I have argued for (Pritchard 2005a, ch. 6). This would ensure that one’s belief that one has lost the lottery is no longer safe, and thus not a case of knowledge, but the problem then is to account for the knowledge that is apparently present in the rubbish chute case. After all, surely there

are *some* near-by possible worlds in which the rubbish chute snags on the way down and yet one continues to believe (on the same basis) that the rubbish is in the basement regardless? The challenge is thus to formulate safety robustly enough to deal with the lottery case, while also allowing the formulation to be liberal enough to count agents as having knowledge in cases like the rubbish chute example.

One could respond to this challenge by claiming that, contrary to first appearances on this score, the agent in the rubbish chute example *does* have a belief which, if it is to count as knowledge at any rate, matches the truth across all nearby possible worlds. After all, suppose for a moment that there is a class of near-by possible worlds in which the bag would snag and yet the agent would continue to believe that it was in the basement (via the same method) regardless, perhaps because there is an imperfection in the lift shaft that the bag is almost snagging on each time it drops. If this is how we are to understand the example, then is it really so plausible to suppose that the agent knows that her rubbish bag is in the basement? Indeed, intuitively, one would only suppose the agent to have knowledge in this case provided there really is no near-by possibility in which the bag would undetectably snag. Implicitly, we assume that the shaft is smooth and that there is nothing for the bag to snag on.

Of course, there are possible worlds that are not too far off in which the rubbish bag does snag, such as worlds in which the shaft gets damaged in such a way that there is now an imperfection that the bag can snag on (but which is such that the imperfection wouldn't cause the agent to alter her belief accordingly, as would happen, for example, if the imperfection in the shaft was manifest to the agent, or was indicated by a sign next to the shaft warning people of problems with the chute that might prevent the chute working as it should). The point, however, is that there is no obvious reason why we should regard such worlds as close-by worlds, and hence the more robust conception of safety as demanding a belief that matches the truth in all near-by worlds can stand.

## 7. ANTI-LUCK EPISTEMOLOGY

One way in which one can motivate the safety-based account—and along the way perhaps offer a more compelling defence of safety-based views against the objection we have just considered—is to explicitly locate the position within an anti-luck epistemology. We have

already noted that one could regard the sensitivity and safety principles as capturing our intuition that knowledge is non-lucky true belief. The idea behind anti-luck epistemology, however, is that one should be more explicit about what this connection involves. In particular, the suggestion—which I have explored at length in Pritchard (2005*a*)—is that we should offer an analysis of luck and of the sense in which knowledge is non-lucky, and then, on this basis, explicitly offer an anti-luck epistemology. What is salient about this sort of project for our purposes is that it seems to directly lend support for safety-based theories of knowledge.

Consider a paradigm case of a lucky event, such as a lottery win. What is it about such an event that makes it lucky? Here, roughly, is one straightforward answer to that question. A lucky event is an event which is of some significance to the agent (or at least in some sense ought to be) which obtains in the actual world but which does not obtain in a wide class of near-by possible worlds where the relevant initial conditions for that event are the same as in the actual world. For example, the thing about winning a lottery (an event that is clearly of significance to one) is that it is the sort of event that wouldn't normally occur, since normally one would lose. Thus there is a wide class of near-by possible worlds in which the relevant initial conditions for the event are the same as in the actual world—in which one continues to purchase a lottery ticket for example, and the lottery continues to be fair with long odds—where one is holding the losing lottery ticket right now.

Contrast this case with that of a paradigm example of a non-lucky event, such as when a skilled archer hits the target. Here we again have an event that is of significance to the agent, but this time this is an event which not only obtains in the actual world, but also in most (if not all) near-by possible worlds in which the relevant initial conditions for that event remain the same (where, for example, the environmental conditions are relevantly similar). It thus isn't lucky by the lights of this specification of luck, just as intuition would predict.

A great deal more needs to be said about this account in order to flesh it out and provide it with the requisite support of course, but this crude outline should suffice for our purposes here. This is because with this rough account of luck in mind we can get a sense of why safety might be the right principle to adopt if one is explicitly seeking to offer anti-luck epistemology.

Note that what we want from an anti-luck epistemology is an account which ensures that, given one has met all the relevant epistemic conditions, one's belief, when true, must be

such that it cannot be a matter of luck that it is true. In other words, the epistemic luck that we are most troubled by is luck in the truth of our beliefs. An inability to eliminate this sort of luck was, recall, what was shown to be amiss by the Gettier-style counterexamples to the traditional tripartite account of knowledge in terms of justified true belief, in that one could have a justified true belief and yet it still be a matter of luck that one's belief was true, as in the stopped clock example described above.

Putting this point together with our rough account of luck indicates that what we want from an anti-luck epistemology is a condition on knowledge which ensures that the event of our forming a true belief (an event which we will take for granted is of significance to the agent, in order to simplify matters) is non-lucky. That is, that it is not the case that although the relevant event (one's forming a true belief in the target proposition) obtains in the actual world, there is a wide class of near-by possible worlds in which the initial conditions for that event remain fixed and yet the event doesn't obtain. Turning this around, and taking it as given that the 'relevant initial conditions' will tend in this case to be captured by the mechanism through which the belief actually arose, we can recast this anti-luck claim as follows: what we seek is a true belief which is such that in most near-by possible worlds in which the mechanism through which the belief actually arose remains the same, the event of believing truly continues to obtain. More simply, what we seek is a true belief which is such that in most near-by possible worlds in which we continue to believe the target proposition in the same way as in the actual world, one's belief continues to be true. But this, of course, is essentially just safety.

Reflection on the anti-luck intuition thus seems to directly motivate a safety-based epistemology, and given the prevalence of this intuition in contemporary epistemology, this is strong support indeed for safety-based views. Returning to the sceptical problem that is our primary focus here for a moment, one can see how safety-based theorists might go about responding to the charge, outlined in the last section, that their view offers a very strange picture of our anti-sceptical knowledge, such that we seem to be able to know the denials of sceptical hypotheses simply in virtue of a stubborn belief in the target proposition and epistemically friendly near-by possible worlds. The answer to this objection will be to point out that if what is most important to knowledge is possession is a true belief which is not luckily true, then our anti-sceptical beliefs, even granted this description of them, can perfectly well meet this rubric. The thought will run that while there may well be something

epistemically lacking about our anti-sceptical beliefs, such an epistemic lack is not decisive when it comes to the issue of knowledge possession, since in this regard mere non-lucky true belief—i.e., safe belief—will suffice.

Setting safety-based views within an anti-luck epistemology might also enable the proponent of the view to deal with the other key problem we raised above, which was the objection that there was no stable account of the safety principle available which could accommodate both the intuition that we have lots of everyday knowledge (such as the knowledge at issue in the rubbish chute example), and the intuition that we also have knowledge in the lottery case. We dealt with that objection above by treating the safety principle as applying to all near-by possible worlds, and this move could be backed-up by anti-luck considerations. After all, given the honorific status of knowledge, it would not be surprising that when it came to knowledge possession we wanted a state which ensured that there were no near-by possible worlds in which one continues to form one's belief in the target proposition on the same basis and yet the belief is false.

Perhaps a better way of responding to this difficulty, however, is to note that in our ascriptions of luck we often put more weight on relevant counterfactual events depending on how modally close they are. For example, consider the following two cases. First, a case in which one is nearly hit by a bullet from a sniper, where the bullet whistles past one's ear. Second, an exactly similar case in which one is nearly hit by a bullet from a sniper, with the one difference that this time the bullet does not come quite as close as in the first case, but in fact hits a tree a few feet away from one. We would clearly judge one's not being hit by the bullet in question to be lucky in both cases, since in each case there is a wide class of near-by possible worlds in which the relevant initial conditions for the event are the same (the environmental conditions, for example) and the bullet does hit (moreover, the event of being hit by a bullet is clearly of significance to one). Nevertheless, I think it is also clear that we would regard the former event as being luckier than the latter event, on account of how the possible world in which one gets hit is closer in the former case than in the latter case.

This has a bearing on how one might modify safety in order deal with Greco's objection. After all, what is salient about the lottery case is that given the nature of how lotteries are decided, the world in which one is presently holding the winning lottery ticket is in fact very close to the actual world in which one is holding the losing ticket. This is in contrast to the rubbish chute example, since while one might disagree about whether,

properly understood, the world in which the bag snags is close enough to count amongst the near-by worlds, it is certainly true by everyone's lights that it won't be particularly close, since if that were the case then the agent concerned certainly wouldn't possess knowledge. With this distinction between the two cases in mind, one can thus deal with the problem in hand by sticking to the original formulation of safety in terms of most near-by possible worlds but modifying one's understanding of this principle so that more weight is given to the closest worlds. One could do this, for example, by insisting that there be no very close near-by possible world in which the agent believes the target proposition (on the same basis) and yet forms a false belief (thereby dealing with the lottery case), while also allowing that of the near-by possible worlds as a whole, it need only be the case that one's belief matches the truth in most of them (thereby ensuring knowledge in the rubbish chute case).<sup>12</sup>

Allying safety-based views to an anti-luck epistemology thus highlights some interesting ways in which the view can be defended against some of its most pressing objections.

## 8. CONCLUDING REMARKS

I want to close by considering two issues of central importance to the sort of views outlined here. The first concerns how this debate connects with one of the dominant research programmes of contemporary epistemology, that of "relevant alternatives" approaches to knowledge. The driving idea behind these proposals is the thought that in acquiring knowledge one does not need to be infallible in the sense of being able to rule-out all error-possibilities associated with the proposition known, since that would certainly be far too restrictive and would amount to a clear invitation to radical scepticism. Instead, it is only a sub-class of all the associated error-possibilities that is relevant, and the task in hand is to offer some plausible and principled specification of what these relevant error-possibilities are.

One can regard sensitivity-based views as being within this general programme, especially insofar as the view is understood as resulting in the denial of the closure principle. Taking 'rule-out' here to be broadly equivalent to 'know to be false', the thought will be that it is an infallibilist and thus non-relevant alternatives line of thought which leads us to

endorse the sceptic's use of closure and hence require of everyday knowledge that we are also able to have knowledge of the denials of sceptical hypotheses. On this view, then, relevant alternatives and the rejection of closure go hand-in-hand.

There is reason to be suspicious of this reading of the relevant alternatives intuition, however. After all, the guiding thought behind this intuition is surely that those error-possibilities that are indeed far-fetched are thereby *irrelevant* to the determination of knowledge, and so can be disregarded with impunity. What is odd about the sensitivity-based reading of the relevant alternatives intuition, however—or at least that reading which results in the denial of closure—is that it does allow far-fetched error-possibilities to be sometimes relevant to the determination of knowledge. In deciding whether one knows that one is not a BIV, for example, the relevant possible worlds will be those *ex hypothesi* far-off worlds in which one is a BIV. But if the possible world in which one is a BIV is indeed far-off, then why is it relevant to the determination of any of one's knowledge, even one's knowledge that one is *not* a BIV?

Safety-based views, in contrast, stick to the core rendering of the relevant alternatives intuition that far-off error-possibilities—and thus far-off possible worlds—are always irrelevant to the determination of knowledge. Interestingly, as we saw above, if one opts for an 'externalist' conception of methods, then it seems that one could retain a sensitivity-based conception of knowledge and hold that far-off sceptical possible worlds, like the BIV world, are always irrelevant to the determination of knowledge. If this is right, then it opens up the possibility that one might be able to stick to this core rendering of the relevant alternatives intuition while also offering a sensitivity-based account. This raises the intriguing question of how the safety principle and the sensitivity principle on this construal are related, a topic which has been insufficiently explored. Could there be a rendering of these principles on which they are extensionally equivalent? If so, then this would be a fascinating development.

The second issue that I want to raise before closing is whether the sceptical problem as we are understanding it really is the most pressing version of this age-old philosophical conundrum. For if it isn't—if there is a more problematic version of the puzzle available—then the anti-sceptical appeal of the proposals sketched here will be rather slim. I suggest that this is indeed the case, in that the core sceptical problem is not a problem that turns on closure, and thus on whether or not we are able in some sense to know the denials of sceptical hypotheses, at all. Instead, I think there are good reasons to suppose that what really

underpins the sceptical argument is an evidential claim to the effect that since our evidence in support of our everyday beliefs does not favour those beliefs over sceptical alternatives, hence it follows that our evidence is insufficient for knowledge. Rescuing knowledge from the sceptic's grasp in the way that sensitivity- and safety-based theorists do—without in the process explaining what the evidential basis of this knowledge is—is thus by-the-by. This, however, is a topic for another occasion.<sup>13,14</sup>

## REFERENCES

- Black, T. (2002). 'A Moorean Response to Brain-in-a-Vat Scepticism', *Australasian Journal of Philosophy* 80, 148-63.
- Dretske, F. (1970). 'Epistemic Operators', *Journal of Philosophy* 67, 1007-23.
- (2005a). 'The Case Against Closure', *Contemporary Debates in Epistemology*, (ed.) M. Steup, 13-26, Oxford, Blackwell.
- (2005b). 'Reply to Hawthorne', *Contemporary Debates in Epistemology*, (ed.) M. Steup, 43-6, Oxford, Blackwell.
- DeRose, K. (1999). 'Solving the Sceptical Problem', *Skepticism: A Contemporary Reader*, (eds.) K. DeRose & T. A. Warfield, Oxford, Oxford University Press.
- Gettier, E. (1963). 'Is Justified True Belief Knowledge?', *Analysis* 23, 121-3.
- Greco, J. (2002). 'Virtue and Luck, Epistemic and Otherwise', *Metaphilosophy* 34, 353-66.
- (forthcoming). 'Problem for Pritchard's Safety', *Synthese*.
- Hawthorne, J. (2005). 'The Case For Closure', *Contemporary Debates in Epistemology*, (ed.) M. Steup, 26-43, Oxford, Blackwell.
- Kornblith, H. (ed.) (2001). *Epistemology: Internalism and Externalism*, Oxford, Blackwell.
- Luper-Foy, S. (ed.) (1987). *The Possibility of Knowledge: Nozick and his Critics*, Rowman & Littlefield, Totowa, New Jersey.
- McDowell, J. (1995). 'Knowledge and the Internal', *Philosophy and Phenomenological Research* 55, 877-93.
- Moore, G. E. (1925). 'A Defence of Common Sense', *Contemporary British Philosophy* (2<sup>nd</sup> series), (ed.) J. H. Muirhead, London, Allen and Unwin.
- (1939). 'Proof of an External World', *Proceedings of the British Academy* 25, 273-300.
- Nozick, R. (1981). *Philosophical Explanations*, Oxford University Press, Oxford.
- Pritchard, D. H. (2002a). 'Recent Work on Radical Skepticism', *American Philosophical Quarterly* 39, 215-57.
- (2002b). 'Resurrecting the Moorean Response to Scepticism', *International Journal of Philosophical Studies* 10, 283-307.
- (2005a). *Epistemic Luck*, Oxford, Oxford University Press.
- (2005b). 'Scepticism, Epistemic Luck and Epistemic Angst', *Australasian Journal of Philosophy* 83, 185-206.
- (2005c). 'The Structure of Sceptical Arguments', *The Philosophical Quarterly* 55, 37-52.

- (2006). ‘Knowledge, Luck, and Lotteries’, *New Waves in Epistemology*, (eds.) D. H. Pritchard & V. Hendricks, Aldershot, Ashgate.
- (*Forthcoming*). ‘How to be a Neo-Moorean’, *Internalism and Externalism in Semantics and Epistemology*, (ed.) S. Goldberg, Oxford, Oxford University Press.
- Russell, B. (1948). *Human Knowledge: Its Scope and its Limits*, London, George Allen & Unwin.
- Sosa, E. (1999). ‘How to Defeat Opposition to Moore’, *Philosophical Perspectives* 13, 141-54.
- Williams, M. (1991). *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism*, Oxford: Blackwell.
- Zagzebski, L. (1999). ‘What is Knowledge?’, *Epistemology*, (eds.) J. Greco & E. Sosa, Oxford, Blackwell.

## NOTES

<sup>1</sup> Since it is knowledge of contingent propositions which is our focus here, we will set to one side the vexed issue of how modal accounts of knowledge can deal with our knowledge of necessary truths.

<sup>2</sup> Nozick actually imposed *two* modal conditions on knowledge. However, since it is only the sensitivity condition that is relevant to the problem of scepticism, for the sake of simplicity we will focus on sensitivity here.

<sup>3</sup> This example is adapted from one discussed by Russell (1948, 170-1), although it was not explicitly advanced as a Gettier-style case.

<sup>4</sup> One might want to modify this principle in a number of ways in order to deal with potential counterexamples of a trivial sort (such as possible cases where the agent doesn’t even believe the entailed proposition), but this unembellished version of closure should suffice for our purposes here.

<sup>5</sup> For more on the contemporary debate regarding scepticism, and on treatments of this type of sceptical argument in particular, see Pritchard (2002*a*).

<sup>6</sup> For a recent exchange regarding the merits, or otherwise, of denying closure, see Dretske (2005*a*; 2005*b*) and Hawthorne (2005).

<sup>7</sup> Williams (1991, ch. 8) was, as far as I am aware, the first to make this point. See also Black (2002) for a recent discussion of this issue.

<sup>8</sup> For more on the epistemic externalism/internalism distinction in general, see Kornblith (2001).

<sup>9</sup> One type of content externalism that is particularly relevant here is disjunctivism, which holds that the content of one’s experiences in cases of veridical perception is different to the content of one’s experiences in counterpart cases in which one is deceived, even though one would be unable to distinguish between the two sets of experiences. See, for example, McDowell (1995).

<sup>10</sup> For more on sensitivity-based theories of knowledge—and Nozick’s view in particular—see Luper-Foy (1987).

<sup>11</sup> For more on this point, see Pritchard (2005*b*).

<sup>12</sup> For more on this point, see Pritchard (2006).

<sup>13</sup> I argue that the source of scepticism lies in an evidential principle known as ‘underdetermination’ rather than in closure in Pritchard (2005*a*, ch. 4; 2005*c*).

<sup>14</sup> I am grateful to the editor of this volume, John Greco, for his comments on an earlier version of this paper, and also for general discussion, over several years, of the topics related to this article.