Conclusive Reasons and Epistemic Luck

Tamar Lando

To cite this article: Tamar Lando (2016) Conclusive Reasons and Epistemic Luck, Australasian Journal of Philosophy, 94:2, 378-395, DOI: 10.1080/00048402.2015.1058830

To link to this article: https://doi.org/10.1080/00048402.2015.1058830

Published online: 25 Aug 2015.
What is it to have conclusive reasons to believe a proposition P? According to a view famously defended by Dretske, a reason R is conclusive for P just in case \( R \) would not be the case unless \( P \) were the case. I argue that, while knowing that \( P \) is plausibly related to having conclusive reasons to believe that \( P \), having such reasons cannot be understood in terms of the truth of this counterfactual condition. Simple examples show that it is possible to believe \( P \) on the basis of reasons that satisfy the counterfactual, and still get things right about \( P \) only as a matter of luck. Seeing where this account of conclusive reasons goes wrong points to an important distinction between having conclusive reasons and relying on reasons that are in point of fact conclusive. It also has wider consequences for whether modal principles like sensitivity and safety can rule out the pernicious kind of epistemic luck, or the kind of luck that interferes with knowledge.

**Keywords:** reasons, epistemic luck, propositional knowledge, safety, sensitivity

1. **Introduction**

Here is one way in which knowledge differs from other epistemic attitudes. In knowing something, *one could not be wrong about it*. When a person knows—as opposed to merely believes, or justifiably believes—that \( P \) is the case, her epistemic credentials are ‘such as to eliminate the possibility of mistake’, as Dretske once said [1971: 1]. It seems right to conclude that, if a person knows that \( P \) on the basis of reasons, her reasons must be conclusive. But what is it to have conclusive reason to believe that \( P \)?

There is a well-known answer to this question. Suppose that S believes \( P \) on the basis of \( R \). Then \( R \) is conclusive for \( P \) if the following counterfactual holds:

\[
(CR) \quad R \text{ would not be the case unless } P \text{ were the case.}
\]

If \( R \) and \( P \) satisfy \( (CR) \), then \( R \) guarantees the truth of \( P \). To have conclusive reason for \( P \), we might say, is just to believe that \( P \) on the basis of reasons that satisfy \( (CR) \). Moreover, when one believes that \( P \) on the basis of such reasons, one thereby knows that \( P \). Even if this cannot be a fully general account of knowledge—even if there may be other ways of knowing that have nothing to do with reasons—we may still think that, where knowledge *does* depend on reasons, \( (CR) \) specifies the crucial condition that must be met.

In what follows, I’ll argue that this counterfactual account of reasons-based knowledge cannot succeed. Simple examples show that it is possible to believe \( P \) on the basis of reasons that satisfy \( (CR) \), and still get things right
about P only as a matter of luck. Given that one does not know what one gets right as a matter of luck, the agent in these examples does not know that P is the case. If that is right, then (CR) is not sufficient for reasons-based knowledge. So, while knowledge may be a matter of having conclusive reason to believe, we cannot understand what it is to have conclusive reasons in terms of the obtaining of that counterfactual constraint. Seeing where things go wrong points the way to an important distinction between having conclusive reason to believe a proposition P, and relying on reasons that are in point of fact conclusive for P.

Talk of conclusive reasons has wider consequences. If the modal condition (CR) fails to rule out cases of the pernicious kind of epistemic luck—the kind of luck that interferes with knowledge—what about other modal conditions advanced in the theory of knowledge more generally? Here I have in mind, in particular, both ‘safety’ and ‘sensitivity’ principles. I argue that the very same examples adduced against (CR) also show that safety and sensitivity do not rule out the kind of luck that interferes with knowledge. By focusing on conclusive reasons and (CR), we get a handle on a broader debate: What exactly is the place of purely external modal constraints in a theory of knowledge, and can any such constraint rule out the pernicious kind of epistemic luck?

2. A Counterexample to the Counterfactual Account

Let us begin with Dretske, who famously advanced a theory of conclusive reasons that relied heavily on the counterfactual condition (CR).

According to Dretske [1971: 12–13], a subject S has conclusive reason R for P just in case

(A) R would not be the case unless P were the case;

(B) S believes, without doubt, reservation, or question, that P is the case and he believes this on the basis of R; and

(C) S knows that R is the case or R is some experiential\(^1\) state of S (about which it may not make sense to suppose that S knows that R is the case; at least it no longer makes much sense to ask how he knows).

To have conclusive reason for believing P is, according to Dretske, to believe P on the basis of reasons that would not obtain unless the proposition believed were true.

I think that it is not difficult to see, against Dretske, that conditions (A)—(C) are not jointly sufficient for reasons-based knowledge. Consider the following example:

\(^1\)In the original version of Dretske’s article [1971], the word ‘experimental’ was used instead of ‘experiential’. This was apparently a typographical error, as it was replaced with the word ‘experiential’ in the reprinting of the article that appears in [Dretske 2000].
Jessica is trying her hand at poker. She has just recently been taught the rules of the game and, despite her immense confidence, is not a very adept player. She takes note of the king of clubs lying open-face on the table, and when her neighbour, John, decides to raise she concludes that John must have another king in his hand, completing a pair. We can imagine that Jessica is either not aware of, or simply does not consider, the other possibilities in the game. As it happens, and unbeknownst to Jessica, John is a rather strange player himself, and plays according to the following strategy (henceforth, ‘Strategy’): Raise whenever there is a king on the table and you have a king, and call in all other cases. Indeed, John always plays by Strategy—he is deeply committed to playing by Strategy.

John’s strategy makes the following claim true: John would not have raised unless he had a king. Moreover, Jessica believes, without a trace of doubt, that John has a king, and she believes this on the basis of the fact that John just raised. Finally, Jessica knows that John raised, because she is sitting right there at the table with him. Jessica satisfies conditions (A)—(C). So, according to Dretske’s account of conclusive reasons, she has conclusive reason to believe that John has a king (and thus also knows that John has a king).

But this is clearly the wrong verdict. A player in similar circumstances to John might raise for any number of reasons. For example, the king on the table might finish his straight, he might have another pair in his hand unrelated to the king on the table, he may merely be bluffing, and so on. Of course John would not have raised under any of those circumstances, but that is only because John is playing by a very peculiar strategy. The fact that he is playing by that strategy is itself something Jessica is entirely ignorant of. In the example, Jessica gets things right about what cards John has, but she does so only as a matter of luck. So, it seems wrong to say that, given what she knows, Jessica has conclusive reason to believe that John has a king (and thus also knows that John has a king).

Dretske himself, recognizing, I think, some of the unintuitive consequences of (A)—(C), tried gamely to defend his account against them. He asks us to suppose that S, for ‘a perfectly silly reason’, or by ‘sheer accident’, comes to the true belief that a reason R is conclusive for P, and comes to believe that P on this basis. Isn’t it obvious, he asks rhetorically [ibid.: 18], that S does not know that P is the case?:

In evaluating the truth of ‘John would not have raised unless he had a king,’ we hold fixed the fact that John plays by Strategy. (It follows that one cannot object to this claim by saying that John might still have raised, if he were not playing by Strategy.) Indeed, Dretske thought that, in evaluating the truth of a claim of the form ‘Not-R unless P’, we hold fixed all circumstances logically and causally independent of P. The fact that John plays by Strategy is logically and causally independent of his having a king. So, that fact, by Dretske’s own lights, is among the circumstances held fixed in evaluating the ‘unless’ claim. Moreover, the decision to play by Strategy is taken before John is dealt any cards. So, by Lewis’s ban on backtracking, in evaluating what would have happened if John had not had a king (a counterfactual), we hold fixed the decision to play by Strategy. Finally, we stipulate in the example that John always plays by Strategy, that he is committed to playing by Strategy, and so on. These facts are meant to ground the intuition that the possible worlds in which John does not play by Strategy are remote ones (or, at any rate, are not among the nearest possible worlds in which John does not have a king).
No, it is not obvious that he does not know that P is the case. I believe that this objection trades on the very confusion we have just discussed; that is, it mistakenly supposes that if S does not know that R is conclusive for P (has no legitimate basis for believing this), then S does not know that P is the case (has no legitimate basis for believing this).

But here it seems, rather, that it is Dretske who confuses things. I agree with him that one need not know *that* one’s reason is conclusive for P in order to have conclusive reason to believe that P. In fact, I would go further: one need not have *any* second-order belief about the strength of one’s reasons at all. But we did not infer that the agent doesn’t know that P from the fact that she doesn’t know that her reason is conclusive. Rather, the ‘silliness’ of her reason caused us to doubt whether R can appropriately be said to be a conclusive reason *for her* in the first place—given what (else) she knows.

Imagine in connection with this what you would say if, for example, you were seated at the poker table with Jessica, and she leaned over and whispered to you that John had a king. ‘Why do you think that?’, you ask. ‘Didn’t you see him just raise?’, Jessica responds. After some back and forth, it becomes clear to you that Jessica has no further evidence to support her strange belief. When, at the end of the game, John reveals that he has been playing by Strategy, you look over at Jessica to see if she knew about this; but she seems just as perplexed as you are. Do you, at this point, think to yourself that Jessica knew that John had a king all along?

To react in this way would, I think, be very generous—indeed, far too generous. Jessica didn’t know at the time: there simply was no way for her to know, given that she was not aware of Strategy. What leads us to deny this is not that Jessica doesn’t know *that* her reason is conclusive, but rather that we don’t think that she has conclusive reason to believe what she does in the first place. She got things right, sure—but only as a matter of luck. In the example, Jessica’s ignorance of the game just so happens to coincide in the right way with John’s idiosyncratic Poker strategy, producing a true belief. ‘House of lunatics’, you mutter to yourself, as you walk away from the game.

### 3. Appropriately Specific Reasons

In the final pages of ‘Conclusive Reasons,’ Dretske himself discusses some problem cases for the claim that (A)–(C) are jointly sufficient for knowledge, and he suggests a way of dealing with them. The cases bear some resemblance to the counterexample just offered, so it’s instructive to look at what Dretske says [ibid.: 20; the case’s name is added by me]:

*Chemistry.* S, upon inspecting an immersed chemical indicator, declares that the solution in which it is immersed is a base. He believes (correctly) that the indicator is the sort which turns from yellow to blue only when immersed in a base. The indicator is Thymol Blue and would not have turned from yellow to
blue (in these conditions) unless the solution were a base... Does S know that
the solution is a base? Before answering this question the reader should be
informed that there is another chemical indicator, Bromophenal Blue, which
also turns from yellow to blue but only when immersed in an acid. S, however,
is quite unaware of the existence of other such indicators. He merely assumes
that a yellow indicator turning blue is a positive test for a base.

S’s ignorance on this last point does not alter the fact that S satisfies the fol-
lowing conditions, instantiating Dretske’s (A)–(C) in section 2:

(A) The indicator would not have changed colour unless the solution was
a base.

(B) S believes that the solution is a base and he believes this on the basis
of the indicator’s change in colour.

(C) S knows that the indicator changed from yellow to blue.

And yet, Dretske says, ‘despite the satisfaction of these conditions, I find it
(in some cases) most implausible to say [S] knows that the solution is a base’
[ibid.: 21].

What goes wrong, according to Dretske, is that, although S’s reason for
believing P is a conclusive one, S is not exploiting those aspects of his reason
in virtue of which it is conclusive. The fact that this indicator is Thymol blue
(and not Bromophenol) is what makes it the case that the indicator’s change
in colour is conclusive for P; but S’s reason for believing P is simply that this
indicator changed colour, not that a Thymol Blue indicator changed colour.
In general, Dretske notes, the fact that an A is a B is, in some cases, conclu-
sive reason for believing P only in virtue of the fact ‘that it is in particular an
A which is B’. In such cases, Dretske says, we should require that the agent’s
reason for belief be not simply that ‘this (something or other) is a B’, but
that the something or other that is a B is also an A. The needed qualification
on (A)–(C), Dretske thinks, can be put in terms of a further restriction on
our reasons. ‘When one has conclusive reasons, then this is sufficient for
knowing that P is the case when those reasons are properly specific, both
with regard to what it is that displays the particular features on which one
relies and on the particular features themselves’ [ibid.: 12]. S need not,
according to Dretske, know that his reason is conclusive for P, but he must
be exploiting those features of his reason in virtue of which it is conclusive.

I think it is easy to see, however, that even when an agent’s reasons for
belief are properly specific in the sense Dretske has in mind, conditions (A)–
(C) are not jointly sufficient for reasons-based knowledge. While S may be
exploiting the correct features of his reason—those features in virtue of which R
is conclusive for P—he may still get lucky in fixing on just those fea-
tures in the first place. Consider a modified version of Dretske’s example:

Chemistry (Modified). S is a chemist, inspecting a chemical solution in which
she has immersed a certain chemical indicator. S believes (correctly) that the
indicator is the sort that turns from yellow to blue when immersed in C1, but not in C2, and so, upon seeing the indicator turn to blue, S comes to believe that the solution is C1. We can add to the story that there is another sort of indicator, which turns from yellow to blue when immersed in C2 but not C1. S is aware of this fact, and comes to believe what she does not just because a yellow indicator turned to blue, but because an indicator of the first type turned to blue. In fact, the solution has been chosen (by someone other than S) from among three different chemical substances, C1, C2, and C3. S knows nothing at all about what happens when the indicator she is inspecting is immersed in C3—a substance that is significantly different in chemical composition from both C1 and C2. As it happens, the type of indicator that S is now inspecting remains yellow when immersed in C3. Thus, the indicator would not have turned from yellow to blue unless the solution was C1.

Under these circumstances, it would be implausible to say that S knows that the solution is C1. Indeed, it is no more plausible to say that S knows that the solution is C1 in the modified example, than to say that S knows that the solution is a base in Dretske’s original example. In each case, there is a relevant alternative to what S believes that S cannot exclude, even though S’s reason is in point of fact incompatible with that alternative. But note that, in the modified example, S’s reason is already specific with regard to what it is that displays the features on which S relies. S’s reason is that an indicator of the first type turned blue—not simply that an indicator, or a yellow indicator, or this indicator, turned blue. Indeed, we can even imagine that S knows the precise chemical composition of the indicator, the temperature of the solution in which it is immersed, etc., and that S relies on all of these features in coming to believe that the solution is C1. Here S’s reason is as specific as one could wish, but none of this mitigates the fact that, given what S knows, she is unable to exclude the possibility that the solution is C3, and thus she fails to know that the solution is C1. If this is right, then the insufficiency of (A)–(C) cannot be fixed simply by requiring that an agent’s reasons be properly specific with respect to what it is that displays the features on which the agent relies.

One might respond by saying that there is no reason to think that C3 is a relevant alternative to begin with—or at any rate an alternative that must be ruled out in order for S to count as knowing. We are seldom, if ever, able to rule out all (logically possible) alternatives to what we believe; and so, given

---

3 One could respond by saying that there is a sense in which the agent is able to exclude these possibilities: namely, her reason itself is incompatible with these possibilities, and it is her reason. We ought to tread carefully here. Just because her reason—the proposition on which she relies—is, or happens to be, incompatible with the alternatives, if she herself cannot reason in such a way as to exclude them—if all she has to say is, ‘I’ve never heard of C3’—then there is also a very important sense in which she cannot rule them out. When we speak of ruling out alternatives in the sense that is important for knowledge, what we have in mind is a competence on the part of the agent—a competence that is manifestly not present when the agent responds in this way. To put things another way around, I’m ready to allow that there is a quite thin sense in which the agent ‘rules out’ alternatives to what she believes when she lays hold of a reason which satisfies (CR), but it is not the sense of ruling out alternatives that we need where knowledge is concerned. This point is further elaborated in section 5.

4 In evaluating the truth of ‘The indicator would not have turned from yellow to blue unless the solution was C1,’ we hold fixed the proper functioning of the indicator. Again, this circumstance is logically and causally independent of the fact that the solution is C1; so, on Dretske’s account it is among the circumstances held fixed in evaluating the ‘unless’ claim (see note 1). Add to this if you wish that this particular indicator, or indicators of this general type, do not easily malfunction.
that we sometimes know things, there must be some alternatives that we can safely ignore. I don’t think that this response will save the sufficiency of (A)–(C). Despite widespread disagreement over what makes an alternative relevant, it is all too easy to supplement the case with the right background circumstances in order to guarantee that the C3 alternative is indeed relevant—whatever one’s notion of relevance. Suppose that one is the sort of person who thinks that the relevance of an alternative depends on objective factors (such as whether there was any ‘objective chance’ that the solution is C3). Then imagine that the solution was chosen at random from among C1, C2, and C3, with each outcome just as likely as any other. Or, suppose, instead, that one is the sort of person who thinks that relevance depends on what alternatives are contextually prominent in a given conversational context. Then imagine that two scientists are attempting to discover, and are genuinely uncertain about, which of the three solutions (C1, C2, and C3) this one is, and that they are wondering whether they should consult with S—whether, that is, S knows. And so on. Altering the background circumstances of the case in any of these ways does not change the fact that the indicator’s shift in colour satisfies the counterfactual constraint: the indicator would not have turned blue unless the solution was C1. So, S has laid hold of a reason that satisfies (CR), and yet is not in a position to rule out a relevant possibility, one that must be excluded if she can be said to know. Insisting that the agent’s reasons be properly specific with respect to what it is that displays the features on which she relies does not help.5

Dretske suggests a further way in which reasons can be made appropriately specific. Consider another of his examples [ibid.: 21; again, the case’s name is added by me]:

**Behaviour.** Suppose K is behaving in such a way that it is true to say that he would not be behaving in that way unless he was nervous. Suppose S purports to know that K is nervous and when asked how he knows this, replies by saying, ‘From the way he is behaving.’ Once again, our three conditions are satisfied, or can easily be assumed to be satisfied. Yet, if we suppose that the distinctive thing about K’s behaviour is that he is doing B1 while performing B2, then if S is relying on B1 (or B2) alone, we should not say that he knows that K is nervous. … [T]he fact is that the crucial aspects (those aspects which make K’s behaviour conclusive) are more specific than those on which S is relying in purporting to know.

Here Dretske’s solution is to require that an agent’s reasons be properly specific not just to what it is that displays the particular features, but to those features themselves. Thus, S’s reason must be that K is performing B1 while performing B2, not just that, for example, K is behaving in the way he is

---

5 One need not be a ‘relevant alternatives’ theorist to see the force of the fact that S is unable to rule out the C3 alternative. Whatever account of knowledge one adopts, it seems right to say that knowing that P requires being able to rule out at least some alternatives to P. But once we concede this, it is just a matter of making plausible that this particular alternative—namely, that the solution is C3—is one that S must be in a position to rule out if she can be said to know that the solution is C1. This, I submit, is something we can do just by fixing the background circumstances of the case in the right way. Moreover, we can do so without affecting the judgment that S’s reason satisfies the relevant counterfactual condition on reasons (i.e. that R would not be the case unless P were the case).
(like _that_). But, again, it’s hard to see how this solves the problem. For suppose that _S_ relies on both _B_₁ and _B_₂ in believing that _K_ is nervous. _S_ knows that, when _K_ is happy, _K_ performs _B_₁ without _B_₂, or _B_₂ without _B_₁, but never the two together. Nevertheless, _S_ does not know that when _K_ is sad (but not nervous) _K_ also does not perform _B_₁ and _B_₂ together. Asked what _K_ does when _K_ is sad, _S_ says, ‘I really couldn’t tell you; he’s never been that way around me.’ (Add to the picture, if you like, that _K_ is very often sad, that the question of whether _K_ is nervous, happy, or sad has just come up in conversation, that participants to this conversation are wondering whether _S_ knows which of these is the case, and so on.) Here it seems to me that, just as before, although _S_ is relying on the right aspects of _K_’s behaviour—on those aspects in virtue of which his behaviour satisfies the relevant counterfactual condition— _S_ cannot be said to know that _K_ is nervous.

Note, finally, that we are not forced to think of _S_ in any of the examples as believing any given false proposition. In particular, _S_ need not believe, in _Chemistry_, that the solution was chosen from among _C₁_ and _C₂_, but _not_ _C₃_, or believe in _Behaviour_ that _K_ is never sad. In _Chemistry_, _S_ may simply fail to consider the _C₃_ possibility, without falsely believing that _C₃_ was not among the items that the solution was chosen from; in _Behaviour_, _S_ may simply fail to consider the possibility that _K_ is sad, in relying on the reason she does.⁶

What exactly do examples like _Poker_, _Chemistry_, and _Behaviour_ show?

It would be a mistake to conclude that _S_’s reason does not, after all, bear the right objective relationship to the proposition believed.⁷ Her reason does satisfy (CR), and when (CR) is satisfied the agent’s reason guarantees the truth of the proposition believed. Her reason, as I will say in what follows, is ‘in point of fact’ conclusive.

What the examples show is, rather, that it is possible to believe _P_ on the basis of an in point of fact conclusive reason and yet fail to know that _P_ is the case. If _R_ is conclusive for _P_, then _R_ guarantees the truth of _P_: _R_ would not be the case unless _P_ were the case. But _R_ cannot also guarantee that anyone who believes _P_ on the basis of _R_ is thereby in a position to know that _P_. The examples above are cases in point: cases in which the right objective relationship obtains between _R_ and _P_, and still, in coming to believe _P_ on the basis of _R_, _S_ gets things right about _P_ (at least partly) as a matter of luck.

---

⁶ The basic distinction here is between believing that a certain possibility does not obtain, and failing to consider that possibility in the first place. Imagine a police detective who has been working on a murder case for a number of years, and solves it when he discovers a clue that points him in an entirely new direction. It is not that at first the detective mistakenly believed that Mrs. Brown did _not_ kill her ex-boss. Rather, the detective had never heard of Mrs. Brown, did not know that she was dismissed from her job some months prior, and so on.

⁷ This thought should be sharply distinguished from the thought that there is something defective about _S_’s epistemic position _vis à vis_ the proposition _S_ believes. There surely _is_ something defective about the latter. This is why we judge that _S_ fails to know in these cases.
knowledge and epistemic luck: sensitivity and safety. These principles share in the broadly modal character of (CR): like (CR), they speak to what would or might have been the case, had things been different from what they are. Moreover, the satisfaction of one or another of these constraints has variously been taken to eliminate the kind of luck that interferes with knowledge. To take a recent passage from Yamada, ‘If one is sensitive to the truth, it is surely no accident if one’s beliefs turn out to be true’ [2011: 75]; and, as Pritchard says, ‘If one meets the safety principle, then it is not a matter of epistemic luck that one’s belief is true’ [2007: 156].

Of course, safety and sensitivity are not specifically tailored to cases of reasons-based knowledge, and so, at least on the face of things, do not purport to tell us anything about what it is to have conclusive reason to believe that P. Nevertheless, in the light of the examples considered above, we do well to ask whether Pritchard and others are right—whether, that is, the satisfaction of these modal constraints is incompatible with the kind of luck that interferes with knowledge. In this section we focus on this question. I’ll argue that the same examples that show that (CR) does not rule out getting things right by ‘accident’ also show that safety and sensitivity do not eliminate pernicious epistemic luck. (The reader uninterested in safety and sensitivity is invited to proceed to section 5.)

According to a simple formulation of sensitivity, S’s belief that P is sensitive just in case if P were not the case then S would not believe that P. This condition first appeared in Nozick’s tracking theory of knowledge, where Nozick argued that sensitivity is a necessary condition on knowledge. Well-known problems attach to this first formulation of sensitivity, in response to which Nozick adopted a more sophisticated version of the principle that appeals to ‘methods’ of belief formation. Suppose that S believes via the method M that P is the case. Then S’s belief is sensitive if it satisfies this:

Sensitivity. If P were not true and S were to use M to arrive at a belief whether P, then S would not believe, via M, that P. 9

Being sensitive to the truth of P does not imply that one would never falsely believe that P. Rather, it implies only that in the nearest possible worlds in which not-P is true, one does not believe (via M) that P.

It is not difficult to see that the sensitivity requirement rules out (as cases of knowledge) many straightforward examples in which the agent gets things right about P only as a matter of luck. To take a familiar case from Gettier [1963: 122], suppose that Smith and Jones have applied for the same job and that Smith has every reason to believe that Jones will get it. Smith also

---

8 Saying just what kind of luck interferes with knowledge is not an easy task. For a useful discussion of some different kinds of epistemic luck, see Unger [1968]; for a useful discussion of some of the pitfalls facing current attempts to characterize luck, see Lackey [2008]. Without taking a stand on how to characterize luck or epistemic luck in general, I’ll use the expression ‘the pernicious kind of epistemic luck’ for the kind of luck that interferes with knowledge.

9 Further complications arise when more than one method is in play, either actually or counterfactually. The cases we consider are ones in which it is either stipulated or intuitively obvious that just a single method is employed.
knows that Jones has ten coins in his pocket. He comes to believe on this basis that

\[ P: \text{The man who will get the job has ten coins in his pocket.} \]

As it turns out, Smith himself will get the job, but—unbeknownst to him—he too has ten coins in his pocket. In this Gettier case, Smith’s belief that \( P \) is true, and yet we judge that Smith fails to know. After all, Smith believes that \( P \) only because Jones has ten coins in his pocket, but \( P \) is true in virtue of the fact that Smith has ten coins in his pocket. Nozick’s sensitivity condition handles the case neatly. The nearest possible worlds in which \( P \) is false are ones in which Smith does not have ten coins in his pocket. But in those worlds Smith continues to believe that \( P \). (He believes that \( P \) because Jones has ten coins in his pocket.) So, Smith’s belief is not sensitive and hence, on Nozick’s account, does not count as knowledge. This is plainly the right verdict. Although sensitivity has fallen out of favour among epistemologists in recent years, it is still widely thought to eliminate the pernicious kind of epistemic luck.

A more current condition on knowledge popularized by Sosa [1999a, 1999b], Williamson [2000], and Pritchard [2005] is the safety principle. As with sensitivity, a sophisticated version of safety takes methods into account. Suppose that \( S \) believes \( P \) via the method \( M \). Then safety is this condition:

**Safety.** In the nearby possible worlds in which \( S \) comes to believe that \( P \) via \( M \), \( P \) is true.

Note that safe belief does not entail sensitive belief (nor vice versa). If \( P \) could not easily have been false, then the nearby worlds where \( S \) believes that \( P \) are ones in which \( P \) is true. So, \( S \)’s belief that \( P \) is safe. But it may nevertheless be the case that, in those distant worlds where \( P \) is false, \( S \) believes that \( P \). Hence, \( S \)’s belief is not sensitive. As with sensitivity, many philosophers (e.g. Latus [2000]; Pritchard [2005]; Yamada [2011]) have claimed that safety rules out cases of the pernicious kind of epistemic luck.

Is this claim defensible? Let’s start by taking another look at *Poker*, where the intuition is perhaps strongest that Jessica gets things right only as a matter of luck. Recall that in that example Jessica believes that John has a king, on the basis of the fact that he raised and there was a king on the table. To bring the case more in line with talk of methods, suppose that Jessica goes around forming beliefs about John, according to the method \( M \): namely, when John raises and there is a king on the table, believe that John has a king. As we said above, John is committed to playing by Strategy—he would play by Strategy, no matter what cards he had. Thus, if John did not have a king, he would not have raised. (Strategy requires that a player raise only if he or she has a king.) If Jessica were then to form a belief about whether John has a king, using the method \( M \), she would not believe that John has a king. So, Jessica’s belief is sensitive: in the nearby possible worlds where John does not have a king, she does not believe (via \( M \)) that he does.
Is Jessica's belief also safe? In the nearby possible worlds where Jessica believes that John has a king via the method M, it is because John raised. But in those worlds John plays by Strategy—he always plays by Strategy. So, given that he raised, John does have a king. If this is right, then Jessica’s belief is safe: in the nearby possible worlds where she believes via M that John has a king, she gets things right. The example shows that neither safety nor sensitivity rules out the pernicious kind of epistemic luck.

One might object at this point that a more natural description of Jessica’s method is to believe of anyone in the game that the person has a king if he or she raises and there is a king on the table. If so, her method is bound to produce false beliefs—if not about John, then about other players in the game. I don’t think that this line of defence will get sensitivity or safety out of the water. First, even if we do take on board this revised description of Jessica’s method, it does not change the fact that Jessica’s particular belief about John is safe. In the nearby possible worlds where she comes to believe the proposition she does (namely, that John has a king), it is because John raised, and John, as we know, only raises on a king when he has a king in his hand. Jessica’s belief about John is safe, and yet she is lucky in getting things right with regard to this particular proposition. Second, by supplementing the example with the right background circumstances, we can generate the intuition that our first description of Jessica’s method is right. Suppose, for example, that Jessica believes that John is the only player at the table who really knows how to play Poker—the others are too dull to get it. Or suppose that she only takes interest in what cards John has; she is simply not engaged in (perhaps she is even resolutely opposed to) figuring out what cards are held by the other players. Then her method really is specific to John. ‘Believe of John, if he raises and there is a king on the table, that he has a king!’ Using this method, Jessica still gets things right as a matter of luck. She is lucky, as we might say, in adopting what happens to be a truth-conducive method.

Cases like Poker differ in important respects from the more familiar Gettier cases. In Gettier cases, the agent is justified in believing what he does but believes it for the wrong reason. Smith, for example, is justified in believing that the man who will get the job has ten coins in his pocket, but he believes this because he thinks that Jones has ten coins in his pocket. In cases like this, it’s quite easy to see how luck enters into the picture. The agent justifiably believes a proposition for one reason, and that proposition ends up being true for entirely different reasons. In Poker, by contrast, Jessica is intuitively not justified in believing that John has a king. But her reason is nevertheless in some sense the ‘right’ one. That John raised is in point of fact conclusive for the proposition that John has a king: that reason would not obtain unless the proposition believed were true. To put the point succinctly, in Gettier cases the agent is justified in believing what she does, and she gets lucky when what she believes turns out to be true for other reasons. In Poker and similar cases, the agent is not justified in believing what she does, and she gets lucky when the proposition she believes turns out to be true for the very reason she believes it. The examples show that this second scenario is not as paradoxical as it sounds. One can get lucky not just in believing the right thing, but in relying on the right reasons.
If this is right, then it is not enough for the agent’s reason to satisfy the right counterfactual constraint, or (in the case of sensitivity and safety) for the agent’s method to be reliable. One can believe that P on the basis of reasons that satisfy the right counterfactual constraint and still get lucky in fixing on just that reason. One can rely on methods that are reliable—that in one way or another modally guarantee the truth of the proposition believed—and still get lucky in adopting just that method.

5. A Negative Moral

Whether we focus on modal constraints that pertain only to reasons, or those which, like sensitivity and safety, pertain to methods, the negative moral is, I think, roughly the same. There is, on the one hand, the objective relationship that obtains between our reason (or method) and the proposition believed, and there is, on the other hand, the agent’s epistemic position vis à vis the proposition believed. The point of the counterexamples is that the two can come apart. An agent can believe that P on the basis of reasons that satisfy the right counterfactual condition—or methods that modally guarantee the truth of P—and yet still not have what it takes, epistemically speaking, to know that P is the case. To know that P is not simply a matter of the objective strength of our reasons or the reliability of our methods.

We can put things another way around by talking, as epistemologists often do, about the possibilities ‘left open’ by an agent’s evidence. On the one hand, there are those possibilities that might be the case given that R is the case. But there is also the way that we as agents are able to reason with respect to those possibilities. This distinction is too often neglected. Suppose that S believes P on the basis of R, and R would not be the case if Q were the case. Then the agent’s reason is incompatible with the alternative Q. Nevertheless, it does not follow that, in believing P on the basis of R, S is in a position to rule out the alternative Q. Whether or not S is able to do so depends, in general, on what other things S knows, on S’s broader epistemic state. We see this directly in the examples. In *Chemistry (Modified)*, the indicator’s change in colour is counterfactually incompatible with the possibility that the solution is C3 (the colour of the indicator would not change if the solution were C3). But, nevertheless, in relying on that reason, S is unable to rule out the possibility that the solution is C3. Asked how she knows that it’s not, S says, ‘I don’t know anything about C3,’ or ‘I hadn’t considered that possibility.’ Seen in this light, what counterfactuals like (CR)—or modal conditions like sensitivity and safety—guarantee is only one part of the equation. They guarantee that certain possibilities are incompatible with one’s reason or method. But they do not also guarantee that an agent who relies on that reason or method is able to reason in any particular way with respect to those possibilities. If knowledge requires being able to rule out alternatives, where this is understood as an epistemic competence, then purely external constraints like (CR), sensitivity, and safety won’t do the job.
6. Having Conclusive Reason

Does this mean that having conclusive reasons is not sufficient for reasons-based knowledge? That sounds implausible. After all, in saying that S has conclusive reason to believe P, we seem to assert that S’s grounds are as good as they come— that her grounds leave no room for doubt. How, then, could S fail to know?

If having conclusive reasons to believe P is just believing P on the basis of a reason that is in point of fact conclusive then, in view of the examples, we must concede that having conclusive reasons is not sufficient for knowledge. An agent can rely on reasons that satisfy (CR) and yet fail to know that P is the case. But, to reverse the point, if having conclusive reasons is sufficient for knowledge, then what it is to have conclusive reasons cannot be (simply) to believe P on the basis of reasons that are in point of fact conclusive— reasons that satisfy (CR).

Consider the following purportedly valid argument: 10

(P1) S knows that P if S has conclusive reason to believe that P.
(P2) A reason R is conclusive for P if and only if R would not be the case unless P were the case.
(C) S knows that P, if S believes that P on the basis of R and R would not be the case unless P were the case.

The examples in sections 2 and 3 show that the conclusion of the argument is untenable. An agent can believe P on the basis of a reason that satisfies the counterfactual in (C) and yet fail to know that P. It looks as though we must reject one of the two premises. I’ve argued already that we have no reason to reject (P2). That premise specifies correctly the relationship that must hold between the propositions R and P if R is conclusive for P. When it’s true that ‘R would not be the case unless P were the case’, R guarantees the truth of the proposition P. (P1), on the other hand, states that having conclusive reason for P is sufficient for knowing that P; and this, too, seems difficult to deny. How could someone whose reasons for P are conclusive fail to know that P? But if (P1) and (P2) are true and (C) is false, there is only one way out. We must deny validity.

Looking at the two premises, it is not difficult to see why the argument is invalid. Premises (P1) and (P2) are concerned with different things. (P1) is concerned with the strength of the agent’s epistemic position vis-à-vis the proposition believed. (P2) is concerned with a counterfactual relationship between propositions that holds independently of what any given agent believes or knows. The argument is invalid because to have conclusive reason for believing P is not (simply) to believe P on the basis of a reason that is in point of fact conclusive.

Some readers may object that putting the point in this way is confused. It suggests that a person may believe P on the basis of a reason, and yet not

---

10 I think that something like this argument is at the heart of Dretske’s [1971] analysis of knowledge, although he never makes it explicit.
have that reason for believing P. The argument, they would say, forces us to give up (P2) rather than to deny validity: what it takes for R to be conclusive for P has been shown to involve more than the satisfaction of the simple counterfactual constraint. But I think that this is the wrong lesson to draw. So long as we are concerned just with an objective relation that holds between propositions—what it is for R to be conclusive for P, independent of any agent S—there seems to me to be nothing wrong with (P2). Where R is the case and the counterfactual in (P2) is satisfied, R guarantees the truth of P.

One might object, in a more radical spirit, that there is simply no sense to be made of talk of one proposition being ‘conclusive’ for another, independent of epistemic agents. Thus, the locution ‘a reason R is conclusive for P (independent of S)’ in (P2) is improper; the only proper locution is ‘S has conclusive reason for P’, as in (P1). Although I am to some degree sympathetic to this line of thought, the point of the present paper is to start with something like Dretske’s view, which gives central place to a relation among propositions, and to see how far this can take us. The counterfactual condition that Dretske identifies does seem to me to specify an important relation that reasons themselves (considered as propositions) may, or may not, satisfy in relation to the proposition believed. But, of course, the main take-away of the paper is that just looking at modal relations that hold among propositions doesn’t in the end get us very far at all in our quest to understand what it is to know on the basis of reasons.

7. Justification in Taking R to Be a Reason

But what is it, then, to have conclusive reason to believe that P?

Presumably, an accurate account of this epistemic notion will steer clear of the counterexamples given above. It will not predict that S has conclusive reason to believe that the chemical solution is C1 but is not able to rule out the possibility that it is C3, or that S has conclusive reason to believe that K is happy but is not able to rule out the possibility that K is sad. One easy way to ensure this would be to say simply that having conclusive reasons for P is, or requires, being able to rule out some set of alternatives to P. But this would be quite unilluminating. It would leave the connection between having conclusive reasons, and reasons themselves, entirely unexplained. Indeed, on such a view, it would look as though having conclusive reasons has nothing to do with reasons.

A different strategy would allow that believing P on the basis of reasons that satisfy (CR) is at least a necessary condition on having conclusive reason to believe that P. We could then go on to ask what further conditions need to be added in order to yield an adequate account of that notion. This has the advantage of ensuring a tight connection between having conclusive reasons, and reasons themselves. But what extra conditions are needed? In the counterexamples we considered above, it is tempting to say that, while the agent’s reason is in point of fact conclusive, it is not a conclusive reason for her. What’s left out isn’t anything about the modal relations the reason
itself stands in, or anything about the strength of the agent’s belief (the agent believes, without a trace of doubt, that P is the case). Rather, it is something to do with the agent’s relationship to her reason as a reason. Helping ourselves to the notion of epistemic justification, it seems natural to require that an agent has conclusive reasons to believe that P only if she is justified in taking her reason to be a reason for P in the first place.

In the spirit of these remarks, we might propose the following.

(HCR) S has conclusive reason R for believing that P if:

(1) S believes P on the basis of R;
(2) R would not be the case unless P were the case;
(3) S knows that R is the case; and
(4) S is justified in taking R to be a reason for P.

(Aside: Whatever our underlying theory of justification, we cannot here understand what it is for S to be justified in taking R to be a reason for P in terms of a demand that S have a conclusive reason for believing that R is a reason for P. First, this demand is too strong. We are often justified in believing what we do not have conclusive evidence for. Second, such a demand would lead to infinite regress. In order to have conclusive reason to believe P, S would have to rely on a reason R that is conclusive for P. But, furthermore, she would have to have conclusive reason to believe that R is a reason for P. And so she would have to rely on a further reason R’ which is conclusive for [R is a reason for P]. And so on. This regress would quickly ensure that no agent ever has conclusive reason to believe that P.)

Now, surely (HCR) avoids some of the counterexamples to Dretske’s proposal discussed above. In Poker, Jessica is not justified in taking the fact that John raised to be a reason for believing that he has a king. Her doing so is grounded in her own ignorance, a failure to conceive of the real possibilities in the game, and not in any appreciation of John’s strategy. She does not satisfy (4), and so, according to (HCR), does not have conclusive reason to believe that John has a king. This is plainly the right verdict. Cases like this point in favour of adopting (1)–(4) as necessary conditions on reasons-based knowledge.

Nevertheless, I think that simple examples show that (1)–(4) are not jointly sufficient. Consider one final example:

11 It is less clear what to say about examples like Chemistry (Modified) and Behaviour. In Chemistry (Modified), S is able to rule out the possibility that the solution is C2, but is not able to rule out the possibility that the solution is C3. Is S justified in taking the indicator’s change in colour to be a reason for believing that the solution is C1? Despite being unable to rule out what is in fact a relevant alternative, an agent can sometimes be justified in believing what she does (and indeed in taking her reason to be a reason). Consider Goldman’s famous barn façade case [1976: 772–3], in which Henry is driving through a part of the countryside littered with papier mâché facsimiles of barns. When Henry looks out of the window and sees what happens to be a real barn, he comes to believe that there is a barn in the field. (He does not know about the papier mâché facsimiles.) Under the circumstances, the possibility that the structure in the field is only a facsimile is a relevant alternative to what he believes. But Henry is presumably justified in believing what he does, despite being unable to rule out this alternative. (Of course, according to ordinary intuitions, Henry does not know that what he sees is a barn, but he is nevertheless justified in believing it.) In examples like Chemistry (Modified) and Behaviour, a lot more would need to be said about the background circumstances of the case before we could determine whether S is justified in believing what she does or in taking her reason to be a reason.
Augie drives to his local KFC, intending to drop in on his friend Dylan who recently got a job there. As he pulls up, he sees that the electric ‘CLOSED’ sign in the window is illuminated and the lights in the restaurant are out. He concludes that Dylan has left for the night, and with grave disappointment returns home. Does Augie have conclusive reason to believe that Dylan is away? Before deciding, consider that until last week this branch of KFC did not have an electric ‘CLOSED’ sign at all. Moreover Dylan, who recently broke up with his girlfriend, has been sleeping in the back room of KFC for the past two weeks. In his desperate circumstances, he entered into a deal with a local crime boss who is planning to burgle KFC: Dylan gets half the money if he cooperates by absenting himself from the premises on the appointed night. The ‘CLOSED’ sign, which the crime boss bought and gave to Dylan a week before, functions as their go-ahead signal. When Dylan leaves the premises, he turns out the lights and illuminates the sign.

In the example, Augie’s reason—that the ‘CLOSED’ sign is illuminated—satisfies condition (2) (because of the agreement between Dylan and the crime boss, the ‘CLOSED’ sign would not be illuminated unless Dylan was away). Moreover, Augie believes that Dylan is away, on the basis of the fact that the sign is illuminated; and he knows that the sign is illuminated, because he sees it right there in the window. Finally, Augie is justified in taking the illuminated sign to be a reason to believe that Dylan is not there. (Consider the way that ‘CLOSED’ signs typically function, that Augie has no reason to think that this particular sign functions any differently, or that Dylan would stay past the hour at which the restaurant closes, and so on.) So, Augie satisfies (HCR).

Nevertheless, it would be wrong to say that Augie has conclusive reason to believe that Dylan is away. The sign he relies on does not function as he thinks it does; the conclusiveness of his reason is located in facts that are entirely unknown to Augie. It’s true that the sign would not be illuminated unless Dylan was away, so the reason Augie relies on satisfies the right modal constraint. But so long as the secret arrangement between Dylan and the crime boss is unknown to Augie—so long, that is, as Augie is ignorant of the very facts that make it the case that his reason is conclusive—Augie cannot be said to know that Dylan is away. Again, one wants to say that, while Augie’s reason is in point of fact conclusive, it is not a conclusive reason for him. If this is right then conditions (1)—(4) are not jointly sufficient for having conclusive reason to believe that P.

Another case that illustrates nicely the insufficiency of (HCR) is the case of the holographic vase, discussed by Nozick and attributed to Shope. A person comes to believe that there is a vase in a box by looking at a hologram (of a vase) projected onto the box. As it turns out, there is a real vase in the box. Indeed, the hologram is part of a machine that turns on a

\[12\] In evaluating the ‘unless’ claim, we hold fixed the agreement between Dylan and the local crime boss. Indeed, the agreement was entered into well before the point at which Dylan absents himself; so, by Lewis’s ban on backtracking, it is among the conditions held fixed in evaluating what would have happened had Dylan not been away. Moreover, we can imagine that Dylan is entirely committed to this agreement, that the decision to enter into the agreement in the first place was determined by several modally robust factors, and so on. These circumstances ground the claim that the worlds in which Dylan and the crime boss do not enter into the agreement are remote ones—or, at any rate, not among the nearest possible worlds in which Dylan is not away.
holographic depiction of a vase only when there is a real vase in the box. If we construe the agent’s ‘reason’ for belief as ‘it looks like there is a vase in the box’, then he satisfies conditions (1)—(3). Moreover, the agent is presum-
bly justified in taking his reason to be a reason (assuming that he has generally good eyesight, and so on). So, he satisfies (HCR). But many philosophers think that the agent in the example does not know that there is a vase in the box. Likewise, I think that these philosophers would deny that the agent has conclusive reason to believe that there is a vase in the box, given that he is is unaware of the mechanism.14

8. Conclusion

Cases like KFC speak, I think, to the deep penetration of the phenomenon of epistemic luck. The problem, on one way of looking at things, is that there can be situations in which the facts that justify the agent in taking R to be a reason for P are wholly distinct from those facts that ground the conclusiveness of R for P. Yes, the agent in KFC is justified in taking his reason to be a reason, and yes, his reason is in point of fact conclusive. But his justification does not link up in the right sort of way with the facts that ground the conclusiveness of his reason. It is in the gap between the two where luck, as it were, resides. Similarly, one can say that, in the Gettier cases, luck resides in the gap between the agent’s justification for believing P and the facts that ground the truth of P. (Smith’s justification for believing that the man who will get the job has ten coins in his pocket is that Jones has ten coins in his pocket; the fact that grounds the truth of the proposition Smith believes is that Smith has ten coins in his pocket.) We have moved, on this way of looking at things, one level up: from luck that appears at the object level, in the gap between justification for P and the facts that ground the truth of P, to luck that appears at the meta-level, in the gap between justification for taking R to be a reason for P and the facts that ground the conclusiveness of R for P.

Where does this leave us? The failure of (CR) calls into question, as I’ve argued, whether any such modal constraint can effectively rule out getting things right as a matter of luck. Moreover, the failure of (HCR) suggests that we cannot correct for the problem simply by tacking on to any such modal constraint a requirement that the agent is justified in taking his reason to be a reason. My own inclination is to think that some more radical solution is called for—a different way of conceiving what it is to have conclusive reason to believe that P. If we take seriously the claim that to have conclusive reason for P is sufficient for knowing that P, and the claim that knowing that P requires ruling out some set of alternatives to P (where this is most plausibly understood as a competence on the part of the agent), then perhaps the way to understand what it is to have conclusive reason to believe

13 It may, in the end, prove to be more natural to treat this case in terms of methods rather than reasons.
14 Nozick himself [1981: 190] suggests that it may, after all, be right to say that the agent does know: ‘This consequence is somewhat counterintuitive; however, we certainly do not want to hold that a person knows that p only if he has no false beliefs about the process via which he comes to believe that p.’ See Shope [1984: 38] for further discussion of the case. Thanks to an anonymous referee for pointing me to this example.
that P is in terms of reasoning rather than reasons. Perhaps, that is, our attention has been mistakenly focused on the propositions on which the agents relies, and the modal relations in which they stand to other propositions, rather than on what it is that epistemic agents are able to do with those reasons, the ways in which they are able to reason with respect to them.

This, of course, does not leave us with a positive account of what it is to have conclusive reasons to believe that P. But it does, I think, point us in a fruitful direction. Where knowledge that P is based on reasons, one must have conclusive reason to believe that P. Rightly understood, having conclusive reasons is not simply a matter of the objective relations that hold between one’s reason and the proposition believed on that basis. What matters is our relationship to our reasons as reasons—and how, using those propositions, we are able to ‘reason’ with respect to the various possibilities that must be ruled out if we can be said to know.

Columbia University

Received: November 2014
Revised: May 2015

References