Why safety doesn't save closure

Marc Alspector-Kelly

Abstract   Knowledge closure is, roughly, the following claim:

> For every agent S and propositions P and Q, if S knows P, knows that P implies Q, and believes Q because it is so implied, then S knows Q.

Almost every epistemologist believes that closure is true. Indeed, they often believe that it so obviously true that any theory implying its denial is thereby refuted. Some prominent epistemologists have nevertheless denied it, most famously Fred Dretske and Robert Nozick. There are closure advocates who see other virtues in those accounts, however, and so who introduce revisions of one sort or another in order to preserve closure while maintaining their spirit. One popular approach is to replace the "sensitivity" constraint at the heart of both of those accounts with a "safety" constraint, as advocated by Timothy Williamson, Duncan Pritchard, Ernest Sosa, Stephen Luper, and others. The purpose of this essay is to show that this approach does not succeed: safety does not save closure. And neither does a popular variation on the safety theme, the safe-basis or safe-indicator account.

Keywords   Safety · Sensitivity · Safe · Sensitive · Safe-basis · Safe-indicator · Closed · Closure · Knowledge

1 Sensitivity and closure

Knowledge closure is, roughly, the following claim:

M. Alspector-Kelly (✉)
Department of Philosophy, Western Michigan University, Kalamazoo, MI 49008-5328, USA
e-mail: malspect@wmich.edu

For every agent S and propositions P and Q, if S knows P, knows that P implies Q, and believes Q because it is so implied, then S knows Q.[1]

Almost every epistemologist believes that closure is true. Indeed, they oftenn believe that it so obviously true that any theory implying its denial is thereby refuted. Some prominent epistemologists have nevertheless denied it, most famously Dretske (1970, 2005a,b) and Nozick (1981). Having discerned other virtues in Dretske's and/or Nozick's accounts, some closure advocates have introduced revisions of one sort or another in order to preserve closure while maintaining their spirit.[2] One popular approach is to replace the "sensitivity" constraint at the heart of both of those accounts with a "safety" constraint, as advocated by Timothy Williamson, Duncan Pritchard, Ernest Sosa, Stephen Luper, and others.[3] The purpose of this essay is to show that this approach does not succeed: safety does not save closure. And neither does a popular variation on the safety theme, the safe-basis or safe-indicator account.

Nozick and, supposedly, Dretske—I will explain the "supposedly" below[4]—require that in order for S to know that P her belief in P must be sensitive, where a belief is sensitive iff were P to be false S would not believe it. This modal conditional is

---

[1] I say "roughly" because it is widely recognized that it is in fact difficult to formulate closure in a manner that is not susceptible to obvious counterexamples (obvious, that is, to both closure advocates and opponents). The closure advocate nevertheless inevitably claims that there is a formulation within near conceptual proximity of the stated but admittedly false version, albeit one the advocate rarely attempts to formulate, but which is obvious and trivial (or, presumably, would be obvious and trivial upon explicit presentation). The history of such claims—that the standard expression of a principle is clearly false but that a conceptual cousin is nearby which just needs a bit of work to formulate, and which is both somehow obvious and can do serious philosophical work—might reasonably give one pause. Consider, for example, the history of the analytic/synthetic distinction.

Hawthorne's (2005) is a notable exception to this breezy attitude toward closure's formulation on the part of its advocates. His version—modeled on that of Williamson (2002)—is now perhaps the standard formulation rather than that offered above. According to it, if you know P and competently deduce Q from P while knowing P throughout, and come to believe Q as a result, then you know Q. These formulations will be equivalent if a belief's being based on a competent deduction is equivalent to its being based on knowledge of the inference relation (and I am uncertain which is better if they are not equivalent). At any rate, the reader is welcome to substitute the Hawthorne/Williamson formulation for that presented here. It won't matter in what follows, since both apply to the cases discussed below.

[2] Aside from the safety view that is the subject of this paper, other good examples are the contexualisms of Stine (1976) and DeRose (1995) and the updated version of Nozick's view advocated in Roush's (2005). Roush's account is a particularly direct version of this approach: her view essentially involves appending knowledge-by-closure to (probabilistic variations of) Nozick's tracking conditions, so that one can know either by tracking or by inferring from what one tracks even though one might well not track the latter. Aside from the bare fact that it preserves closure by fiat, as it were, it is hard to see why one attracted to a Nozickian tracking account would endorse a view according to which we are in a position to know a lot that we don't track. (Given the prevailing sentiment in favor of closure, however, perhaps that's enough.)

[3] Williamson (2002), Pritchard (2005, 2008, 2009), Sosa (1999a,b), Luper (2006). One of the earliest statements of such a view—perhaps the earliest—is Luper 1984 (in which a version of what is referred to here as the safe-basis account is defended). I should note that significant variety exists within this family of views (even beyond that which I will mark between safety-belief and safe-basis views). Luper (1984), for example, refers to a sequence of causes, each of which is safe vis-à-vis the proposition known, terminating in the belief; Pritchard (2008, 2009) distinguish between worlds that are very close nearby worlds from those that are nearby but not very close (see note 14). I pass over these differences because, so far as I can discern, the points made here concerning the relation between safety and closure apply to all such variations.

[4] See Sect. 4.

almost universally interpreted in accordance with the Lewis/Stalnaker account of such conditionals: in the nearest world in which P is false S does not believe that P.[5] As Nozick and Dretske both concede, sensitivity implies closure failure: it is possible for an agent to have a sensitive belief that P, to infer Q from it, and still have an insensitive belief that Q.

For example: you believe that the Broncos beat Buffalo yesterday; you just read that result in today's Kalamazoo Gazette. That belief is sensitive: if the Broncos had lost, the newspaper would have reported that they had lost. (In the nearest world in which the Broncos lost, the Gazette would report that fact and you would believe that instead.) Since the Broncos did in fact win, the Gazette didn't erroneously report that they won as a result of an accident while typesetting. Suppose that you believe that the erroneous report did not in fact happen. (It follows, after all, from what you believe, namely, that the Broncos beat Buffalo, and you believe in closure.) That belief is not sensitive: if the paper had erroneously reported the score, you would still believe that they hadn't. So if knowledge requires sensitivity, you know that the Broncos won. But you don't know what follows from this, that the Gazette didn't erroneously report the score.

I happen to think that this is the right result. My purpose here is not, however, to attack closure, but only to argue that safety does not preserve it. Notwithstanding such examples, the safety theorist considers closure to be worth preserving. But she still likes the idea that knowledge requires some sort of ruling-out of counterfactual worlds in which you believe P when P is false. So she offers safety instead.

2 Characterizing safety

There are a variety of expressions of safety—were S to believe P, P would be true, S's belief that P is not easily wrong, and so on—but they come down to the same thing when characterized in terms of possible worlds: a belief is safe iff within the space of nearby possible worlds—worlds similar to the actual according to some measure— every world in which the agent believes it, it is true; equivalently, there is no such world in which the belief is false but the agent still believes it.

I am aware of no reasonably precise characterization of the boundary between worlds nearby and remote (call this the "nearness boundary") in the relevant literature. But this much is clear: that boundary does not vary with content of P.[6] There are,

---

[5] Lewis (1973), Stalnaker (1968). Nozick does not insist on this account, and in fact tentatively proposes instead that Q is true throughout the nearest "P neighborhood" of the actual world. See Nozick (1981, p. 680, fn. 8). Dretske's "official" view—his information-theoretic account formulated in Dretske (1981)—is in not expressed in counterfactual terms at all, but is instead expressed probabilistically: S knows that P iff S's belief that P is caused (or causally sustained) by (a state which manifests) the information that P, where a state manifests the information that P when the probability of P given the existence of that state is 1. Notwithstanding this fact, and the fact that probability and possible-world remoteness are not equivalent (see footnote 18), Dretske has, to my knowledge, demonstrated no resistance to the subjunctive formulation of his view (modulo the point in Sect. 4 referred to in the previous footnote), and I here continue the tradition of doing so. (The relationship between the two formulations, however, deserves exploration.)

[6] This contrasts with sensitivity: the boundary between the sphere centered on the actual world in which there are no not-P worlds and the greater sphere (doughnut?) beyond—set by the distance to the nearest possible not-P world(s)—varies with the content of (not-)P.

as a result, two very different ways in which a belief can be safe. If there are worlds in which it is false which lie within that boundary, then safety requires that the agent not believe it in any such world. Call these "near-safe" beliefs. In these cases the safety requirement is more stringent than sensitivity and implies it. If a belief is near-safe then there are nearby worlds in which P is false (by definition of "near-safe"). Therefore the nearest world in which P is false—one of those worlds—is itself nearby. Since the belief is near-safe, the agent believes P in none of those nearby not-P worlds. Therefore, she does not believe it in that nearest world. Therefore, her belief in P is sensitive. Sensitivity does not, however, imply near-safety. Suppose that the nearest not-P world is nearby. Sensitivity requires that she not believe it in that world; but that is consistent with her believing it in more remote but still nearby not-P worlds, thereby violating near-safety.

A belief's being near-safe is not assured by the content of the proposition believed: one person can believe that proposition and know it, while another can believe it and not know it. Something—some feature of the way things actually are with the agent and/or her environment—must therefore rule out the existence of such worlds.[7]

However, if there are no nearby worlds in which the proposition believed is false—if the nearest such world lies beyond the nearness boundary—then the belief is automatically safe: it is not easily wrong simply because the proposition believed is not easily false. Call these "far-safe" beliefs. A far-safe belief is therefore known as soon as it is believed: it is impossible for one person to believe it and know it and another to believe it and not know it.[8] That knowledge therefore imposes no conditions on the agent or her environment whatsoever (beyond, of course, those facts which ensure that the proposition's negation is remote).

It is hard to imagine two more disparate knowledge-conditions. One constrains the agent and/or her environment in order to ensure that certain modal conditions hold, conditions more stringent than those imposed by sensitivity. The other imposes no such conditions whatsoever, beyond mere believing itself: with far-safe beliefs, to believe is to know, no matter why you believe it.

3 Two concerns in passing

While not directly concerned with the subject-matter at hand—the relation between safety and closure—I cannot resist presenting two brief concerns about the safety account just characterized. We will return to the main program in Sect. 4.

---

[7] The mythology of possible worlds tends to encourage the image that the truth-maker of modal claims is settled independently of the facts on the ground of the actual world. The actual world is, after all, only one world among many and enjoys no causal contact with them; how could the nature and distribution of all those infinitely many other worlds depend on how things happen to be on the actual world? But the structure of possible-worlds space is a function of the actual facts: it is because of the actual molecular structure of glass that it is fragile, because Bob is actually standing right next to the cliff's edge that the world in which he falls is unsettlingly nearby, and so on. Modal/dispositional properties have categorical/actual bases.

[8] This assumes that a belief's being safe is sufficient for its being known. We will examine that assumption more closely in Sect. 8.

First, the fact that mere believing can suffice for knowledge of far-safe beliefs is a very worrisome aspect of the view. Safety theorists respond to this worry by pointing out that we do seem somehow committed to believing their paradigmatic far-safe belief—that I am not a brain in a vat—notwithstanding the apparent utter lack of evidence for it (by design, of course).[9] Perhaps such an in-principle-undetectable possibility would have to be known by courtesy, as it were, assuming it to be known at all.

But in other cases such knowledge-by-courtesy is simply untenable. For example, in 1999a Sosa presents the famous garbage-chute case against the sensitivity account. I drop a bag of garbage down the garbage chute in my apartment and it arrives as expected in the basement. I believe, and (Sosa insists) I know, that the bag is in the basement. But if the bag were not in the basement I would still believe that it is, since it would have failed to arrive in virtue of having been caught on a protruding nail (or by having suffered some fate along those lines) in which case I would still think it made it all the way down.[10] Sosa (and others, including Pritchard 2008) claim that the safety theorist is better positioned to deal with this case, since the belief that it is in the basement is, they claim, safe.

It is not, however, near-safe. If it were, then "the bag is in the basement" would be easily false: there would be a nearby world in which the bag is not in the basement, presumably as a result of the same protruding nail that brought the sensitivity account to grief. But then I would still believe it was in the basement, rendering my belief unsafe.[11] So if my belief is safe at all, it is far-safe. But far-safe beliefs are known merely by being believed. So my reason for believing that the bag is in the basement— that I dropped it down the chute—is irrelevant; I know it whatever my reason might be, so long as I believe it.

Now vary the example a bit. Suppose that the custodian, having noticed my putting the garbage bag in the hallway closet instead of down the chute, later (and out of my sight) begrudgingly removed it from the closet and dropped it down the chute. I believe that my bag is in the basement because I bizarrely (and unreasonably) think that I teleported it there by placing it in the closet. My belief remains far-safe: if there was no nearby world in which the bag failed to arrive in the basement in the original scenario, nothing in these alterations changes that. But the suggestion that I know *this* proposition—that my bag is in the basement—merely in virtue of my believing it, despite my being utterly unaware that it went down the chute—is rather hard to swallow.

---

[9] Sosa, for example, takes the sensitivity theorist to task for failing to make sense of your (unknown, according to her) belief that you are not a brain in a vat in Sosa (1999a).

[10] There are issues here concerning the status of "backtracking" conditionals, the relation between subjunctive conditionals and their Lewis/Stalnaker interpretations, and related matters that I will not pause to investigate, granting instead for the sake of argument that sensitivity fails in this case.

[11] Presumably "were P true then S would believe P" admits of no more backtracking than does "were P false S would not believe P". In neither case can we alter the past fact that the bag was dropped into the chute: we can look only to worlds in which the bag was dropped when evaluating either counterfactual. So if there is a nearby world in which the bag is nevertheless not in the basement, this must be because something happened to it on the way down. Consistency, at least, appears to require this of the safety theorist who wields this example against sensitivity.

Nor is this an isolated concern. For similar reasons, the safety theorist must treat my knowing that there is not a mule cleverly disguised to look like a zebra in the zebra paddock (as per Dretske's famous example in his 1970) as far-safe. (If there were a nearby world in which it is a disguised mule I would still think that it isn't, rendering the belief unsafe.) But then even if I don't visit the zoo at all—even if I know absolutely nothing about zoos or their inclinations vis-a-vis the disguising of animals—so long as I nevertheless believe, for whatever bizarre reason, that there isn't a disguised mule in the zebra paddock of that particular zoo, I will know it. And similarly for any number of parallel, easily constructible cases.

The present concern assumes that safety is offered as sufficient for knowledge. The safety theorist might suggest instead that while safety is necessary for knowledge, it is not sufficient.[12] She might, for example, suggest that knowledge requires both that the belief be safe and that the agent possess an adequate reason for that belief. Doing so will allay the concern being canvassed. Such a view might, however, strike one as ad hoc. A natural understanding of the purpose of requiring that the agent possess an adequate reason is to eliminate a belief's being merely luckily true; that is, after all, a traditional motivation for including a justification component in the analysis of knowledge. But insofar as safety is understood—as it is by Pritchard, for example—to ensure that the belief is not true merely by luck, then such an understanding is inapplicable vis-à-vis far-safe beliefs, which are guaranteed to be non-lucky no matter what the agent's reasons might be. On the other hand, if the adequate-reason requirement is not imposed in order to eliminate luck, then it is unclear why the safety theorist would impose it (except merely as a stopgap in way of resolving this problem).[13]

My second concern in passing is that, given the dramatic shift in the requirements for knowledge and when the boundary between worlds nearby and far is crossed, one would expect a clear, non-arbitrary specification of the line separating them. But, as indicated earlier, I am aware of no such specification. This would perhaps be a more acceptable state of affairs if the boundary is vague (as indeed it intuitively is). But there is no obvious corresponding continuum in the shift in the requirements for knowledge on the safety account: if the negation of the putatively known proposition is far, mere belief is enough, whereas if it is near, much more is required. Knowledge itself might be a vague property, but if so that does not translate into vagueness in the applicability of these very different requirements for knowledge when it exists.[14]

---

[12] I will be considering this view in more detail in Sect. 8.

[13] Safety theorists are well aware of this problem. Logically necessary propositions, for example, are inevitably far-safe because there can be no nearby worlds in which the proposition is false (simply because there are no worlds, near or far, in which the proposition is false at all). As a result, some safety theorists—Pritchard, for example—restrict the equivalence of knowledge and safe belief to "fully" contingent propositions capable of being false in nearby possible worlds (see, e.g., Pritchard 2007). That restriction will not however resolve the present problem, which concerns clearly contingent propositions.

[14] Pritchard's version of safety might be employed to respond to this concern. In Pritchard (2008) he suggests that safety perhaps requires only that S not believe P in the nearest of the nearby worlds, although S must not believe P in *most* of the nearby worlds overall. Perhaps one could generalize this into a continuum by saying that the further out one travels in possible-worlds space away from the actual world, the greater the proportion of not-P worlds in which S believes P to those in which S does not believe P can be, consistently with S's knowing P (until eventually one reaches a distance at which any such proportion is so consistent).

Sosa (1999a) and others have complained that advocates of relevant alternatives theory (RAT) rarely offer a precise characterization of the distinction between relevant and irrelevant alternatives. But it would appear that the same complaint can be lodged against Sosa's own safety account.[15] This is especially so given that the separation between worlds near and far plays a similar role in the safety account to that played by the relevant/irrelevant alternatives distinction in RAT theories: both are intended to prevent distant sceptical scenarios from having an adverse impact on everyday knowledge, notwithstanding our apparent lack of evidence against them.[16]

4 Closure-failure and safety

Put those concerns aside, however, and consider the question with which we are directly concerned: does safety preserve closure?

Assume that S knows P and knows that P implies Q, believing Q as a result. Assuming also that knowledge requires safety, S's belief that P is safe (as is S's belief in the inference). If closure fails, then S's resulting belief in Q must be unsafe. If it is unsafe, the nearest not-Q world cannot lie outside the nearby-words boundary; that would ensure its far-safety. So S's belief in Q can only be unsafe because there are nearby worlds in which not-Q and S believes Q. That P implies Q requires that every world in which not-Q is one in which not-P. Therefore if there is a nearby not-Q world in which S believes Q, it is also a not-P world. Since S's belief in P is safe, S does not believe P in such a world. The question whether safety implies closure turns, then, on whether the fact that S does not believe P in every nearby world in which Q is false implies that S also does not believe Q in every such world. If it does, then safety implies closure; for there would be no nearby world in which S believes Q and Q is false, as would be required for a counterexample to closure.

One might argue for this—that S's not believing that P in nearby not-Q worlds means that S does not believe Q in those worlds—as follows. (This reasoning is my best guess as to why it is often believed that safety implies closure.) We are, after all, talking about closure, wherein S believes Q because she infers it from P. But if S believes Q because she infers it from P, then S believes P. So the counterexample must describe a nearby world in which S infers Q from P, and therefore believes P, and in which P is safe, but where S's belief in Q is not. But that's impossible. Again, if S's belief in Q is unsafe, then the world demonstrating that this is so is a nearby world in which Q is false but S believes it true. But, again, if Q is false then so is P in that

---

[15] It should be noted that Sosa no longer imposes a safety requirement on knowledge, insisting instead that the belief be "apt". See Sosa (2007). In this paper I am concerned with Sosa-as-safety-advocate.

[16] It is, moreover, not clear that the complaint succeeds against, at least, Dretske's version of RAT. On his view, the alternatives to P that are irrelevant for S's knowledge of P are simply those that don't show up in the nearest possible world in which P is false. "The thermometer is broken while reading 101 when my temperature is 98.6" is not a relevant alternative to "my temperature is 101" (as indicated by the intact thermometer) because the thermometer is not broken in the nearest world in which my temperature is not 98.6. In *that* world—the nearest world in which my temperature is not 98.6—that different temperature is reflected in the thermometer's reading. So although "my temperature is not 98.6" is relevant—on Dretske's view the negation of P itself is always relevant—the thermometer's reading puts me in a position to rule it out.

(nearby) world. If S's belief that P is safe, then S must not believe P in that world. But then she can't have inferred Q from P, which requires that she *does* believe P. But we're assuming that she did infer Q from P. So there is no such world. So S's belief in Q must be safe if her belief in P is. Q.E.D.

But this reasoning is flawed. It is true that S did *in fact* infer Q from P. But it does not follow from that actual fact that she *also* inferred Q from P in every *counterfactual* nearby world in which she believes Q. She could, consistently with her actually inferring Q from P, believe Q for other reasons entirely in those worlds. So she could believe Q for other reasons entirely in a nearby world in which Q is false while in *no* nearby world—not that one nor any other—does she believe a false P. So her belief in P could be safe while her belief in Q is unsafe; safety does not preserve closure.

For example: my Lamborghini is in parking lot space 3B and, being paranoid, I am continuously monitoring a closed-circuit monitor trained on it. It is not stolen, and I do not believe that it is. However, one car, chosen randomly, is stolen by the parking lot attendant each night (but they are never stolen by anyone else; aside from these inside jobs it's a very secure lot).[17] I have no idea that this is the case and trust the attendant implicitly, so that if I see him drive the car away on the monitor I will assume that he is just moving it to another space. A world in which the attendant steals my car is a nearby world: the world in which it is stolen is just like the actual except that the attendant randomly chose my car rather than someone else's.[18] In every nearby world in which my car is not in 3B (because the attendant stole it) I will believe that it is not in 3B, since I will see the attendant drive it away on the monitor. So my belief that it is in 3B is safe. That belief also implies that it is not stolen, which is why I actually believe that it isn't stolen. However, in a world in which the attendant does steal it I will still believe that it was not stolen (thinking that he only moved it to another space). Therefore there is a nearby world in which my car is stolen in which I believe that it is not stolen; my true belief that my car is not stolen is not safe. Closure fails.

As it happens, closure also fails in this case on the sensitivity view. My belief that my car is in space 3B is sensitive because the nearest world in which it is not in that space is one in which I don't believe that it is (thanks to the operation of the monitor). However, in the nearest world in which the proposition that it isn't stolen is false (that is, in which it is stolen), which is the world in which the attendant steals it, I still believe it is not stolen.

5 Safe-basis views and closure

Although surely a disappointment, this result might not be that much of a surprise to some advocates of safety. Indeed, Sosa concedes that safety does not preserve closure.

---

[17] Nor is there a nearby world in which the car is stolen while remaining in 3B (by some sort of illicit transfer of ownership, say).

[18] Improbability and world-remoteness are sometimes confused in the literature: that a proposition is improbable is sometimes incorrectly cited as demonstrating its remoteness. But they are not the same. Advocates of safety, moreover, need to keep them separate. For I don't know that I have won the lottery before the draw, notwithstanding the improbability of my doing so. That it is improbable had therefore better not make my winning remote, since that would make my belief that I will lose safe. (See Pritchard 2008 for discussion).

However, he thinks that a "safe-basis" or "safe-indicator" account does do so, as do Luper and Pritchard (Sosa 1999a,b; Luper 1984, 2006; Pritchard 2008). I will refer to this version as the "safe-basis" account, in contrast to the "safe-belief" account we have been examining.

According to the safe-basis account the modal property is predicated, not of the belief itself, but of the basis (or grounds, source, or reason) for the belief: if basis R of a belief that P held, then P would be true. Equivalently: the belief that P has a basis R that it would have only if true. Expressed in terms of possible worlds: There is no nearby world in which S believes that P on the basis of R and in which P is false.

It is compatible with this that there nevertheless exist nearby possible worlds in which S believes that P on a *different* basis than R, and further that P is false in some such world. So that a belief has a safe basis does not imply that the belief itself is safe.[19] It is in this way that the safety account will avoid running afoul of, for example, Nozick's grandmother case (famously also a counterexample to sensitivity). In that case a grandmother, seeing her grandson walking about hale and hearty in front of her, knows as a result that he is not ill, although if he were ill (and so out of her view in a hospital bed) her relatives would still convince her that he was well (to avoid the shock that the bad news would produce).[20] Assume that her grandson's profligate lifestyle is such that illness is a proximate threat, and her belief that her grandson is well is unsafe: there is a nearby world in which he is ill and she believes him well. However, the basis of her belief—her view of him walking about—is safe: there is no nearby world in which he is ill but she enjoys the same view.[21]

Similarly, the basis of the grandmother's belief is sensitive: in the nearest world in which he is ill, she will not believe him well on the basis of her view of him walking about (although she will believe him well). So a sensitive-basis account also avoids Nozick's grandmother case (without Nozick's infamous resort to method-relativity).

This is in fact Dretske's account. Notwithstanding his being frequently characterized as requiring that a known belief be sensitive, he requires instead that a belief that P be *caused* by a condition that is sensitive to P's holding (which is compatible with the insensitivity of the belief itself). The grandmother case (and others often cited against Nozick, such as the red barn case) demonstrates the significance of the difference.[22]

The Lamborghini case of earlier also provides an example of a belief that is unsafe but with a safe basis. My belief that my car is not stolen is, we saw, unsafe, since there is a nearby world in which the attendant stole it while I watch it all on the monitor and assume that he is only moving it to another space as a result. But notice that the actual basis for my belief that it is not stolen is different than it is in this nearby world: it is instead that the car sits in 3B, as indicated on the monitor, from which I infer that it

---

[19]   However, that a belief has an unsafe basis does imply that the belief itself is unsafe. For in a nearby world demonstrating that a belief has an unsafe basis the belief is false but has the same basis as the actual, which requires that it be falsely believed.

[20]   Nozick (1981, p. 179).

[21]   This example demonstrates that the safety theorist cannot require that a belief be both itself safe and that it have a safe basis in order to be known (which would preserve closure); the grandmother intuitively knows that her grandson is well, notwithstanding the fact that her belief to that effect is unsafe.

[22]   Dretske distinguishes his view from Nozick's with respect to this case in Dretske (2005a).

is not stolen. Since there is no nearby world in which I believe that it is in 3B when it isn't—thanks to the monitor—there is also no nearby world in which I infer from it's being in 3B that it is not stolen when it is stolen. So although my belief that it is not stolen is unsafe, it nevertheless has a safe basis. As a result, this case does not constitute a counterexample to closure on the safe-basis account.

Sosa (1999b) and Luper (2006) claim, moreover, that the safe-basis account preserves closure in general, since R safely indicates P only if R safely indicates Q, where Q is anything entailed by P. That is correct; for if R safely indicates P, then in every nearby world in which R exists P is true. But in those same worlds Q is true as well. Therefore in every nearby world in which R exists Q is true, and so R is a safe basis for belief in Q. Therefore, so long as the bases of P and Q are the same, closure is preserved.

As a demonstration that the safe-basis account preserves closure, however, this reasoning is also fallacious. Assume that the agent infers Q from P; the basis of the agent's belief in Q is then that it follows from P. But the basis of S's belief in P is *not* that it follows from P. If it is an inferential belief, it is inferred from some other proposition; if not, it is based on perception, or memory, or the like. So the actual bases of P and Q are distinct. So the fact that the safe basis relation preserves closure *in the sense that* if R is a safe basis for P then R is a safe basis for anything that follows from P (including Q) is irrelevant.[23]

There is no reason why there could not be a nearby world in which my belief in P has a different basis than it has in the actual world, and in which world P is false. (Recall that such a world is allowed by P's having a safe basis, since it's having such a basis is compatible with the belief itself being unsafe, and so with a nearby world in which it is false but I believe it anyway.) So the actual basis of my belief in P could be safe, compatibly with the existence of a nearby world in which P is false but I still believe it, albeit on a different basis than in the actual world. Further, P's being false in that world allows Q to be false in that world. And since, as we just noted, the bases of my beliefs in P and Q are different in the actual world, the basis of P could differ in that world from its basis in the actual world while the basis of Q remains the same as in the actual world. So Q could be false in that world while it has the *same* basis as in the actual world. Such a world would be one in which demonstrates that Q has an unsafe basis while P has a safe basis, and so that closure fails.

Here is an example based on the Lamborghini case. In the actual world I believe that my car is in 3B because I am watching it on the closed-circuit monitor, as before. And, as before, there is a nearby world in which the attendant steals it. However, in

---

[23] Luper presents the argument of the previous paragraph in his 2006. In that same paper, he takes Dretske to task for suggesting that the fact that perception is not closed (that one perceives that *p* and that *p* implies *q* does not require that one perceives that *q*) is evidence that knowledge is similarly not closed. As Luper points out, the principle "If S knows *p* noninferentially via perception, and S believes *q* because S knows that *q* is entailed by *p*, then S knows *q* noninferentially via perception" is false, but only trivially so; one obviously can't know something non-inferential by inference. That truism, he rightly points out, hardly counts as evidence against knowledge-closure. But by the same token S's knowledge that *q* does not derive from the same indicator as S's knowledge that *p*: *q* is known by inference, while *p* is known by perception. So the fact that a safe indicator of *p* is also a safe indicator of *q* does nothing to demonstrate that the safe-indicator account preserves closure.

that world the agent will *first* cut the wires to the monitor. I will then call him in a panic to find out what's going on, and he will reassure me (from his cell phone while joyriding in my car) that the power has blown in the parking lot but the car is still in 3B. And I believe him since I trust him implicitly. I then infer from my car's being in 3B that it is not stolen, just as I do in the actual world.

My belief that my car is in 3B is based on my perception of the image of it sitting in 3B on the closed-circuit monitor. That basis is safe: there is no nearby possible world in which the monitor produces that image and my car is not in 3B.[24] That it is sitting in 3B implies that it is not stolen. But the basis of my belief that it is not stolen is not safe: there is a world in which I believe it on the same basis, having inferred it from its sitting in 3B, but in which it is false, since the attendant is falsely assuring me that it is in 3B by phone while joyriding in my car. So closure is not preserved on the safe-basis account either.

And as before, it so happens that closure also fails on the corresponding version of the sensitivity account, the sensitive-basis account. On that account, S's knowledge that P requires that the basis of S's belief that P be sensitive, that is, that in the nearest possible world in which P is false, R does not exist. The basis of my belief that my car is in 3B is sensitive: in the nearest world in which my car is not in 3B—being the world in which the agent steals the car and cuts the wires to the monitor—I do not have the reason I do have—namely, the image on the monitor—for believing that it is in 3B. However, the basis for my belief that my car is stolen is not sensitive, so long as we assume that the nearest world in which my car is stolen is the one in which the attendant stole it. For then my reason for believing that my car is not stolen—namely, inferring it from my belief that the car is in 3B—remains.

6 Basis-individuation

The advocate of the safe-basis account might counter that the basis of the belief that the car is not stolen in the world in which the attendant steals the car after having cut the wires is not in fact the same as in the actual world. For in the actual world the basis is inference-from-the-car's-being-in-3B-*as-established-by-viewing-the-monitor,* not merely inference-from-the-car's-being-in-3B. Obviously that is not the same basis as in the counterfactual world described above (wherein I believe it is in 3B because the attendant tells me so while joyriding). In general the basis, so specified, ensures that in the nearby worlds in which the belief in Q has that basis the belief in P will have the same basis as in the actual world (simply because the basis of P is incorporated into that of Q). P will therefore be true (since that basis is safe), and so ensure that Q is true as well (since it implies Q). So the basis of the belief in Q, so specified, is guaranteed to be safe and closure is preserved.

It is difficult to say anything decisive here about which is the correct description of the basis. For we face a version of the generality problem: the basis could be

---

[24] Assume that the security is such that only the attendant has a chance of stealing it, and he wouldn't dream of attempting to monkey with the monitor to play a misleading loop of tape when he can just cut its wires and reassure me by phone.

individuated more or less finely, and so tailored to suit the purposes of the theorist. We could construe the basis of the belief in Q more broadly as "by inference", which would bring the earlier counterexample back in (since in the relevant counterfactual world the agent did infer Q), and so closure would fail. Or, we could construe the basis as "inference-from-the-car's-being-in-3B-as-established-by-viewing-the-monitor-in-the-*actual*-world", so that there can be no possible worlds at all other than the actual in which the belief in Q has the same basis, which ensures—surely too easily—that there are no such worlds in which Q is false.

Without attempting to resolve that issue in general, there are nevertheless reasons to resist the safe-basis theorist's proposed specification of the basis of the belief in Q. First, it is rather ad hoc to select a characterization of that basis which encompasses the basis of the belief in P. Why should Q's basis, which, it is agreed by all, is *different* than that of the P's, nevertheless necessarily encompass P's basis? One would, prima facie, expect their bases to be characterized independently.

Second, the proposal does not line up with the relevant speech behavior. If asked why I believe that my car isn't stolen, I will say "because it is still in 3B (and it can't be in 3B if it's stolen)." My interlocutor might *then* ask why I believe that it is in 3B, and my answer to that question will be different in the actual and counterfactual worlds. But intuitively those are (different) answers to the question why I believe *P*—that the car is in 3B—not answers to why I believe *Q*—that it isn't stolen. Intuitively, the topic has simply changed from a question concerning the basis of one belief to the basis of another.

The point can be made more generally. I believe that the current U.S. president's last name begins with a letter in the second half of the alphabet, and so (now, at least) do you. And you and I believe this for what is intuitively the same reason, namely, that "Obama" begins with "O", which is in the second half of the alphabet. Our reasons would differ if you inferred this instead from your belief that the president's name is Josaphat Xerxes. But it remains the same reason if I learned that the president's name is "Obama" by reading about it on the net while you saw it on TV. Indeed, if we insist that the basis of a belief always includes the basis of every belief it is inferred from, then even if you and I infer Q from P and P from R, so long as we believe R on different non-inferential grounds, our beliefs in Q still have different reasons. The overall result will be that it will be only very rarely correct to say that two people believe P for the same reason.

In general, we appear to individuate the basis or reason of a belief arrived at by inference by reference to the mode of inference and the proposition(s) from which the belief is inferred, but *not* also by reference to the sources of those beliefs themselves. If the safe-basis advocate who wishes to preserve closure wants to conveniently depart from this in her characterization of the cases that cause her difficulty, she needs to formulate a principled reason for doing so that doesn't obviously derive from that wish.

7 Good news, bad news

However much all of these additional counterexamples to closure might disturb the safety or safe-basis advocate, the sensitivity theorist might be pleased with the

presentation of yet more grist for her anti-closure mill. But such complacency would be misplaced. For in both versions of the Lamborghini case we have discussed, I intuitively know *both* P and Q. In each, I know that the car is in 3B in virtue of seeing it there on the monitor, and infer from its presence there that it has not been stolen. That is, intuitively, enough for knowledge that the car has not in fact been stolen, whatever the car's fortunes in non-actual possible worlds might be. So these are not just counterexamples to closure on the four views (safety, safe-basis, sensitivity, and sensitive-basis) we have discussed. They are also counterexamples to each of those views as theories of knowledge, whether or not those views affirm or deny closure.

Nevertheless, these cases do demonstrate that no advantage vis-a-vis closure is to be gained by shifting from sensitivity (or sensitive-basis) views to safety (or safe-basis) views. That supposed advantage is, I suspect, the most influential consideration among those cited in favor of safety over sensitivity. It is, in particular, a crucial component in the safety theorist's response to scepticism.

According to the safety theorist we know not only that we have hands and similar mundane facts, but also that (for example) we are not brains in vats (BIVs), which hypothesis does not therefore threaten our mundane knowledge. The disadvantage of this approach is simply that it is unintuitive that we do know that we are not BIVs—which hypothesis is, after all, expressly designed to neutralize all possible empirical evidence we have available. The benefit, however, is that we need not resort to the denial of closure in order to protect our mundane knowledge from sceptical attack in the manner of Dretske and Nozick (according to whom we know that we have hands but don't know that we are not BIVs, closure failing from the former to the latter).

The fact that closure fails on the safety account as well, however, eliminates that supposed advantage of the safety theorist's response to scepticism. This is not the place to explore these different responses to scepticism in detail. And it is certainly true that Dretske's and Nozick's views have problems of their own (most famously, perhaps, being the "abominable conjunction" problem).[25] But since the widely advertised advantage of the safety account—that it can respond to scepticism without violating closure—is accompanied by violations of closure elsewhere on that account, it is rather less clear that safety represents a significant advance over its sensitivity predecessor.

8 Safety never tried to save closure

The safety advocate might at this point simply deny that the safety condition (of either the belief itself or its basis) is intended to preserve closure, insisting that some other condition of knowledge has that job instead.[26] There are two versions of this response. On the first, safety is necessary for knowledge, but not sufficient. This is likely the most common version; Williamson, for example, explicitly disavows the project of

---

[25] The problem, named as such by Keith De Rose in his 1995, is the infelicity in asserting (for example) "I know I have hands but I don't know that I'm not a handless BIV".

[26] Thanks go to two referees whose comments prompted this section.

specifying sufficient conditions for knowledge.[27] On this version, closure is preserved because P in the counterexamples is not known, notwithstanding its being safe, since it fails to satisfy an additional closure-preserving condition.

This version, however, violates the quite robust intuition noted above to the effect that both P and Q in the counterexamples are known. Indeed, the intuition is, I suspect, especially strong that I know P in particular; surely continuous scrutiny of the closed-circuit monitor image provides me with knowledge that my car is in 3B.

Indeed, the strategy of preserving closure by denying knowledge of P ultimately leads to scepticism. The counterexamples can easily be revised with my knowledge of P based on unaided perception rather than via the closed-circuit monitor. In the first Lamborghini case, for example, suppose my (extreme) paranoia leaves me unsatisfied with the closed-circuit monitor: I now sit in the parking lot itself with the car in full view, watching it continuously. (However, when I watch the attendant drive it away in order to steal it, while assuring me he is moving it to another space, I still trust him as before.)[28] Denial of knowledge of P now requires denying knowledge of my car's presence in front of my watchful eyes. Generalized, the result is (at least) external-world scepticism.

So no response to these examples that involves denying knowledge of P will be tenable. But the attempt by the safety theorist to preserve closure by appeal to safety as a necessary but insufficient condition can only proceed by denying knowledge of P.[29] So closure cannot be recovered by construing safety as a necessary but insufficient condition of knowledge.[30]

The second version of the response under consideration avoids this difficulty. On this version, safety is sufficient for knowledge, but not necessary. While likely to be less popular among safety advocates—it is indeed unclear whether such a position would count as a within the family of safety accounts at all—such a position could be modeled on Roush's modified tracking account.[31] In way of reconciling closure with Nozick's tracking account, Roush simply counts as known both those beliefs that track the truth *and* those that are known consequents of those that track the truth. So while tracking is not a necessary condition of knowledge, the only exceptions to its being required are precisely those which ensure that closure succeeds. The safety theorist could attempt a similar move: known beliefs are those that are safe *or* the known consequents of such beliefs, whether or not those consequents are safe. On this

---

[27]   Indeed, he disavows provision of any analysis of knowledge whatsoever. Notwithstanding that stance, he does nevertheless appear committed to safety as at least a necessary condition. See Williamson (2002).

[28]   In the second version of the case, imagine that instead of cutting the cables to the vehicle before stealing it, the attendant cuts the lights in the parking lot, and proceeds to provide me with verbal assurance by cell-phone the car remains in 3B. (Assume also that I am deaf and so cannot hear the car's departure.)

[29]   Such a view cannot instead affirm knowledge of (both P and) Q since Q fails to satisfy the (on this view) necessary condition of being safe.

[30]   It might also be worth noting that the additional condition contemplated in Sect. 3—that the agent possess an adequate reason for believing the proposition, in addition to its being safe—will not help preserve closure. For in the counterexamples the agent does possess what appears to be such a reason. In the first Lambourghini case, for example, I have what appears to be a perfectly good reason to believe that the car is in 3B, namely, that I (appear to) see it on the monitor.

[31]   See footnote 2.

version, closure is preserved in the examples above because both P and Q are known notwithstanding Q's not being safe, since it is the known consequent of safe beliefs.

I leave it to the safety advocate to decide whether construing safety as an unnecessary condition for knowledge recovers enough of the safety account to be worth endorsing. At any rate the attempt to, as it were, shift the burden of sustaining closure onto another characteristic of knowledge amounts to conceding that point I aim to establish here: safety itself does not preserve closure. It cannot therefore be cited as an advantage of a safety account (over sensitivity accounts, in particular) that its adoption preserves closure where other views do not.

Moreover, even if the resulting view also provides a response to scepticism, that advantage also cannot be ascribed to the safety condition. For if closure is preserved in the account by means independent of the safety condition, that fact alone grounds a "Moorean" response to scepticism. "I am not a (handless) BIV", for example, follows from my (presumably safe) belief that I have hands. So whether or not the former anti-sceptical belief is safe, the view under consideration implies that it is known, since it is a known consequence of a known belief. The invocation of safety is therefore superfluous: the independent closure-satisfying condition grounds a Moorean response that renders the "neo-Moorean" response invoking safety otiose. (Compare Roush's account, which provides precisely this response to scepticism, while grounded in a Nozickian sensitivity account.[32])

So while the safety condition could be a sufficient but unnecessary component of an overall theory of knowledge that preserves closure, it is no longer clear that component makes any contribution worth citing over a corresponding view incorporating a sensitivity component. And if the safety advocate insists that safety be construed as at least a necessary condition of knowledge, if not a necessary and sufficient condition, and the she is also not a sceptic (as indeed none are), then closure failure is a consequence of both the safe-belief and safe-basis accounts as much as it is of the corresponding sensitivity accounts. So it remains the case that no obvious advantage is to be gained by endorsing a safety-based account over its sensitivity-based predecessors, at least with respect to the status of closure.[33]

References

DeRose, K. (1995). Solving the skeptical problem. *Philosophical Review, 104*, 1–52.
Dretske, F. (1970). Epistemic operators. *Journal of Philosophy, 67*, 1007–1023.
Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge: MIT Press.
Dretske, F. (2005a). The case against closure. In M. Sosa & E. Steup (Eds.), *Contemporary debates in epistemology* (pp. 13–26). Oxford: Blackwell.
Dretske, F. (2005b). Reply to Hawthorne. In M. Sosa & E. Steup (Eds.), *Contemporary debates in epistemology* (pp. 43–46). Oxford: Blackwell.
Hawthorne, J. (2005). *The case for closure*. In M. Sosa & E. Steup (Eds.), *Contemporary debates in epistemology* (pp. 26–43). Oxford: Blackwell.
Lewis, D. (1973). *Counterfactuals*. Cambridge: Cambridge University Press.

---

[32] See Roush (pp. 51–57).

Luper, S. (1984). The epistemic predicament: Knowledge, Nozickian tracking, and scepticism. *Australasian Journal of Philosophy, 62*(1), 26–49.

Luper, S. (2006). Dretske on knowledge closure. *Australasian Journal of Philosophy, 84*(3), 379–394.

Nozick, R. (1981). *Philosophical explanations*. Oxford: Oxford University Press.

Pritchard, D. (2005). *Epistemic luck*. Oxford: Oxford University Press.

Pritchard, D. (2007). Anti-luck epistemology. *Synthese, 158*, 277–297.

Pritchard, D. (2008). Sensitivity, safety, and anti-luck epistemology. In J. Greco (Ed.), *The oxford handbook of scepticism* (pp. 437–455). Oxford: Oxford University Press.

Pritchard, D. (2009). Safety-based epistemology: Whither now?. *Journal of Philosophical Research, 34*, 33–45.

Roush, S. (2005). *Tracking truth: Knowledge, evidence, and science*. Oxford: Clarendon Press.

Sosa, E. (1999a). How to defeat opposition to Moore. *Philosophical Perspectives, 13*, 141–154.

Sosa, E. (1999b). How must knowledge be modally related to what is known. *Philosophical Topics, 26*, 373–384.

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. 1). Oxford: Oxford University Press.

Stalnaker, R. (1968). A theory of conditionals. *American Philosophical Quarterly*, monograph no. 2: 98–112.

Stine, G. (1976). Skepticism, relevant alternatives, and deductive closure. *Philosophical Studies, 29*, 249–261.

Williamson, T. (2002). *Knowledge and its limits*. Oxford: Oxford University Press.