

Safety, sensitivity and differential support

José L. Zalabardo¹ 

Received: 28 March 2017 / Accepted: 21 November 2017 / Published online: 2 December 2017
© The Author(s) 2017. This article is an open access publication

Abstract The paper argues against Sosa’s claim that sensitivity cannot be differentially supported over safety as the right requirement for knowledge. Its main contention is that, although all sensitive beliefs that should be counted as knowledge are also safe, some insensitive true beliefs that shouldn’t be counted as knowledge are nevertheless safe.

Keywords Ernest Sosa · Epistemology · Knowledge · Skepticism · Safety · Sensitivity

1 Sensitivity and safety

In “How to Defeat Opposition to Moore”, Ernest Sosa mounted an attack on the main premise of a prominent sceptical argument. According to the sceptical premise that Sosa seeks to reject, I don’t know that I am not a brain in a vat, artificially stimulated to produce the experiences of a normal embodied human being, such as my own.

Sosa targets in particular the line of support for the sceptical premise based on the contention that my belief that I’m not a brain in a vat (\sim BIV) doesn’t have the property that has come to be known as *sensitivity*:

S’s belief that p is *sensitive* just in case, if p were false, S wouldn’t believe p .

I’m grateful to Luke Fenton-Glynn, Charlie Neil and Daniel Rothschild for their comments on this material. This research has been supported by the Ministerio de Economía y Competitividad of the Government of Spain (Projects FFI2012-38908-C02-01 and FFI2014-57409). This support is gratefully acknowledged.

✉ José L. Zalabardo
j.zalabardo@ucl.ac.uk

¹ Philosophy Department, UCL, Gower Street, London WC1E 6BT, UK

Sosa accepts that my belief in \sim BIV is not sensitive, but he takes issue with the claim that sensitivity is a necessary condition for knowledge. He advocates replacing sensitivity with its contrapositive, to which he refers as *safety*:

S's belief that p is safe just in case, if S believed p, p would be true.

My belief in \sim BIV, Sosa argues, is safe. Hence replacing sensitivity with safety as a necessary condition for knowledge would deprive the sceptic's main premise of this line of support.

Sosa sketches a line of reasoning which, if successful, would establish the conclusion that "one cannot differentially support sensitivity as the right requirement" (Sosa 1999: p. 146). According to this conclusion, every virtue that can be claimed for sensitivity as a necessary condition for knowledge would be matched by conferring this status on safety instead. My interest in this claim here is restricted to its extensional aspects—to the view that the safety condition gets the extension of the concept of knowledge right wherever the sensitivity condition gets it right. I am going to refer to this thesis by saying that safety *dominates* sensitivity, or as *domination*. Domination can be formulated as the following two claims:

D1 If a belief that has the status of knowledge is sensitive, then it is also safe.

D2 If a true belief that doesn't have the status of knowledge is insensitive, then it is also unsafe (unless it exhibits a shortcoming that would deprive it of the status of knowledge independently of its modal profile).

According to D1, the safety requirement doesn't exclude from the extension of knowledge any belief that shouldn't be excluded, unless the sensitivity requirement does so too. According to D2, the safety requirement excludes from the extension of knowledge every belief that should be excluded, unless the sensitivity requirement doesn't exclude it either. The parenthetical qualification in D2 is needed because no modal condition, including sensitivity as well as safety, can be expected on its own to be a sufficient condition for knowledge. Sensitive and/or safe true beliefs could be excluded from the extension of knowledge by other shortcomings, as, for example, the possession by the subject of strong (misleading) evidence against her belief.¹ Since this qualification will have to be accepted by the sensitivity theorist no less than by the safety theorist, we can say that if D1 and D2 hold, then, from an extensional point of view, the sensitivity requirement is no better than the safety requirement.

I'm going to argue that D1 is true, but D2 is false, and hence that domination fails. There is one respect in which the sensitivity requirement is superior to the safety requirement. Some beliefs that should be excluded from the extension of knowledge fail the sensitivity test but pass the safety test, even though they don't exhibit any non-modal shortcomings.

¹ I am indebted on this point to an anonymous referee for this journal.

2 From sensitivity to safety

In this Sect. 1 want to address the question whether there can be, contrary to D1, instances of knowledge that are sensitive but unsafe. We can see what this kind of case would look like if we translate the sensitivity and safety subjunctives to the possible-world idiom. The standard translation key goes as follows, with $p \rightarrow q$ representing the subjunctive conditional, *if p obtained, then q would obtain*:

K1: $p \rightarrow q$ is true just in case in all the worlds² in which p is true that are at the shortest distance from actuality q is also true.³

Using this key, sensitivity is rendered as:

S's belief that p is sensitive just in case, in all the worlds in which p is false that are at the shortest distance from actuality, S doesn't believe p.

Applying this translation key to the safety requirement produces an unwelcome result. It gives us:

S's belief that p is safe just in case, in all the worlds in which S believes p that are at the shortest distance from actuality, p is true.

This is not what we want. If S believes that p, then the nearest world in which S believes that p is the actual world.⁴ Hence S's belief that p will be safe just in case it is true: every true belief will be safe. This renders safety completely ineffectual as a necessary condition for knowledge.

Sosa gets around this problem by using a different translation key for safety—one that seems better suited to subjunctives with true antecedents. It requires that the consequent of the conditional should be true, not only in the actual world, but also in every world in which the antecedent is true lying no further than a certain fixed distance d from actuality:

K2: $p \rightarrow q$ is true just in case in every world in which p is true that is at a distance of d or less from actuality, q is also true.

Using this translation key, safety becomes the following non-trivial condition:

S's belief that p is safe just in case, in every world in which S believes p that is at a distance of d or less from actuality, p is true.

² My use of the plural here is meant to accommodate the possibility that there are multiple p-worlds that are equally distant from actuality, with no p-world being closer. On this point, see (Lewis 1986b: pp. 6–7).

³ Construing world similarity as a distance facilitates exposition. Similar points could be made on the more plausible construal of the notion as a partial ordering. The notion of world similarity raises important problems that might undermine its usefulness in the formulation of conditions on knowledge. See, for example, (Kment 2006). Here I'm going to leave these difficulties to one side.

⁴ See David Lewis's strong centering assumption (Lewis 1986b).

S's true belief that p will still be unsafe if there are nearby worlds (i.e. at a distance of d or less) in which S falsely believes p .⁵

We are now in a position to consider what an instance of knowledge that's sensitive but unsafe would look like. It would have to have as its content a (true) proposition that's false in worlds at a distance of d or less. In order to be sensitive, in the $\sim p$ -worlds at the shortest distance from actuality, S would not believe p , but in order to be unsafe, there would have to be other $\sim p$ -worlds, further away than the nearest ones, but still at a distance of d or less, in which S believes p .

We can easily describe true beliefs with these features. Suppose, for example, that I'm looking at a vase on a stand that could easily not be there. In normal circumstances, my true belief that there is a vase on the stand will be both sensitive and safe. It will be sensitive because in the nearest worlds in which the vase is not there I don't believe that it is, and it is safe because in all the nearby worlds in which I believe the vase is there it is actually there. But suppose now that the stand is rigged with a holographic projector connected to a thermostat in the following way: if the vase is not there and the temperature is 19C or more, it produces a perfectly convincing hologram of the vase. If the temperature is less than 19C no hologram is produced. Suppose also that, as a matter of fact, the temperature is 18C. Notice first that this circumstance wouldn't undermine the sensitivity of my belief. Assuming that the presence of the vase has no significant effect on the temperature of the room, in the nearest worlds in which there is no vase on the stand I don't believe that there is a vase on the stand, since in those worlds the temperature is 18C, as in the actual world, and hence there is no hologram. However, my belief will no longer be safe, since there are fairly close possible worlds—those in which the temperature is just a bit higher—in which I believe falsely that there's a vase on the stand, since I'm fooled by the hologram.⁶

I think that cases of this kind are clear instances of sensitive but unsafe true belief, as the notions have been formulated.⁷ However, this doesn't by itself make them counterexamples to D1. This would require, in addition, that we can intuitively recognise these cases as instances of knowledge. On this point, the examples fall short. It seems to me that in these cases our intuitions are pulled in both directions, and are not sufficiently robust to adjudicate the fate of D1.

I want to suggest, however, that the point is moot, since the kind of case under discussion is only made possible by a policy for translating subjunctives into the language of possible worlds that is highly questionable. The problem concerns the decision to use different translation keys for sensitivity and safety. This could only be

⁵ Notice that on this construal of the notion of safety, the distance d is fixed and the same for every proposition, irrespective of whether the d -sphere contains any worlds, other than the actual world, in which S believes that p . If there are no such worlds, the S's true belief that p will be (trivially) safe.

⁶ In order for the nearest worlds in which there is no vase on the stand and the temperature is 19C to be further away from actuality than the nearest worlds in which there is no vase on the stand but the temperature is 18C, in the world-similarity relation approximate similarity of particular fact has to count for something. This is allowed, but not required, by Lewis's criteria (Lewis 1986a: pp. 47–48). We could easily describe examples that aren't subject to this proviso.

⁷ This is so on the plausible assumption that the distance from actuality of the nearest worlds in which there is no vase on the stand and the temperature is 19C isn't greater than d .

justified on the grounds that, in the cases that interest us (true belief), the antecedent of the sensitivity subjunctive will be false but the antecedent of the safety subjunctive will be true. But it seems wrong to make the truth conditions of a subjunctive depend on the truth value of its antecedent. Take, for example, the subjunctive “if the currency were devalued, interest rates would go up”. On the current proposal, if the currency is actually devalued, the truth of the subjunctive will require interest-rate rises, not only in actuality, but also in the range of conditions obtaining in all the worlds at a distance of d or less in which the currency is devalued, e.g. at a range of levels of taxation, inflation, etc., but if the currency is not devalued, it would suffice for the truth of the subjunctive that interest rates go up in the specific circumstances obtaining in the nearest worlds in which the currency is devalued, e.g. with the precise levels of taxation, inflation, etc. present in those worlds. I find this counterintuitive. If, in order to avoid making safety redundant, we look at a wide range of worlds to determine the truth value of the subjunctive, we need to apply the same approach to sensitivity.

For every contingent proposition p , let $CT(p)$ denote the distance from actuality of the closest world in which p is true. With the help of this function, we can easily formulate a translation key that yields the intended results:

K3: $p \rightarrow q$ is true just in case in every world in which p is true that is at a distance of $CT(p) + d$ or less from actuality, q is also true.⁸

For subjunctives with true antecedents, K3 yields the same translations as K2, since $CT(p) = 0$ whenever p is true. Hence the notion of safety for actually obtaining beliefs remains unchanged. However, for subjunctives with false antecedents, the translations generated by K3 differ from those generated by K1. Sensitivity is now formulated as:

S 's belief that p is sensitive just in case, in every world in which p is false that is at a distance of $CT(\sim p) + d$ or less from actuality, S doesn't believe p .⁹

On this rendition, in order for S 's belief that p to be sensitive, S needs to refrain from believing p , not only in the nearest $\sim p$ -worlds, but also in every other $\sim p$ -world lying at a distance of no more than d beyond these.¹⁰ On this construal, sensitivity is harder to achieve.

⁸ This construal of subjunctives is at odds with the standard Stalnaker-Lewis analysis of counterfactuals, which makes the truth-value of a counterfactual depend only on the truth value of the consequent in the nearest worlds in which the antecedent is true, along the lines of K1 (Lewis 1973; Stalnaker 1968).

⁹ Notice that this definition has no application to necessarily true propositions. If p is necessarily true, then CT yields no value for $\sim p$. This explains why, as Nozick noticed, the sensitivity constraint doesn't exclude any belief in a necessary truth from the extension of knowledge (Nozick 1981: p. 186).

¹⁰ Robert Nozick seems to endorse this kind of construal, although his statements to this effect are less clear than one would wish. He explicitly endorses an approach along these lines in the case of adherence. He takes S 's (true) belief that p to be adherent if “[h]e believes that p for some distance out in the p neighborhood of the actual world” (Nozick 1981: p. 176). He then suggests that sensitivity is to be treated in the same way: “similarly, condition 3 [sensitivity] speaks not of the whole not- p neighborhood of the actual world, but only of the first portion of it” (Nozick 1981: p. 176). If we assume that this first portion may contain worlds that are further away from actuality than the nearest not- p worlds, then this construal of sensitivity is Nozick's.

I am claiming that this is the construal that we need to use in order to consider the relationship between safety and sensitivity.¹¹ And on this construal, there is no scope for counterexamples to D1, since sensitivity is now strictly stronger than safety.

To see this, it will help if we see both sensitivity and safety as properties that rule out nearby possible worlds in which *S* falsely believes that *p*, i.e. worlds in which *S* believes that *p* but *p* doesn't obtain. Call these worlds *error worlds*. Both sensitivity and safety impose a lower bound on how close to actuality error worlds can be found. The only difference between the two properties is that, while the lower bound imposed by sensitivity may vary with the proposition believed, safety imposes a fixed lower bound—the same for every proposition. On the one hand, *S*'s belief that *p* is sensitive just in case the nearest error world is at a distance greater than $CT(p) + d$. On the other hand, *S*'s belief that *p* is safe just in case the nearest error world is at a distance greater than *d*.

On this way of looking at things, it's easy to see why sensitivity is stronger than safety. *CT* only yields non-negative values. Hence for any proposition *p*, $CT(p) + d \geq d$. Therefore, if there are no error worlds at a distance of $CT(p) + d$ or less, as sensitivity requires, there won't be any error worlds at a distance of *d* or less, as required for safety. We can conclude that every sensitive belief is also safe.

This puts paid to any hope of finding counterexamples to D1. These would have to be beliefs with the status of knowledge that are sensitive but not safe. There can't be any of these because every sensitive belief, with or without the status of knowledge, is also safe.¹²

3 From safety to sensitivity

Let's turn now to D2—the principle that every insensitive true belief that doesn't have the status of knowledge is also unsafe. Our question is whether there are counterexamples to this principle—safe but insensitive true beliefs to which we shouldn't accord the status of knowledge. I am going to argue that cases of this kind can be easily described. To see this, notice that what's required for the safety of a belief in a true proposition *p* changes drastically depending on whether or not $CT(\sim p)$ exceeds *d*. When $CT(\sim p) \leq d$, *S*'s doxastic dispositions have to be such that in the $\sim p$ -worlds at a distance of *d* or less from actuality *S* doesn't believe *p*. If she did, there would be error worlds at a distance of *d* or less from actuality, and *S*'s belief would be unsafe. However, when $CT(\sim p) > d$, safety demands *nothing* of *S*'s doxastic dispositions. If there are no $\sim p$ -worlds at a distance of *d* or less from actuality, then a fortiori

¹¹ My claim here is that this is the best construal of the subjunctives that are widely used to formulate sensitivity and safety. However, as Sherrilyn Roush has shown, these notions can also be formulated in terms of conditional probabilities (Roush 2005). I have defended this approach in (Zalabardo 2012). For a discussion of the relationship between sensitivity and safety (and adherence) in the probabilistic setting, see (Zalabardo 2012: pp. 115–118).

¹² The outcome that matters for our purposes is that, on this score, safety is no worse than sensitivity. Every instance of knowledge that passes the sensitivity test will also pass the safety test. However, we have also established that safety is no better than sensitivity. Any true belief that shouldn't be accorded the status of knowledge but passes the sensitivity test will also pass the safety test.

there are no error worlds at a distance of d or less from actuality. Hence *any* belief in p will be safe, independently of the subject's doxastic dispositions with respect to p , and, specifically, independently of whether they render her belief sensitive.¹³ For sensitivity, unlike safety, imposes demands on the subject's doxastic dispositions independently of the value of $CT(\sim p)$. No matter how high this value might be, in order for S 's belief in p to be sensitive, her doxastic dispositions vis à vis p will have to be such that they prevent belief in p in some worlds—i.e. the $\sim p$ worlds at a distance from actuality of $CT(\sim p) + d$ or less.

Some consequences of this disparity are welcomed by Sosa. It is the reason why the sensitivity requirement excludes my belief in $\sim BIV$ from the extension of knowledge, but safety doesn't. The sheer distance from actuality of the nearest BIV -worlds—beyond any plausible value we might set for d —means that my belief in $\sim BIV$ —and indeed any belief in this proposition by an embodied subject in the kind of world we think we inhabit—will be safe, independently of my doxastic dispositions with respect to this proposition. Sensitivity, by contrast, imposes demands that are not met by my doxastic dispositions concerning $\sim BIV$. This is why replacing sensitivity with safety as a necessary condition for knowledge opens the possibility of adopting the Moorean response to the sceptical argument that Sosa recommends.

If, as Sosa maintains, I know I'm not a brain in a vat, this safe but insensitive belief of mine is not a counterexample to D2.¹⁴ I want to argue, however, that other safe but insensitive beliefs cannot be plausibly seen as instances of knowledge.

A counterexample to D2 would be a case in which a subject S believes a true proposition p and the following conditions obtain:

1. S doesn't know p
2. S 's belief that p is not sensitive
3. S 's belief that p is safe
4. S 's belief that p doesn't exhibit any non-modal shortcoming that would exclude it from the extension of knowledge

A case of this kind would have to be excluded from the extension of knowledge by a necessary condition for the instantiation of the concept, and while the sensitivity condition would achieve this, the safety condition would fail to do so.

As we pointed out just now, a sufficient condition for S 's belief that p to be safe is that $CT(\sim p)$ is greater than d . Hence we will have a counterexample to D2 if we could find a case satisfying conditions 1, 2 and 4, above, as well as:

$$3^*. CT(\sim p) > d$$

Sosa says very little about the extent of the d -sphere. All we know is that it includes worlds that could easily have been actual, and that it excludes BIV -worlds. Hence the only way we can be sure that a proposition p satisfies condition 3* is if the nearest worlds in which it's false are no closer to actuality than BIV -worlds. But this is all we need to generate counterexamples to D2. Consider the following proposition:

¹³ In (Zalabardo 2009) I introduced the notion of *freedom from the risk of error*, which allows a gradual transition between the two ways of satisfying safety.

¹⁴ I agree with Sosa on this point, but for different reasons. See (Zalabardo 2012: pp. 136–139).

BIV6: MI6 secretly keeps a collection of envatted brains, artificially stimulated to produce the experiences of normal embodied human beings.

Let's assume that BIV6 is false. I claim that it's not open to Sosa to deny that CT(BIV6) > d. The worlds in which MI6 keeps a collection of envatted brains are surely no closer to actuality than the worlds in which I am an envatted brain. The former are at least as different from the actual world as the latter, on any plausible measure of similarity. Of course some BIV-worlds are different from the actual world in pretty radical respects that are not matched by BIV6-worlds. However, there will also be BIV-worlds that are fairly similar to the actual world, except for the fact that brain envatment takes place. Presumably Sosa would want to place these BIV-worlds at a distance greater than d, since otherwise they wouldn't pass the safety test, and the switch from sensitivity to safety would be of limited help against the sceptic. But this difference from actuality is also present in BIV6-worlds. I can't see why the fact that in BIV-worlds I am the victim of this procedure should place these worlds further away than BIV6-worlds, in which others suffer this fate. It follows from this that any belief in \sim BIV6 will satisfy condition 3* and a fortiori that it will be safe. Hence, in order to get a counterexample to D2 it will suffice to find a belief in \sim BIV6 that satisfies conditions 1, 2 and 4.

\sim BIV6 is perfectly knowable. Someone with the right level of clearance will be able to know \sim BIV6. And even without this, someone with the right knowledge of what's technically possible in this area will be able to know \sim BIV6. It's even plausible to say that most of us know \sim BIV6 in this way: I know that MI6 doesn't keep envatted brains because it's technically impossible to do that. But does *everyone* who believes \sim BIV6 know this? I'm going to argue that this question should be answered in the negative.

Consider Roger, who believes \sim BIV6, but for slightly unorthodox reasons. Roger doesn't believe that brain envatment is technically impossible. In fact he believes it's a common occurrence, since he heard about Putnam's thought experiment and got the wrong end of the stick. However, he is convinced that MI6 doesn't engage in these activities, the reason is that he has a friend who tells him that he works for MI6 and is always prepared to answer his questions about the service. As it happens, Roger's friend is just a cleaner in the MI6 headquarters, with no access to any classified information, and gives random but coherent answers to Roger's questions, just to humour him. When Roger asked him if MI6 kept any envatted brains, he assured him that they didn't. It is on these grounds that Roger believes \sim BIV6.¹⁵

¹⁵ Kelly Becker has deployed similar cases in support of his claim that the safety condition must be supplemented by process reliabilism. In his first case, John has the true belief that he is not in the Matrix, but the reason he believes this is that he hates Keanu Reeves (Becker 2007: p. 73). In Becker's second case, Smith has the true belief that the Earth revolves around the sun because she is a sun worshiper (Becker 2007: p. 75). Both seem to be safe true beliefs that should be excluded from the extension of knowledge. However, Becker's cases, as they stand, cannot be invoked to undermine D2. As Becker explains, the safety theorist might try to invoke the need to relativize safety to methods in order to argue that John's belief is not safe. Alternatively, the safety theorist might concede that the belief is safe, but claim that it is excluded from the extension of knowledge by exhibiting the kind of non-modal shortcoming that D2 contemplates. Thus, for example, she might argue that beliefs are excluded from the extension of knowledge whenever a non-cognitive factor, such as John's hatred of Keanu Reeves, has played an important role in the formation of the belief, independently of the reliability of the process or the sensitivity or safety of the belief. The

It seems obvious to me that Roger's true belief in \sim BIV6 doesn't have the status of knowledge, and hence that it satisfies condition 1. Roger's friend's testimony about MI6 activities is not a source of knowledge, and Roger doesn't know any of the truths that he comes to believe on this basis. Hence Roger's true belief should be excluded from the extension of knowledge.

It doesn't seem plausible to expect that the exclusion will be effected by the kind of non-modal shortcoming contemplated in condition 4. We should be able to fill in the details of the case in such a way that Roger's belief satisfies any non-modal conditions that we might want to impose for testimonial knowledge. It follows that Roger's belief would have to be excluded from the extension of knowledge by the safety condition. However, the safety condition fails to achieve this, since Roger's belief, like any other belief in \sim BIV6, is safe. All that we need to show now, in order to have a counterexample to D2, is that the sensitivity condition succeeds where the safety condition fails—i.e. that Roger's belief is insensitive.

This can be easily shown. Consider the closest worlds (up to a distance of CT(BIV6) + d) in which MI6 does keep a collection of envatted brains. We can expect that in at least some of these worlds Roger's friend still tells him that MI6 doesn't keep envatted brains, and Roger still believes what his friend tells him. It follows that Roger's actual belief in \sim BIV6 is insensitive, as desired.

In sum, Roger's belief in \sim BIV6 is a safe but insensitive true belief to which we shouldn't accord the status of knowledge, although it doesn't exhibit any non-modal shortcomings. It follows that we have a counterexample to D2: some of the true beliefs that should be excluded from the extension of knowledge are excluded by the sensitivity condition but not by the safety condition. I conclude that Sosa's claim that safety dominates sensitivity should be rejected. Notice, however, that this is not an indictment of Sosa's safety-based account of knowledge. We have shown that safety cannot do by itself all the work that sensitivity can do in excluding from the extension of knowledge beliefs that shouldn't be there. But it is open to Sosa to claim that the job is done by safety in conjunction with other requirements, and his positive contributions to the analysis of knowledge appear to take this form.¹⁶ My claim is much more limited. Sosa's overall account may well be able to exclude from the extension of knowledge all the beliefs that would be rightly excluded by the sensitivity requirement, but contrary

Footnote 15 continued

same approach might be used for Smith's sun worship. Notice, anyway, that as Becker describes Smith's case, it is irrelevant to D2, as Smith's belief is sensitive. In the nearest worlds in which the earth doesn't revolve around the sun, she doesn't believe that it does, since "worlds in which this is false are more distant than ones in which Smith even exists" (Becker 2007: p. 75).

¹⁶ I'm not claiming either that sensitivity emerges victorious from the contest with safety. It may well be that the sensitivity condition gets the extension of knowledge wrong in ways that outweigh the kind of problem posed for the safety condition by beliefs like Roger's.

to what he claimed in 1999, the safety requirement can't do this on its own. Sensitivity *can* be differentially supported as the right requirement.¹⁷

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Becker, K. (2007). *Epistemology modalized*. New York: Routledge.
- Blome-Tillmann, M. (2017). Sensitivity actually. *Philosophy and Phenomenological Research*, 94, 606–25.
- Kment, B. (2006). Counterfactuals and explanation. *Mind*, 115, 261–310.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. (1986a). Counterfactual dependence and time's arrow. In P. Papers (Ed.), *Philosophical papers, vol. II*. Oxford: Oxford University Press.
- Lewis, D. (1986b). Counterfactuals and comparative possibility. In P. Papers (Ed.), *Philosophical papers, vol. II*. Oxford: Oxford University Press.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge: Harvard University Press.
- Roush, S. (2005). *Tracking truth*. Oxford: Oxford University Press.
- Sosa, E. (1999). How to defeat opposition to moore. In J. E. Tomberlin (Ed.), *Philosophical perspectives, 13, epistemology*. Malden, MA: Blackwell.
- Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory*. Oxford: Blackwell.
- Zalabardo, J. L. (2009). How I know I'm not a brain in a vat. *Royal Institute of Philosophy Supplement*, 64, 65–88.
- Zalabardo, J. L. (2012). *Scepticism and reliable belief*. Oxford: Oxford University Press.

¹⁷ Michael Blome-Tillmann has recently defended the superiority of safety over sensitivity by endorsing the following construal of safety:

(SAFE') S's belief that p (via method M) is safe =_{df} [S couldn't easily have formed a false belief (via M)] (Blome-Tillmann 2017: p. 614).

A simple modification of Roger's case will show that my objection to D2 cannot be sidestepped by adopting Blome-Tillmann's construal of safety. Suppose that Roger's friend is a reliable informant on all topics except on the question whether MI6 keeps envatted brains. On this question, Roger's friend believes falsely that MI6 does keep envatted brains, but for some reason he is strongly committed to ensuring that Roger doesn't find out, so he tells him, insincerely, but, as it happens, truly, that MI6 doesn't keep envatted brains, and this is so in all nearby worlds, as well as in at least some of the nearest worlds in which BIV6 is true. We now have a counterexample to D2 with safety construed as Blome-Tillmann proposes—a true belief that shouldn't be classified as knowledge and is insensitive but safe (Roger couldn't easily have formed a false belief based on his friend's testimony), while not exhibiting any non-modal shortcomings. I'm indebted on this point to an anonymous referee for this journal.