

*Luminous Margins**

Brian Weatherson

2004

Abstract: Timothy Williamson has recently argued that few mental states are luminous, meaning that to be in that state is to be in a position to know that you are in the state. His argument rests on the plausible principle that beliefs only count as knowledge if they are safely true. That is, any belief that could easily have been false is not a piece of knowledge. I argue that the form of the safety rule Williamson uses is inappropriate, and the correct safety rule might not conflict with luminosity.

1 Luminosity

In *Knowledge and Its Limits* Timothy Williamson argues that few conditions are *luminous*.¹ A condition is luminous iff we know we are in it whenever we are. Slightly more formally, Williamson defines

A condition C is defined to be *luminous* if and only if (L) holds:

(L) For every case α , if in α C obtains, then in α one is in a position to know that C obtains (95).

Intuitively, the argument against this is as follows. The following three conditions are incompatible.

Gradual Change There is a series of cases, each very similar to adjacent cases, that starts with a case where C clearly obtains, and ends with a case where C clearly doesn't obtain.

Luminosity Whenever C obtains you can know it does.

Safety Only safe beliefs count as knowledge, so whenever you can know that C obtains, C obtains in all very similar cases.

Luminosity and Safety entail

*Thanks to Tamar Szabó Gendler, John Hawthorne, Chris Hill, Ernest Sosa and the *AJP*'s referees.

¹Williamson (2000) Ch. 4; all references to this book unless otherwise specified.

Tolerance Whenever C obtains, it obtains in all very similar cases.

But Tolerance is incompatible with Gradual Change, since Tolerance entails that if the first member of the series is a case where C obtains, then every successive member is also a case where C obtains. Williamson argues that for any interesting epistemic condition, Gradual Change is a clear possibility. And he argues that Safety is a general principle about knowledge. So Luminosity must be scrapped. The counterexamples to Luminosity we get from following this proof through are always borderline cases of C obtaining. In these cases Luminosity fails because any belief that C did obtain would be unsafe, and hence not knowledge.

I will argue, following Sainsbury (1995), that Williamson has misinterpreted the requirement that knowledge be safe. The most plausible safety condition might be compatible with Gradual Change and Luminosity, if we make certain plausible assumptions about the structure of phenomenal beliefs.

One consequence of the failure of Luminosity is that a certain historically important kind foundationalist analysis of knowledge fails. This kind of foundationalist takes the foundations to be luminous. Although I think Williamson's argument against Luminosity does not work, my objections are no help to the foundationalist. As I said, my objection to Williamson rests on certain assumptions about the structure of phenomenal beliefs. It is a wide open empirical and philosophical question whether these assumptions are true. If this kind of foundationalism provided a plausible *analysis* of knowledge, then it would be a wide open question whether our purported knowledge rested on any foundations, and hence a wide open question whether we really had any knowledge. But this is a closed question. It is a Moorean fact that we know many things. So while I object to Williamson's claim that we have no luminous mental states, I do not object to the weaker claim that we *might* not have any luminous mental states, and this claim is enough to do much of the philosophical work to which Williamson puts Luminosity.

2 Williamson's Example

Williamson suggests that (L), the formal rendition of Luminosity, fails for all interesting conditions even if we restrict the quantifier to those that are 'physically and psychologically feasible' (94), and I will assume that is what we are quantifying over. To argue that (L) fails for any interesting C, Williamson first argues that it fails in a special case, when C is the condition *feeling cold*, and then argues that the conditions that lead to failure here are met for any other interesting C. So I will also focus on the special case.

Mr Davis's apartment faces southwest, so while it is often cold in the mornings it always warms up as the midday and afternoon sun streams in. This morning Mr Davis

felt cold when he awoke, but now at noon he is quite warm, almost hot. But the change from wake-up time to the present is rather gradual. Mr Davis does not take a hot bath that morning, nor cook a hot breakfast, but sits reading by the window until the sun does its daily magic. Assume, for the sake of the argument, that *feeling cold* is luminous, so whenever Mr Davis feels cold, he knows he feels cold. Williamson argues this leads to a contradiction as follows. (I've changed names and pronouns to conform with my example.)

Let t_0, t_1, \dots, t_n be a series of times at one millisecond intervals from dawn to noon. Let α_i be the case at t_i ($0 \leq i \leq n$). Consider a time t_i between t_0 and t_n , and suppose that at t_n Mr Davis knows that he feels cold. ... Now at t_{i+1} he is almost equally confident that he feels cold, by the description of the case. So if he does not feel cold at t_{i+1} , then his confidence at t_i that he feels cold is not reliably based, for his almost equal confidence on a similar basis one millisecond earlier that he felt cold was misplaced ... His confidence at t_i was reliably based in the way required for knowledge only if he feels cold at t_{i+1} . In the terminology of cases...:

(I_i) If in α_i he knows that he feels cold, then in α_{i+1} he feels cold. (97)

Given (L), all instances of (I_i), and the fact that Mr Davis feels cold when he awakes, we get the false conclusion that he now feels cold. So if we accept all instances of (I_i), we must conclude that (L) is false when C is *feeling cold* and 'one' denotes Mr Davis. Why, then, accept (I_i)? One move Williamson makes here is purely defensive. He notes that (I_i) is different from the conditionals that lead to paradox in the Sorites argument. The antecedent of (I_i) contains the modal operator *knows that* absent from its consequent, so we cannot chain together instances of (I_i) to produce an implausible conditional claim. If that operator were absent then from all the instances of (I_i) it would follow that if Mr Davis feels cold at dawn he feels cold at noon, which is false. But by strengthening the antecedent, Williamson weakens (I_i) to avoid that conclusion. But the fact that (I_i) is not paradoxical is not sufficient reason to accept it.

3 Reliability

It is useful to separate out two distinct strands in Williamson's argument for (I_i). One strand sees Williamson arguing for (I_i) by resting on the principle that beliefs constitute knowledge only if they are reliably based. The idea is that if Mr Davis's belief that he feels cold is a bit of knowledge, it is reliable, and if it is reliable it is true in all similar situations, and hence it is true in α_{i+1} . The other strand sees him appealing to a vague but

undoubtedly real requirement that beliefs must be *safely* true in order to be knowledge. Neither argument is successful, though the second kind of argument is better than the first.

Williamson acknowledges Conee and Feldman's arguments that no reliabilist epistemologist has yet solved the generality problem (100). But he takes this to be reason to abandon not the concept of reliability, but the hope of providing a reductive analysis of it. Williamson thinks we can get a long way by just resting on the intuitive concept of reliability. This seems to be a mistake. There are two ordinary ways of using 'reliable' in the context of discussing beliefs, and neither provides support for (I_i) .

First, and this is clearly not what is needed, sometimes 'reliable' just means true. This is the sense of the word in which we can consistently say, "It turned out the information that old Ronnie provided us about where the gov'nor was eating tonight was reliable, which was plenty surprising since Ronnie hadn't been right about anything since the Nixon administration." This is the sense in which reliable means just what the etymology suggests it means, something that can be relied upon. And that means, in practice, *true*. But that won't help at all, for if 'reliable' just means true, then nothing follows from the fact that knowledge is reliable that does not follow from the fact that it is factive.

Second, there is a distinctively philosophical sense in which reliable means something more like true in a wide range of circumstances. This is the sense in which a stopped clock is not even reliable twice a day. At first, this might look to help Williamson a little more. But if philosophical usage is to be key, the second look is more discouraging. For in its philosophical usage, reliability does not even entail truth. And if reliability does not entail truth in the actual situation, it surely does not entail truth in nearby situations. But Williamson's argument for (I_i) requires that reliability in α_i entails truth in α_{i+1} . So on neither of its natural readings does the concept of reliability seal the argument here, and since we have no unnatural reading to fall back upon, the argument from reliability for (I_i) fails. To be fair, by chapter 5 of Williamson's book the concept of reliability that seems to be employed is little distinguishable from the concept of safety. So let us turn to those arguments.

4 Safety

Williamson at times suggests that the core argument for (I_i) is a straight appeal to intuition. "[E]ven when we can appeal to rigorous rules, they only postpone the moment at which we must apply concepts in particular cases on the basis of good judgement. ... The argument for (I_i) appeals to such judgement." (101) The appeal to intuition is the royal road to scepticism, so we would be justified in being a little wary of it. Weinberg,

Stich, and Nichols (2001) discovered that undergraduates from the same social class as Williamson, Mr Davis and I would frequently judge that a subject could not know that mule was a mule unless he could tell it apart from a cleverly painted zebra. The judgements of that class are not obviously the basis for a sane epistemology.

Williamson undersells his argument by making it an appeal to judgement. For there is a principle here, if not a rigorous rule, that grounds the judgement. The principle is something like Ernest Sosa's safety principle. The idea is that a belief does not constitute knowledge if it is false in similar situations. "[N]ot easily would S believe that p without it being the case that p ." (Sosa 1999, 142) There is much to be said here about what is a *similar* situation. (David Lewis (1996) discusses a concept of similarity in the context of saying that worlds can be *salient*, in his sense, in virtue of being similar to salient worlds.) It might turn out that there is no account of similarity that makes it plausible that this is a constraint on knowledge. But for present purposes I am prepared to grant (a) that only safe beliefs count as knowledge, and (b) that α_{i+1} is a similar situation to α_i .

This might seem like too much of a concession to Williamson, for it already conflicts with some platitudes about knowledge. Consider a case that satisfies the following three conditions. Some light reflects off a leopard some distance away and strikes our eyes. The impact of that light causes, by the normal processes, a belief that a leopard is nearby to appear in our belief box. Beliefs, including leopard-related beliefs, that we form by this kind of process are on the whole very reliable. You might think these conditions are sufficient for our belief to count as *knowledge* that a tiger is present. The proponent of Safety denies this. She says that if, for example, there are several cheetahs with a particularly rare mutation that make the look much like leopards around, and if we saw them at similar distance we would have mistaken them for leopards. Since we could easily have had the belief that a leopard is nearby while there were no leopards, only cheetahs, nearby, the belief is not safe and so does not count as knowledge.

There are two reasons to think that safety is too strong here, neither of which strike me as completely compelling. (I'm still conceding things to Williamson here. If there's a general objection to Safety then his argument against Luminosity does not get off the ground. That's not my position. As I'll soon argue, I think Williamson has misinterpreted Safety.) The first reason is a worry that if we deny knowledge in a case of reliable veridical perception, we are conceding too much to the sceptic. But the proponent of Safety has a very good reason to distinguish this case from my current veridical perception of a table - my perception is safe and the perception of a leopard is not. So there is no slippery slope to scepticism here. The second is that the allegedly similar case is not really that similar, because in that case the belief is caused by a *cheetah*, not a *leopard*. But to regard cases where the evidence is different in this way as being dissimilar is to

make the safety condition impotent, and Sosa has shown that we need *some* version of Safety to account for our intuitions about different cases.²

So I think some version of Safety should be adopted. I don't think this gives us (I_i), for reasons related to some concerns first raised by Mark Sainsbury (1995). The role for Safety condition in a theory of knowledge is to rule out knowledge by lucky guesses. This includes lucky guesses in mathematics. If Mr Davis guesses that 193 plus 245 is 438, he does not thereby know what 193 plus 245 is. Can Safety show why this is so? Yes, but only if we phrase it in a certain way. Assume that we have a certain belief B with content p . (As it might be, Mr Davis's belief with content $193 + 245 = 438$.) Then the following two conditions both have claims to being the correct analysis of 'safe' as it appears in Safety.

Content-safety B is safe iff p is true in all similar worlds.

Belief-safety B is safe iff B is true in all similar worlds.

If we rest with content-safety, then we cannot explain why Mr Davis's lucky guess does not count as knowledge. For in all nearby worlds, the content of the belief he actually has is true. If we use belief-safety as our condition though, I think we can show why Mr Davis has not just got some mathematical knowledge. The story requires following Marian David's good advice for token physicalists and rejecting content essentialism about belief (David (2002); see also Gibbons (1993)). The part of Mr Davis's brain that currently instantiates a belief that 193 plus 245 is 438 could easily have instantiated a belief that 193 plus 245 is 338, for Mr Davis is not very good at carrying hundreds while guessing. If, as good physicalists, we identify his belief with the part of the brain that instantiates it, we get the conclusion that this very belief could have had the *false* content that 193 plus 245 is 338. So the belief is not safe, and hence it is not knowledge.

This lends some credence to the idea that it's belief-safety, not content-safety, that's the important safety criteria. When talking about Mr Davis's mathematical hunches, belief-safety is a stronger condition than content-safety. But when talking about his feelings, things may be reversed.

Let me tell you a little story about how Mr Davis's mind is instantiated. Mr Davis's phenomenal beliefs do not arise from one part of his brain, his belief box or mind's eye,

²I assume here a relatively conservative epistemological methodology, one that says we should place a high priority on having our theories agree with our intuitive judgments. I'm in favour of a more radical methodology that makes theoretical virtues as important as agreement with particular intuitions Weatherson (2003). On the radical view Safety might well be abandoned. But on that view knowledge might be merely true belief, or merely justified true belief, so the argument for Luminosity will be a non-starter. But the argument of this paper does not rest on these radical methodological principles. The position I'm defending is that, supposing a standard methodological approach, we should accept a Safety principle. But as I'll argue, the version of Safety Williamson adopts is not appropriate, and the appropriate version does not necessarily support the argument against Luminosity.

tracking another part, the part whose states constitute his feeling cold. Rather, when he is in some phenomenal state, the very same brain states constitute both the phenomena and a belief about the phenomena. Mr Davis's brain is so wired that he could not have any sensation of radiant heat (or lack thereof) without his *thereby* believing that he is having just that sensation, because he could not have felt cold without that feeling itself being a belief that he felt cold. In that case, belief-safety will not entail (I_i) . Imagine that at α_i Mr Davis feels cold, but at α_{i+1} he does not. (I assume here, with Williamson, that there is such an i .) At α_i he thereby believes that he feels cold. The content of that belief is a *de se* proposition that is false at α_{i+1} , so it violates content-safety. But in α_{t+1} that part of his brain does not constitute his feeling cold (for he does not feel cold), and thereby does not constitute his believing that he feels cold. By hypothesis, by that time no part of his brain constitutes feeling cold. So the belief in α_i that he feels cold is not false in α_{i+1} ; it either no longer exists, or now has the true content that Mr Davis does not feel cold. So belief-safety does not prevent this belief of Mr Davis's from being knowledge. And indeed, it seems rather plausible that it *is* knowledge, for he could not have had just *this* belief without it being true. This belief violates content-safety but not belief-safety, and since we have no reason to think that content-safety rather than belief-safety is the right form of the safety constraint, we have no reason to reject the intuition that this belief, this more or less infallible belief, counts as a bit of knowledge.

This story about Mr Davis's psychology might seem unbelievable, so let me clear up some details. Mr Davis has both phenomenal and judgemental beliefs about his phenomenal states. The phenomenal beliefs are present when and only when the phenomenal states are present. The judgemental beliefs are much more flexible, they are nomically independent of the phenomena they describe. The judgemental beliefs are grounded in 'inner perceptions' of his phenomenal states. The phenomenal beliefs are not, they just are the phenomenal states. The judgemental beliefs can be complex, as in a belief that *I feel cold iff it is Monday*, while the phenomenal beliefs are always simple. It is logically possible that Mr Davis be wired so that he feel cold without believing he feels cold, but it is not an accident that *he* is so wired. Most of his conspecifics are similarly set up. It is possible that at a particular time Mr Davis has *both* a phenomenal belief and a judgemental belief that he feels cold, with the beliefs being instantiated in different parts of his brain. If he has both of these beliefs in α_i , then Williamson's argument may well show that the judgemental belief does not count as knowledge, for it could be false in α_{i+1} . If he has the judgemental belief that he is not cold in α_i , then the phenomenal belief that he is cold may not be knowledge, for it is plausible that the existence of a contrary belief defeats a particular belief's claim to knowledge. But that does not mean that he is not *in a position to know* that he is cold in α_i .

Some may object that it is conceptually impossible that a brain state that instantiates

a phenomenal feel should also instantiate a belief. And it is true that Mr Davis's phenomenal states do not have some of the features that we typically associate with beliefs. These states are relatively unstructured, for example. Anyone who thinks that it is a conceptual truth that mental representations are structured like linguistic representations will think that Mr Davis could not have the phenomenal beliefs I have ascribed to him. But it is very implausible that this is a *conceptual* truth. The best arguments for the language of thought hypothesis rest on empirical facts about believers, especially the facts that mental representation is typically productive and systematic. If there are limits to how productive and systematic Mr Davis's phenomenal representations are, then it is possible that his phenomenal states are beliefs. Certainly those states are correlated with inputs (external states of affairs) and outputs (bodily movements, if not actions) to count as beliefs on some functionalist conceptions of belief.

A referee noted that we don't *need* the strong assumption that phenomenal states can be beliefs to make the argument here, though it probably is the most illumination example. Either of the following stories about Mr Davis's mind could have done. First, Mr Davis's phenomenal belief may be of the form "I feel ϕ ", where "I" and "feel" are words in Mr Davis's language of thought, and ϕ is the phenomenal state, functioning as a name for itself. As long as the belief arises whenever Mr Davis is ϕ , and it has the phenomenal state as a constituent, it can satisfy belief-safety even when content-safety fails. The second option involves some more contentious assumptions. The phenomenal belief may be of the form "I feel thus", where the demonstrative picks out the phenomenal state. As long as it is essential to the belief that it includes a demonstrative reference to that phenomenal state, it will satisfy belief-safety. This is more contentious because it might seem plausible that a particular demonstrative belief could have picked out a different state. What won't work, of course, is if the phenomenal belief is "I feel F ", where F is an attempted description of the phenomenal state. That certainly violates every kind of safety requirement. I think it is plausible that phenomenal states could be belief states, but if you do not believe that it is worth noting the argument could possibly go through without it, as illustrated in this paragraph.

Mr Davis is an interesting case because he shows just how strong a safety assumption we need to ground (I_i) . For Mr Davis is a counterexample to (I_i) , but his coldness beliefs satisfy many plausible safety-like constraints. For example, his beliefs about whether he feels cold are sensitive to whether he feels cold. Williamson (Ch. 7) shows fairly conclusively that knowledge does not entail sensitivity, so one might have thought that in interesting cases sensitivity would be too strong for what is needed, not too weak as it is here. From this it follows that any safety condition that is strictly weaker than sensitivity, such as the condition that the subject could not easily believe p and be wrong, is not sufficient to support (I_i) . Williamson slides over this point by assuming that the

subject will be almost as confident that he feels cold at α_{i+1} as he is at α_i . This is no part of the description of the case, as Mr Davis shows.

My argument above rests on the denial of content essentialism, which might look like a relatively unsafe premise. So to conclude this section, let's see how far the argument can go without that assumption. Sainsbury responds to his example, the lucky arithmetic guess, by proposing a different version of safety: mechanism-safety.

Mechanism-safety B is safe iff the mechanism that produced B produces true beliefs in all similar worlds.

I didn't want to rest on this too much because I think it's rather hard to say exactly what the mechanism is that produces Mr Davis's belief that he feels cold. But if it's just his sensory system, then I think it is clear that even at α_i , Mr Davis's belief that he feels cold satisfies mechanism-safety. The bigger point here is that content-safety is a very distinctive kind of safety claim, but it's the only kind that justifies (I_i).

5 Retractions

To close, let me stress how limited my criticisms of Williamson here are. Very briefly, the argument is that there can be some self-presenting mental states, states that are either token identical with the belief that they exist or are constituents of (the contents of) beliefs that they exist, and these beliefs will satisfy all the safety requirements we should want, even in borderline cases. If some conditions are invariably instantiated by self-presenting states, then those conditions will be luminous. And I think it is a live possibility, relative at least to the assumptions Williamson makes, that there are such self-presenting states. But there aren't very many of them. There is a reason I picked *feels cold* as my illustration. It's not laughable that it is self-presenting.

On the other hand, it is quite implausible that, say, *knowing where to buy the best Guinness* is self-presenting. And for states that are not self-presenting, I think Williamson's anti-luminosity argument works. That's because it is very plausible (a) that for a belief to be knowledge it must satisfy either belief-safety or mechanism-safety, (b) a non-self-presenting state satisfies belief-safety or mechanism-safety only if it satisfies content-safety, and (c) as Williamson showed, if beliefs about a state must satisfy content-safety to count as knowledge, then that state is not luminous. So epistemic states, like the state of knowing where to buy the best Guinness, are not luminous. That is to say, one can know where to buy the best Guinness without knowing that one knows this. And saying that (for these reasons) is to just endorse Williamson's arguments against the KK principle. Those arguments are an important

special case of the argument against luminosity, and I don't see how any of my criticisms of the general argument touch the special case.

Williamson describes his attacks on luminosity as an argument for cognitive homelessness. If a state was luminous, that state would be a cognitive home. Williamson thinks we are homeless. I think we *may* have a small home in our phenomenal states. This home is not a mansion, perhaps just a small apartment with some afternoon sun, but it may be a home.

Don't be fooled into thinking this supports any kind of foundationalism about knowledge, however. It is true that *if* we have the kind of self-presenting states that Mr Davis has (under one of the three descriptions I've offered), then we have the self-justifying beliefs that foundationalism needs to get started. But it is at best a wide-open philosophical and scientific question whether we have any such states, while it is not a wide-open question whether we have any knowledge, or any justified beliefs. If these states are the only things that could serve as foundations, it would be at least conceptually possible that we could have knowledge without self-justifying foundations. So the kind of possibility exemplified by Mr Davis cannot, on its own, prop up foundationalism.

References

- David, Marian. 2002. "Content Essentialism." *Acta Analytica* 17: 103–14. doi: 10.1007/bf03177510.
- Gibbons, John. 1993. "Identity Without Supervenience." *Philosophical Studies* 70 (1): 59–79. doi: 10.1007/bf00989662.
- Lewis, David. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74 (4): 549–67. doi: 10.1080/00048409612347521. Reprinted in his *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 1999, 418–446. References to reprint.
- Sainsbury, Mark. 1995. "Vagueness, Ignorance and Margin for Error." *British Journal for the Philosophy of Science* 46: 589–601. doi: 10.1093/bjps/46.4.589.
- Sosa, Ernest. 1999. "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13: 141–53. doi: 10.1111/0029-4624.33.s13.7.
- Weatherson, Brian. 2003. "What Good Are Counterexamples?" *Philosophical Studies* 115 (1): 1–31. doi: 10.1023/A:1024961917413.
- Weinberg, Jonathan, Stephen Stich, and Shaun Nichols. 2001. "Normativity and Epistemic Intuitions." *Philosophical Topics* 29 (1): 429–60. doi: 10.5840/philtopics2001291/217.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.

Published in *Australasian Journal of Philosophy*, 2004, pp. 373-383.