LUMINOSITY AND THE SAFETY OF KNOWLEDGE

BY

RAM NETA AND GUY ROHRBAUGH

Abstract: In his recent *Knowledge and its Limits*, Timothy Williamson argues that no non-trivial mental state is such that being in that state suffices for one to be in a position to know that one is in it. In short, there are no "luminous" mental states. His argument depends on a "safety" requirement on knowledge, that one's confident belief could not easily have been wrong if it is to count as knowledge. We argue that the safety requirement is ambiguous; on one interpretation it is obviously true but useless to his argument, and on the other interpretation it is false.

I. Introduction

It is commonly thought that we have special epistemic access to our own mental states. Few philosophers are still inclined to explain this special access in terms of infallibility, certainty, incorrigibility, or transparency, but most defend the existence of a "Cartesian core" of states that satisfy a weaker condition: if you are in some such state and you possess the requisite cognitive capacities, you are thereby in a position to know that you are in that state. Timothy Williamson (1996, 2000) has argued that even this weaker claim of privileged first-person access is false. Williamson's argument relies on an epistemological principle, that knowing requires that one could not have easily been mistaken. This "safety" requirement on knowledge is a crucial premise in many of the arguments of *Knowledge and its Limits* and leads Williamson to important conclusions about skepticism, the KK principle, and the paradox of the Surprise Examination.[1] The requirement has also been propounded by Ernest Sosa (1999, 2000, forthcoming) and appears to be gaining wider currency.

Pacific Philosophical Quarterly 85 (2004) 396–406

We argue that the safety requirement is mistaken. Although the requirement sounds plausible – even platitudinous – we identify counter-examples to it, and we explain away the intuitions that appear to support the requirement. Williamson, in contrast, has no resources for explaining away our apparent counter-examples to his thesis. Our counter-examples do not in any way rely on the intuitions that favor luminosity, so our defense of luminosity is not merely a *modus tollens* of Williamson's *modus ponens*. We aim to defend the luminosity thesis against Williamson's argument in a way that is not question-begging.[2]

II. Williamson's argument against luminosity

In order to isolate the role of safety in Williamson's argument, let us set out his argument in detail. His target is the claim that some non-trivial conditions are *luminous*, in that they satisfy the following definition:

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

Williamson claims to offer a *reductio ad absurdum* of (L) for a representative case of C, the mental condition "feels cold." Here is how the argument goes.

Consider a series of times t_0 through t_n, one millisecond apart, through which a subject S changes from feeling cold to not feeling cold, and throughout which she does whatever she is in a position to do in order to know whether she is cold. By hypothesis, then,

(1) At t_0, S feels cold.
(2) At t_n, S does not feel cold.

Now assume, for *reductio*, that feeling cold is a luminous condition (recalling S's efforts):

(3) If in α_i S feels cold, then in α_i S knows that she feels cold.

What must be true of S in order for her to know that she feels cold? On Williamson's view, at least this much: S must be confident that she feels cold and her confidence must be reliably based. Thus, if S knows at t that she feels cold, then she is confident at t that she feels cold, and her confidence is reliably based.

But what is involved in her confidence being *reliably based*? Again, at least this much: in all cases that are *sufficiently similar* to the case at t, S is confident that she feels cold only if she feels cold. It seems that we can

sidestep the question of how to specify the notion of "sufficient similarity" of cases by stipulating that the case at t+1 is as similar as we like to the case at t. At t+1, a mere millisecond later than t, S is almost as confident as she was at t, and her feeling of coldness is almost the same as it was at t, and so on. Thus, for her confidence at t that she feels cold to be reliably based, her confidence at t+1 that she feels cold at t+1 must be correct. And so:

(4) If in α_i S knows that she feels cold, then in α_{i+1} S feels cold

Together, (3) and (4) imply:

(5) If in α_i S feels cold, then in α_{i+1} S feels cold.

But we can use (5) for induction on t. Thus:

(6) At t_1, S feels cold. (from 1, 5)
(7) At t_2, S feels cold. (from 6, 5)

And so on, until we reach,

(C) At t_n, S feels cold,

which contradicts our assumption, (2). Since (1) and (2) are true by hypothesis and (4) is a instance of Williamson's safety requirement, Williamson concludes that (3) is false and, thus, that "feeling cold" is not a luminous condition. Since nothing in the argument depends upon the special features of feeling cold, no potentially variable condition – mental or otherwise – is luminous.

The weight of the argument clearly falls on (4). On Williamson's view, if S's confidence that she feels cold at t+1 is confidence in a false proposition, then it follows that her nearly indistinguishable confidence that she feels cold at t is too unreliable for her true belief at t to constitute knowledge. Williamson puts it thus:

> [I]f one believes outright to some degree that a condition C obtains, when in fact it does, and at a very slightly later time one believes outright on a very similar basis to a very slightly lower degree that C obtains, when in fact it does not, then one's earlier belief is not reliable enough to constitute knowledge (Williamson, 2000, p. 101).

Now, why does Williamson accept this claim about what is required for knowledge?

Williamson's broader idea is that the widely accepted connection between knowledge and reliability should be understood as imposing a

"safety" requirement on cases of knowledge: that one's belief could not have easily been wrong and one's confidence could not have easily been misplaced. This colloquial formulation of the requirement is, in turn, cast in the terminology of possible cases: one can know p in a case α only if in every possible case sufficiently similar to α, one confidently believes p only if p is true, where similarity is similarity in the initial conditions of the cases and context partially determines what degree of similarity counts as sufficient (Williamson, 2000, p. 124).

III. Does knowledge require safety?

So the only reason that Williamson offers for accepting premise (4) is that, for the kind of example at issue, it is an obvious specification of the general safety requirement on knowledge.[3] But is the safety requirement correct? It depends on how we are to interpret the notion of "sufficiently similar." A case β might be thought sufficiently similar to a case α in which p is true only if p is also true in β, for they are similar with respect to the truth of p. On this stringent interpretation of "sufficiently similar," Williamson's safety requirement is obviously true: in all cases that are, in this sense, sufficiently similar to α, p will also be true, and so in all such cases, S confidently believes p only if p is true. However, this trivial version of the safety requirement will not support (4). The crucial pair of cases for Williamson are these: at t one feels cold and knows it, while at t+1, one does not feel cold but confidently believes that one does. (4) denies the possibility of such a pair of cases, but their possibility is consistent with this trivial interpretation of the safety requirement. So if the safety requirement is to imply (4), it cannot be interpreted in this trivial way.

The safety requirement must, therefore, be interpreted in such a way that it does not rule out variation in the truth-value of the belief between sufficiently similar cases. So construed, however, the requirement is implausibly strong. There are pairs of possible cases which are initially similar in just about every respect except for the truth of the proposition believed, and in which my misplaced confidence in one case does not prevent me from having knowledge in the other case. We give two examples.

(A) I am drinking a glass of water which I have just poured from the bottle. Standing next to me is a happy person who has just won the lottery. Had this person lost the lottery, she would have maliciously polluted my water with a tasteless, odorless, colorless toxin. But since she won the lottery, she does no such thing. Nonetheless, she *almost* lost the lottery. Now, I drink the pure, unadulterated water and judge, truly and knowingly, that I am drinking pure, unadulterated water. But the toxin would not have flavored the water, and so had the toxin gone in, I would still

have believed falsely that I was drinking pure, unadulterated water. The actual case and the envisaged possible case are extremely similar in all past and present phenomenological and physical respects, as well as nomologically indistinguishable. (Furthermore, we can stipulate that, in each case, I am killed by a sniper a few moments after drinking the water, and so the cases do not differ in future respects.) Despite the falsity of my belief in the nearby possibility, it seems that, in the actual case, I know that I am drinking pure, unadulterated water.

(B) I am participating in a psychological experiment, in which I am to report the number of flashes I recall being shown. Before being shown the stimuli, I consume a glass of liquid at the request of the experimenter. Unbeknownst to either of us, I have been randomly assigned to the control group, and the glass contains ordinary orange juice. Other experimental groups receive juice mixed with one of a variety of chemicals which hinder the functioning of memory without a detectable phenomenological difference. I am shown seven flashes and judge, truly and knowingly, that I have been shown seven flashes. Had I been a member of one of the experimental groups to which I was almost assigned, I would have been shown only six flashes but still believed that I had been shown seven flashes due to the effects of the drug. It seems that in the actual case I know that the number of flashes is seven despite the envisaged possibility of my being wrong. And yet these possibilities are as similar in other respects as they would have to be for the experiment to be well designed and properly executed.

We take (A) and (B) to be clear examples of knowledge. But, in both cases, the subject's knowledge is not safe: there is a nearby possibility in which the subject holds the same belief but that belief is false. Readers might worry that it is not clear how to judge the proximity of a possible world to the actual world. While we sympathize with this worry, it is no more a worry for the critics of the safety requirement than for its proponents. For in determining whether or not the safety requirement captures our intuitive judgments concerning whether or not particular cases of true belief count as cases of knowledge, we must determine whether or not there is a nearby possibility of error in those cases. We have formulated cases (A) and (B) in such a way that, on any measure of proximity that is designed to capture our intuitive judgments concerning the truth-values of various counterfactuals, it seems clear that the envisaged unactualized possibilities of error will count as nearby possibilities. In section V, we return to this issue.

IV. Henry and the barns

Although we take (A) and (B) to be clear examples of knowledge, they will remind some philosophers of the "fake barns" example that Alvin

Goldman (1976) made famous. In Goldman's case, Henry sees a real barn, but this real barn is the only real barn in a countryside that is full of fake barns. Consequently, his belief that there is a real barn before him fails to count as knowledge. If Henry's situation is viewed as analogous to the situations in (A) and (B), one might conclude that (A) and (B) are not examples of knowledge and, thus, not counter-examples to Williamson's safety principle.

It seems to us that Henry's situation is *not* analogous to the situations in (A) and (B). Henry's *actual circumstances* are epistemically unfavorable in a way in which the actual circumstances of the agents of (A) and (B) are not. There really are fake barns around Henry, and this fact, combined with Henry's limited powers of fake-barn discrimination, make it difficult for him to know that he sees a real barn. In contrast, the threats to knowledge in (A) and (B) remain purely counterfactual: even though things *could* have gone epistemically less well, and almost did go epistemically less well, in point of fact, the threat was avoided and the actual case remains epistemically unproblematic. The lottery winner did not put toxin in my glass, and I did not ingest a drug which interferes with my memory. Henry's case is different. Even though he managed to end up looking at the real barn, the existence of the fake barns in his actual environment prevent him from knowing.

The principled difference between the original Henry case and cases (A) and (B) may be emphasized by modifying the original Henry case so that the threat to Henry's knowledge is purely counterfactual, like the threats in cases (A) and (B). Consider then:

(C) Henry is standing before a real barn, one of many real barns in a countryside that contains no fake barns. Indeed, let us say that no fake barns have ever been built anywhere within thousands of miles of this countryside. Nonetheless, this countryside was *almost* chosen to be the location for a movie. Had it been so chosen, all of the real barns would have been replaced with fake barns, and Henry would now be looking at a fake barn. Nonetheless, the countryside was not actually chosen as the site for the movie. It seems that Henry knows that he is standing before a real barn.[4]

Case (C) involves the nearby possibility of just the sort of situation which prevents Henry from knowing in Goldman's original case. Despite this, Henry's belief in (C) seems to be a case of knowledge. Although the (almost actualized) epistemic threat is of the same sort as that which defeats Henry's ability to know that he is looking at a real barn in the original case, the fact that the threat remains non-actual in (C) seems to make a difference.

Again, we can understand the disanalogy between the original Henry case and cases (A) and (B) by considering how we might modify (A) and (B) to make them more closely analogous to the original Henry case.

(A′) I am drinking a glass of water which I have just poured from the bottle, a bottle selected from a refrigerator full of water bottles. Unbeknownst to me, a malicious person has polluted almost every bottle in my refrigerator with a tasteless, odorless, colorless toxin. I happen to have selected the only bottle of pure, unadulterated water. I judge truly that I am drinking pure unadulterated water. Still, I would have judged the same had I selected one of the other bottles. I do not know that I am drinking pure, unadulterated water.

(B′) I am taking part in a long series of psychological experiments, in each of which I am to report the number of flashes I recall being shown to me after ingesting a glass of liquid. In this one case, I have been assigned to the control group and the liquid is ordinary orange juice. I am shown seven flashes and judge, truly, that I have been shown seven flashes. In some of the other trials in which I have participated, I have been assigned to an experimental group in which the liquid also contains a drug which interferes with memory, and the beliefs I formed on those trials were false.

The original Henry case is analogous to (A′) and (B′), but not analogous to (A) and (B). (A) and (B) are, rather, analogous to (C). Henry's inability to know in the original Henry case casts no doubt on our verdict that cases (A) and (B) are cases of knowledge.

Now, we might wonder: what is it that makes these two kinds of case epistemologically different? By virtue of what do (A), (B), and (C), on the one hand, differ from (A′), (B′), and the original Henry case, on the other? Why do the former count as cases of knowledge whereas the latter do not? We will return to these questions in section VI.

V. Context relativity

The original Henry case is germane in another way to our use of cases (A) and (B) to undermine the safety requirement. Goldman's own diagnosis of the original Henry case invokes the machinery of "relevant alternatives," and the relevance of an alternative is said to be relative to context, both the context of the epistemic subject and the context of the epistemic attributor. Williamson explicitly allows that context can influence the degree of similarity required for sufficient similarity. This raises the question: can Williamson appeal to the contextual relativity of "sufficient similarity" in order to show that (A) and (B) do not call into question the safety requirement?

No. In each of our two cases (A) and (B), there are other possible cases in which both of the following two things hold: (i) the subject's having the same belief in those possible cases is clearly relevant to whether or not the

subject's belief in the actual case is knowledge, and (ii) those very same possible cases are clearly *less similar* to the actual case than the alternative possible cases we considered in the examples above. Consider:

(A″) The actual case is just like the actual case described in (A). But now consider an alternative possibility in which my sense of taste leads me to form the belief that I am drinking pure, unadulterated water even though I am drinking raw sewage. Clearly, this possible case is much less similar to the actual case described in (A) than is the possible case in which the lottery ticket holder loses and so attempts to pollute my water. (We may assume the presence of the lottery ticket holder and her losing the lottery are constant across the two possible cases.) And yet if my sense of taste would still have led me to form the belief that I was drinking pure, unadulterated water in that more distant possible case, then my sense of taste is not trustworthy, and my gustatory belief is not knowledge.

(B″) The actual case is just like the actual case described in (B). But now consider an alternative possibility in which my memory, even without the influence of some drug, is so poor that it leads me to form the belief that there were seven flashes, even though there were only two of them. That alternative possible case is much less similar to the actual case described in (B) than is the possible case in which I am put in an experimental group and consume a drug which hinders my memory slightly. Any yet, if my memory would still have led me to form the belief that there were seven flashes in that more distant possible case, then my memory is not trustworthy and my memorial belief is not knowledge.

Even if context can alter the degree of similarity required for sufficient similarity, a stricter context cannot exclude the alternative possibilities described in (A) and (B) as irrelevant without also excluding the alternative possibilities described in (A″) and (B″) as irrelevant. But the alternative possibilities described in (A″) and (B″) are obviously relevant to whether or not the subject knows. And so Williamson cannot appeal to the contextual relativity of "sufficient similarity" to explain away apparent counterexamples like (A) and (B). It seems that Williamson has no choice then but to reject our intuitions regarding (A) and (B).

VI. Why we shouldn't expect knowledge to be safe

Examples (A), (B), and others like them in profile appear to be counterexamples to Williamson's safety requirement. Of course, if Williamson offered some principled argument for his particular version of the safety requirement on knowledge, then we would have to rethink our intuitions about the sort of examples that we have just given, but he gives no such

argument. Instead, his defense of the requirement appears to rest on the intuitive appeal of the general principle, that one can only know if one could not have been easily wrong. However, as we have pointed out, that general principle is obviously true on one interpretation, an interpretation on which it cannot play the role that Williamson needs it to play in his argument. We suggest that the plausibility of the general principle is partly due to our tendency to interpret it in this way and to confuse this trivial version of the principle with the other, substantial version which Williamson's argument requires.

Although our argument against safety appeals to intuitions about cases, Williamson cannot justly complain (as some philosophers would) about this appeal to intuitions, for Williamson himself also appeals to intuitions to support the safety requirement. He does not, however, explain away the intuitions that conflict with his thesis, e.g., intuitions about examples like (A) and (B).

We conclude that knowledge, in general, is not governed by a safety requirement of the kind that Williamson needs for his argument. Anyone should grant that knowing p in circumstances c requires that one confidently believe p in at least *some* circumstances very similar to c in which p is also true. But it is implausible to claim that knowing p in c requires that one would not have confidently believed p in even one circumstance very similar to c in which p is false – at least as long as we do not measure similarity in terms of the truth of the proposition believed.

This conclusion should not be surprising. Knowledge is an important cognitive achievement. Like other achievements worth pursuing, it must be earned and is not assured. Indeed, the most dramatic achievements are those which are earned despite substantial risk of failure. The horse which wins by a nose, the leap across a chasm which almost results in a fatal plunge, and the Nobel Prize which could easily have gone to a competitor are all achievements earned despite the nearby possibility of failure. In general, earned achievements are not safe from failure, and knowledge is no different on this score.[5] When one succeeds in forming a true belief in an epistemically respectable way, the nearby possibility of having gone wrong is not a reason to revoke the title of knowledge.

Thinking about knowledge as an achievement helps us to address the question we raised at the end of section IV, *viz.*, how should we understand the difference between (A), (B), and (C), on the one hand, and (A′), (B′), and the original Henry case, on the other? Here's our answer: In the former cases, but not the latter, the subject's belief is formed on an epistemic basis that is adequate, given his actual circumstances, to make his belief count as a particular kind of epistemic achievement. In the latter cases, his belief is *not* so formed. In the latter cases, the subject's true belief fails to count as an epistemic achievement, as opposed to mere, unearned success.

There are, of course, many different ways of trying to spell out what makes a subject's way of forming a belief epistemically adequate or inadequate in a given set of circumstances.[6] But if we think of knowledge as an achievement, then we should expect the epistemic adequacy or inadequacy of someone's belief to depend not upon what *might* have or *would* have happened had things been different, but rather upon what actually did happen. Achievements, it seems, are occurrent facts. Whether or not you win a race, for instance, is a matter of what actually happens, not of what might have or would have happened had things been different.

VII. Conclusion

Williamson's argument highlights the continuing absence of a robust positive account of the distinctive epistemological features of introspection. It also encourages us to take seriously the possibility that our acceptance of the usual assumptions about this realm are wrong. Finally, it represents a serious contribution to our attempts to understand the connection between knowledge and reliability in concrete terms. However, Williamson's specific proposal, the safety requirement, is implausibly strong when understood in a way that will serve his purposes. In its absence, we see no reason to accept (4), and so no reason to accept Williamson's argument against luminosity nor yet any reason to reject the notion of a Cartesian core of luminous mental states.[7]

Department of Philosophy
University of North Carolina
Chapel Hill

Department of Philosophy
Auburn University

NOTES

[1] C.f. Williamson 2000, chapters 5, 6, and 8, in which the safety requirement is used, respectively, to argue against the KK principle, to defuse the paradox of the Surprise Examination, and to argue that some important skeptical arguments are ill-motivated because they presuppose that one is always in a position to know what one's evidence is and whether one is rationally forming beliefs on the basis of one's evidence.

[2] After submitting this paper for review, we encountered two very useful recent papers that challenge Williamson's anti-luminosity argument: Brueckner and Fiocco (2002) and Conee (forthcoming).

[3] Earl Conee has correctly pointed out (in personal communication) that (4) could be true even if the safety requirement is false. Williamson could therefore have tried to defend (4) even without adhering to the safety requirement. While he could have tried to do this, he didn't. We stick to examining his actual argument, partly because we suspect that the safety requirement is what really animates Williamson's overall epistemological project.

[4] Lest it be objected that this alternative possibility branches from actuality too far back in the past, let's add that the movie crew can work very fast.

The same point is made by a case in which an earthquake strikes the region just as Henry comes upon it, one which *almost* collapses the backs of all the barns, leaving only facades.

[5] We further explore the connection between knowledge and achievement in our manuscript "Knowledge, Safety, and Achievement".

[6] One of us would explain this difference in terms of the defeasibility or indefeasibility of the subject's grounds for belief. See Neta (2002).

[7] We thank Tim Black, Jonathan Cohen, Earl Conee, Eric Marcus, Peter Murphy, Duncan Pritchard, Baron Reed, Matthias Steup, Matthew Weiner, and an anonymous referee for *Pacific Philosophical Quarterly* for their helpful comments and questions.

REFERENCES

Brueckner, Anthony and Fiocco, M. Oreste (2002). "Williamson's Anti-luminosity Argument," *Philosophical Studies* 110, pp. 285–293.

Conee, Earl (Forthcoming). "The Comforts of Home," *Philosophy and Phenomenological Research*.

Goldman, Alvin (1976). "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73, pp. 771–91.

Neta, Ram (2002). "S knows that p," *Noûs* 36, pp. 663–681.

Sosa, Ernest (1999). "How to Defeat Opposition to Moore," in Tomberlin (1999).

Sosa, E. (2000). "Contextualism and Skepticism," *Philosophical Issues XI: Supplement to Noûs*.

Sosa, E. (Forthcoming). "Relevant Alternatives, including Contextualism," *Philosophical Studies*.

Tomberlin, James (ed.) (1999). *Philosophical Perspectives 13: Epistemology*. Cambridge, MA and Oxford: Blackwell.

Williamson, Timothy (1996). "Cognitive Homelessness," *Journal of Philosophy* 93, pp. 554–73.

Williamson, T. (2000). *Knowledge and its Limits*. New York: Oxford University Press.