Knowledge, Safety, and Meta-Epistemic Belief

Jacob Ross

Abstract: This paper raises problems both for the view that safe belief is necessary for knowledge and for the view that it is sufficient. Focusing on 'meta-epistemic beliefs,' or beliefs about the epistemic status of one's own beliefs, it is shown that the necessity claim has counterintuitive implications and that the sufficiency claim implies a contradiction. It is then shown that meta-epistemic beliefs raise similar problems for a wide range of accounts of knowledge, and hence that they provide a powerful test for theories of knowledge.

Many philosophers have defended the following:

> ***Necessity Thesis***: Safely believing that *p* is a necessary condition for knowing that *p*.[1]

And some have defended the following:

> ***Sufficiency Thesis***: Safely believing that *p* is a sufficient condition for knowing that *p*.[2]

In this paper, I argue that *meta-epistemic beliefs*, or beliefs about the epistemic status of one's own beliefs, raise problems for both these claims. In relation to meta-epistemic beliefs, I argue, the Necessity Thesis has counterintuitive implications and the Sufficiency Thesis implies a contradiction. Moreover, I argue that such beliefs raise similar problems for a wide range of accounts of knowledge, and so they provide a powerful test for theories of knowledge.

We may say that one's belief is *safe* if and only if, in having this belief, one could not easily have been wrong. This idea is typically modeled in terms of possible worlds, as follows: *S*'s belief that *p* is safe in world *w* if and only if there is no possible world close to *w* in which *S* believes that *p* and *p* is false (Williamson, 2000, 126-27). I will adopt this standard conception of safety throughout most of what follows, though later I will consider a more sophisticated alternative.

Let's begin by considering the Necessity Thesis. To see a problem for this view, suppose there is a person named Prudence who believes the following proposition:

q: Prudence believes only what she knows.

Clearly there are possible worlds in which Prudence *falsely* believes that *q*. This would obtain in any world in which, in addition to believing that *q*, Prudence believes some other proposition that happens to be false. Furthermore, it seems intuitively that there are possible worlds in which Prudence *truly believes that p*. It seems this could obtain, for example, if initially Prudence believed only the most undeniable truths (e.g., that *1 = 1*, that *she exists*, or that *she is in pain*) and then, after carefully and rigorously reflecting on her own beliefs, she came to believe that *q*.

The two scenarios described above seem to be not only possible but *physically* possible, and realizable by physical systems consisting of finitely many molecules. And so it seems we should be able to construct a sequence of worlds, w_1 through w_n, with the following features:

(i) In all these worlds, Prudence believes that *q*;

(ii) In w_1, *q* is true;

(iii) In w_n, *q* is false;

(iv) Any two consecutive worlds in this sequence differ only with respect to a single molecule.

We can construct such a sequence by beginning with a world, w_1, where Prudence believes that q, q is true, and one of the other propositions she believes is that *she is in pain*. (To ensure that the latter proposition is knowable to Prudence, we may assume that her pain in w_1 is extreme agony, and hence far from a borderline case of pain). Then, as we move along the sequence of worlds from w_1 to w_n, we hold everything fixed except the parts of her brain responsible for pain sensations, and we modify these, one molecule at a time, until we reach a world, w_n, where Prudence is not in pain.

Now suppose, for *reductio*, that the Necessity Thesis is true. And consider a sequence of worlds, $w_1, \ldots, w_n$, satisfying conditions (i) through (iv). Let m be an integer ($2 \leq m \leq n$), and suppose that, in w_m, q is false. Since q is false in w_m, we may infer that, in w_{m-1}, Prudence's belief in q is unsafe. For, by (iv), w_m and w_{m-1} differ only with respect to a single molecule, and so they will count as close on any plausible similarity metric. Hence, w_{m-1} is close to a world w_m in which Prudence falsely believes that q, and so Prudence's belief in q is unsafe in w_{m-1}. We may therefore infer, from the Necessity Thesis, that in w_{m-1} Prudence does not know that q. And this implies that, in w_{m-1}, Prudence believes something she doesn't know. Hence, in w_{m-1}, the truth conditions for q are not satisfied, and so q is false.

Thus, from the supposition that q is false in w_m, we have derived the conclusion that q is false in w_{m-1}. We may therefore infer that for any m ($2 \leq m \leq n$), if q is false in w_m, then q is false in w_{m-1}. Hence, since (iii) implies that q is false in w_n, we may infer, by mathematical induction, that q is false in every world in the sequence, including w_1. And this contradicts (ii). And so the Necessity Thesis is inconsistent with the assumption that there is a sequence of worlds satisfying conditions (i) through (iv).

The defender of the Necessity Thesis might reply as follows:

> We should reject the assumption that there is a world, w_1, in which Prudence truly believes that *q*. For, even in the best case, where Prudence comes to believe that *q* on the basis of the most thorough investigation of her other beliefs, there is still something problematic about her belief in *q*. In order for *q* to be true, all of Prudence's beliefs must constitute knowledge, and so every proposition she believes, *including q*, must be true. Thus, the truth of *q* belongs to its own truth conditions. This makes *q* ungrounded. Hence, what we should say about *q* in the best case, w_1, is not that it is true, but rather that it lacks any determinate truth value.

This response won't save the Necessity Thesis. For, from the assumption *q* is false in w_n, and that every world in the sequence is close to its successor, we can prove, by mathematical induction, that *q* is false in every world in the sequence, including w_1. Hence, the Necessity Thesis conflicts not only with the claim that *q* is true in w_1, but also with the claim that *q* has no determinate truth value in w_1. Thus, the defender of the Necessity Thesis is committing to the counterintuitive claim that, in any world where Prudence believes that she believes only what she knows, her belief is *false*.

Let us now turn to the Sufficiency Thesis, which faces a problem that is related but more severe. To see this problem, consider a case in which someone named Humility believes exactly $n + 1$ propositions. The first *n* propositions are those belonging to some set, *B*, and the remaining proposition is one that does not belong to *B*, namely:

r: For some proposition *p* not in *B*, Humility believes, but does not know, that *p*.

The first step in my argument will be to show that this belief must be safe. Suppose, for *reductio*, that Humility's belief in *r* is not safe. There must therefore be a nearby possible world in which Humility falsely believes that *r*. Let *w* be such a world. Clearly, in *w*, *r* is *false*, since we have stipulated that Humility falsely believes that *r* in this world. But it likewise follows that, in *w*, *r* is *true*. For, since Humility falsely believes that *r* in *w*, she believes something false, which implies that she believes something she doesn't know. And since *r* does not belong to *B*, it follows that there is some proposition not in *B* (namely *r*) that Humility believes but does not know. Hence, in world *w*, the truth conditions for *r* are satisfied, and so *r* is true. Thus, the supposition that Humility's belief in *r* is unsafe implies a contradiction. And so this belief must be safe.

We can now show that the Sufficiency Thesis likewise implies a contradiction. For, since we have shown that Humility's belief in *r* is safe, it follows from the Sufficiency Thesis that Humility knows that *r*. Hence, since knowledge is factive, *r* must be *true*. However, since *r* is the only proposition not in set *B* that Humility believes, Humility's knowing that *r* implies that there is no proposition not in *B* that Humility believes but does not know. Thus, the truth conditions for *r* are not satisfied, and so *r* must be *false*. Thus, the Sufficiency Thesis implies a contradiction.

One might try to solve this problem by moving to a more sophisticated account of safety. Consider, for example, Williamson's more recent suggestion that one's belief in a proposition *p* formed on a basis *b* should count as safe iff there is no nearby world in which one falsely believes a proposition *close to p* on a basis *close to b* (Williamson 2009b, p. 325). Unfortunately, this conception of safety won't solve the problem. For we can stipulate that the basis for Humility's belief in *r* is her recognition that it is impossible for her to falsely believe that *r*. If this is the basis of her belief, then the basis is infallible, and so it seems she couldn't believe anything false on a relevantly similar basis.[3] Thus, in the version of the case just described, even the more sophisticated conception of safety will imply that Humility's belief in *r* is safe. Hence, even on this conception of safety, the Sufficiency Thesis implies that Humility knows that *r*. And this implication, as we have seen, is incoherent.

I conclude by noting that this last example raises problems, not only for safety accounts, but for a wide range of accounts of knowledge. For, in the case just described, Humility has a belief that is demonstrably true and which she believes for that very reason. Hence, on a natural understanding of justification, her belief is *justified*. And so the simple Justified True Belief account seems to imply, incoherently, that she knows that *r*. And many of the JTB+X accounts that have been proposed to solve the Gettier problem have the same implication, since, in relation to many of these accounts, Humility satisfies condition X. Her belief is *sensitive* (since she can't falsely believe that *r*, it follows that, in the nearest world where *r* is false, she doesn't believe that *r*). Her belief is *reliable*, since it is formed by a process that is guaranteed to form only true beliefs. Her belief is not based on any *false lemmas*. Her belief is *undefeated*, since there is no further truth such that Humility's knowing this truth would override or undercut her present justification for believing that *r*. And her belief is *apt*, since it is accurate in a way manifesting, or attributable to, her skill in forming such beliefs.[4] Thus, this case appears to rule out a wide range of theories of knowledge, and so it provides a powerful test for any such theory.

School of Philosophy

University of Southern California

References

Brendel, E. (2012) 'Knowledge: Safe or Virtuous?' In *Knowledge, Virtue, and Action*, ed. T. Henning and D. Schweikard (London: Routledge): 227-44.

Goldman, A. (1979) 'What is Justified Belief?' In *Justification and Knowledge*, ed. G. Pappas (Dordrecht: D. Reidel): 1-24.

Harman, G. (1973) *Thought* (Princeton: Princeton University Press).

Lehrer, K. and T. Paxson Jr. (1969) 'Knowledge: Undefeated Justified True Belief.' *Journal of Philosophy* 66: 225-237.

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

Nozick, R. (1981) *Philosophical Explanations* (Cambridge, MA: Harvard University Press).

Sosa, E. (1999) 'How to Defeat Opposition to Moore.' *Philosophical Perspectives* 13: 141-54.

----------. (2007) *A Virtue Epistemology: Apt Belief and Reflective Knowledge* (Volume I) (New York: Oxford University Press).

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

----------. (2009a) 'Probability and Danger.' *The Amherst Lecture in Philosophy* 4: 1–35. <http://www.amherstlecture.org/williamson2009/>.

----------. (2009b) 'Replies to Critics.' In *Williamson on Knowledge*, ed. P. Greenough and D. Pritchard (Oxford: Oxford University Press): 279-384.

[1] Classic defenses of the Necessity Thesis can be found in Sosa (1999), Williamson (2000), and Pritchard (2005).

[2] Many passages in Williamson suggest that he regards safe belief as both necessary and sufficient for knowledge. He says, for example, that he has 'developed and defended . . . a conception of knowledge as safety from error' (2009a, p. 9). This strong view is explicitly defended in Brendel (2012).

[3] See Williamson, 2009b, p. 325 fn. 13, where Williamson suggests that two bases that differ in their reliability don't count as close. Hence, an infallible basis will count as close only to another infallible basis.

[4] For the sensitivity condition, see Nozick (1981). For the reliability condition, see Goldman (1979). For the no false lemmas condition, see Harman (1973). For the defeasibility condition, see Lehrer and Paxson (1969). And for the aptness condition, see Sosa (2007).