AN ARGUMENT FOR THE SAFETY CONDITION ON KNOWLEDGE

Michael J. SHAFFER

ABSTRACT: this paper introduces a new argument for the safety condition on knowledge. It is based on the contention that the rejection of safety entails the rejection of the factivity condition on knowledge. But, since we should maintain factivity, we should endorse safery.

KEYWORDS: safety, knowledge, factivity

The safety condition on knowledge is a necessary condition for knowing that, recently, has been most systematically defended by Williamson, Sosa and Pritchard.[1] But it came into prominence in virtue of Nozik's analysis of knowledge, which was itself a reaction to earlier reliabilist accounts of knowledge and justification.[2] So, the safety condition is supposed to reflect the basic idea of the sort of reliability associated with bona fide knowledge that distinguishes it from mere belief and lucky true belief. The safety condition can be understood simply and informally as follows:

If *A* knows that *p*, then *A* could not easily have falsely believed that *p*.

This relatively non-technical gloss on safety and it can be made more precise as follows:

(Safety) $(w_i \vDash K_A p) \rightarrow \neg\, [\langle w_i \rangle \vDash (B_A p \,\&\, \neg p)]$.

Here '$\langle w_i \rangle$' is the set of world sufficiently close to w_i, '$K_A p$' represents *A*'s knowing that *p*, and '$B_A p$' represents *A*'s believing that *p*. So understood, the safety condition is the claim that if *A* knows that *p* at w_i, then *A* does not believe that *p* when *p* is false in worlds sufficiently similar to w_i. This regimentation captures the

[1] See Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2001), Ernest Sosa, "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999): 141-54, Duncan Pritchard, "Anti-Luck Epistemology," *Synthese* 158 (2007): 277-98, "Knowledge, Luck, and Lotteries," in *New Waves in Epistemology*, eds. Vincent Hendricks and Duncan Pritchard (London: Palgrave Macmillan, 2008), 28-51, "Safety-Based Epistemology: Whither Now?" *Journal of Philosophical Research* 34 (2009) 33-45, and *Knowledge* (London: Palgrave Macmillan, 2009).

[2] See Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981).

core idea of the safety condition well.

One main issue involved in the debate about safety is determining what worlds count as close worlds and there is considerable controversy both about how to parse closeness and whether particular accounts of the factors involved in judging closeness are intuitively supported. For the purposes of this paper this does not, however, matter. Whatever turn out to be the correct factors involved in judgments of closeness it should be clear that any such account of closeness must be reflexive, that is to say $w_i \in \langle w_i \rangle$. This is because, whatever the details involved, closeness is a similarity relation and every world is *maximally* similar to itself.

In any case, according to those who defend this condition on knowledge, safety is supposed to have independent merit as an intuitively plausible condition on knowledge. But, it would be advantageous to have a substantial argument in favor of this condition rather than having to depend on such weak and merely intuitive support for the principle and/or in light of conflicting and accounts of the closeness relation. The purpose of this paper is to provide such an argument and it is based on Kripke's recognition that safety and factivity are intimately related. Kripke made the relevant observation that is crucial to this argument in a 1986 talk in reference to Nozik's account of knowledge. In short, the argument presented here in support of safety involves the Kripke-inspired recognition that denying safety entails denying the factivity (or veridicality) condition of knowledge. It proceeds then by showing that since we should not deny factivity, we should endorse safety. Let us then look at Nozik's analysis of knowledge.

Nozik introduced the following account of knowledge as a particular form of epistemological reliabilism. *A* knowns that *p*, if and only if,

(1) *p* is true.

(2) *A* believes that *p*.

(3) If *p* weren't true, *A* wouldn't believe that *p*.

(4) If *p* were true, *A* would believe that *p*.[3]

(3) is, of course, Nozik's version of the safety condition. But, Kripke has pointed out that (2) and (3) jointly entail (1), in addition to pointing out a variety of other problems plaguing Nozik's analysis.[4] This point about the relationship between (1), (2) and (3) is particularly interesting because Kripke's observation can be

[3] Nozick, *Philosophical Explanations.*

[4] Saul Kripke, "Nozick on Knowledge," in *Saul Kripke: Collected Papers vol. 1* (Oxford: Oxford University Press, 2011), 162-224.

leveraged into a substantive argument for the safety condition on knowledge. This can be accomplished chiefly by considering what the denial of safety involves.

So what does denying safety entail? Denying safety entails this:

(Unsafe Knowledge) $(w_i \vDash K_A p)$ & $[<w_i> \vDash (B_A p \,\&\, \neg p)]$.

Knowing p at a given world is compatible with falsely believing p in worlds close to that given world. What then is the problem with respect to factivity? In order to see the problem we must have a clearer understanding of factivity in hand. The factivity condition on knowledge can be simply and informally understood as follows:

If A knows that p, then p is true.

As it is typically understood in epistemic logic, the factivity condition can then be parsed quasi-formally as follows:

(Factivity) $(w_i \vDash K_A p) \rightarrow [(w_i \vDash p)$ & $(w_j \vDash p$, for all w_j that are accessible from $w_i)]$.

To see the important implications of factivity consider the following basic model theory for standard epistemic logic. Let W be a set of worlds such that each $w_i \in W$, and R be the relation of epistemic possibility relating worlds. $<W, R>$ is then a frame in the usual sense and propositions will be subsets of W such that p is true in w_i if and only if $w_i \in p$. Let $R(w_i)$ be defined as follows: $R(w_i) = \{x \in W: R\, w_i\, x\}$. p is known at w_i then if and only if p follows from $R(w_i)$. In other words p is known at w_i if and only if p is true in all worlds that are epistemically accessible from, or are epistemic alternatives to, w_i. A world w_i is an *epistemic alternative* to world w_j for A just in case the accessibility relation holds between w_i and w_j. A bit more formally, factivity is the following condition on knowledge:

(Factivity) $(w_i \vDash K_A p) \rightarrow R(w_i) \subseteq p$.

Factivity holds in all frames in which the accessibility relation is reflexive, that is to say that factivity is an axiom of epistemic logic just in case w_i is accessible from itself. This is the case for all systems of epistemic logic at least as strong as the system KTD.

The issue then is that it should be clear that if one simultaneously accepts factivity and unsafe knowledge then one is committed to contradiction. This will be the case if there is at least one world where p is false that is close to a given world where p is known that is also an epistemic alternative to that world, and there is *always* at least one such world.[5] Consider a given proposition p known at

[5] There will actually be many such worlds.

w_1 and the definition of unsafe knowledge. Since the notion of closeness involved in the safety condition is reflexive, if *p* is known at w_1, then it can be the case that *p* is false at w_1. Why? This is simply because unsafe knowledge permits an agent to have knowledge of a proposition in a given world w_1 even when the agent falsely believes the proposition in worlds that are close to w_1. But, since closeness is reflexive, w_1 is itself one of those close worlds. So, unsafe knowledge permits an agent to know in w_1 even when the agent falsely believes the proposition in question in w_1. However, by factivity and the reflexivity of the epistemically access relation, if *p* is known at w_1 it also follows that *p* is true at w_1, since w_1 is a member of the set of worlds that are epistemically accessible from w_1. So, jointly endorsing unsafe knowledge and factivity leads to contradictions and one must go. But, since factivity is such a deeply entrenched and orthodox condition on knowledge and its denial invites all sorts of Morrean-like worries about false knowledge claims of the form "I know that *p*, but $\neg p$", we should simply treat Kripke's observation about Nozik's conditions (1), (2) and (3) as a reductio of the denial of safety and thereby as a substantive argument in favor of safety. In other words, since such Moorean "knowledge" claims clearly involve contradictions and are infellicitous we should maintain factivity and reject the denial of safety. What Kripke;s recognition allows us to see then is that arguments that support factivity are, ipso facto, arguments that support safety.