OPEN ACCESS

University of BRISTOL

Okasha, S. (2013). On a flawed argument against the KK principle. Analysis, 73(1), 80-86. 10.1093/analys/ans133

Link to published version (if available):
10.1093/analys/ans133

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research

General rights

On a flawed argument against the KK principle

Abstract

It is often claimed that externalist analyses of knowledge imply the falsity of the KK principle. However one of the standard arguments for this claim in the literature rests on a straightforward intensional fallacy. The fallacy is exposed and discussed.

Externalists in epistemology typically reject the KK principle – which says that if a subject knows that *p*, then she knows that she knows that *p*. Indeed many epistemologists write as if it is obvious that the KK principle fails on an externalist analysis of knowledge. Surprisingly however, one of the standard arguments found in the literature for this claim is fallacious. The aim of this paper is to expose and document the fallacy.

Externalism says that whether a subject's true belief constitutes knowledge depends on an 'external' condition, i.e. one to which the subject need have no epistemic access. Thus for example reliabilists analyse '*S* knows that *p*' as '(i) *S* believes that *p*; (ii) *p* is true; and (iii) *S*'s belief that *p* was formed by a reliable process'. This is a species of externalism since people rarely know or believe anything about the reliability of their belief-forming processes (and may not even possess the concept 'reliable belief-forming process'); so the reliabilist's third condition is an external one. For concreteness I focus on reliabilism, but the main points to be made apply to any externalist account of knowledge.

Since a person who believes that *p* will typically not know whether their belief that *p* was reliably formed, it *seems* tempting to infer that the KK principle must therefore fail. Numerous authors have succumbed to this temptation. Michael Williams (1995) writes:

> "For externalists, the relevant conditions must *be* fulfilled, but the believer need not...*know* (or justifiably believe) that they are. *Externalism thus drives a wedge between knowing something or other and knowing that one knows it.*" (p. 96 my emphasis)

Similarly, Fred Drestke (2004) writes:

> "modest contextualism (and, hence, externalism) provides an illuminating explanation of why KK fails. It fails because factual knowledge, according to modest contextualism, depends for its existence on circumstances of which the knower may be entirely ignorant. *So the knower can know that P without knowing (as required by KK) that he knows that P*." (p.176, my emphasis)

Alexander Bird (1998) agrees:

> "[for reliabilists] if I get knowledge by a reliable method **m** (e.g. an encyclopedia), it is required only that **m** be reliable as a matter of fact. It is not required that I know that the method is reliable...a reliable method will yield true beliefs whether or not I know that it will...*That is, I can know without knowing that I know. The K-K principle is rejected.*" (p.220, my emphasis)

Similarly Peter Pagin (2011):

> "From an internalist perspective in epistemology, the KK principle has some plausibility, but from an externalist perspective, it is rejected: that your belief that *A* is reliable does not entail that your belief *that* your belief that *A* is reliable is *itself* reliable." (p. 13, emphasis in original)

And Richard Schantz (2004):

> "[externalists] insist that unreflective first-level knowledge does not require that the believer knows anything about her own reliability...of course, if she has no beliefs on the reliability of her cognitive mechanism, she lacks something. And what she lacks, is second-level knowledge or justification: she does not know, or justifiably believe, that her belief is justified or amounts to knowledge. *No wonder then that externalists generally dismiss the KK-thesis*" (p. 9, my emphasis).

And finally David Hemp (2006):

> "It is also natural for externalists to reject [the KK principle]...This can be seen more clearly by focusing on the reliabilist theory of knowledge. If one's belief that p is produced by a reliable process that one knows nothing about, then one may have no way of knowing that this belief constitutes knowledge, *and thus no way of knowing that one knows that p*." (my emphasis)

The above authors all think that from externalism, plus the fact that people typically do not know or believe that their belief-forming mechanisms are reliable, the falsity of the KK principle follows. But to reason this way is to commit an intensional fallacy. In effect, the above authors reason as follows:

1. Kp (assumption)

2. $\neg K[Bp$ is reliable$]$ (assumption)

3. $Kp \equiv [Bp \ \& \ p$ is true $\& \ Bp$ is reliable$]$ (externalism)

4. $Kp \Rightarrow KKp$ (KK principle, assumed for *reductio*)

5. KKp (from 1, 4)

6. $K[Bp \ \& \ p$ is true $\& \ Bp$ is reliable$]$ (from 3, 5)

7. $K[Bp$ is reliable$]$ (from 6, assuming knowledge distributes across conjunction)

8. $\neg[Kp \Rightarrow KKp]$ (from 1, 2, 3, 4, 7, by *RAA*)

This argument *seems* to show that on an externalist account of knowledge, a person who knows that p (premise 1), but who does not know that their belief that p was reliably formed (premise 2) does not know that they know that p – and thus that the KK principle fails.[1] But the argument is fallacious: to derive (6) from (3) and (5) involves substitution within an intensional context. Grant the externalist the equivalence of Kp and $[Bp \ \& \ p$ is true $\& \ Bp$ is reliable$]$ as in (3); it does *not* follow that someone who knows that they know that p, i.e. KKp, must then know the conjunction $[Bp \ \& \ p$ is true $\& \ Bp$ is reliable$]$, nor therefore that they must know the third conjunct.[2] To think otherwise is to commit an intensional fallacy. So the fact that a person who knows that p will typically not know (or even believe) that their belief that p was reliably formed, has no bearing on the question of whether they know that they know that p, even granting externalism.

---

[1] P. Pagin's argument, quoted in the text, is a slight variant. Pagin argues that to know that one knows that p, one's belief that one's belief that p is reliable would have to be reliable. So Pagin moves from KKp to $[B[Bp$ is reliable$]$ is reliable$]$. This involves a two-step inference, firstly from KKp to $K[Bp$ is reliable$]$ as above, and from there to $[B[Bp$ is reliable$]$ is reliable$]$, via a second application of the externalist definition of knowledge. The first step involves the intensional fallacy.

[2] This assumes that the reliabilist's definition of knowledge is meant as a statement of necessary co-extension, not of intensional equivalence. This is how reliabilism is standardly understood, by its proponents and opponents; see Shapiro (2006) for discussion of this point.

The fallacy above involves substituting the externalist definition of knowledge for the second occurrence of 'K' in the expression 'KK*p*'. It is of course legitimate to make this substitution for the *first* occurrence of 'K'. That is, from the equivalence K*p* ≡ [B*p* & *p* is true & B*p* is reliable], and from KK*p*, we can validly infer B[K*p*] & K*p* is true & B[K*p*] is reliably formed. This shows us the real requirements for knowing that one knows that *p*, by externalist lights: one must believe that one knows that *p*, actually know that *p*, and one's belief that one knows that *p* must be reliably formed. Whether this last condition obtains will depend on the reliability of the mechanisms by which one's beliefs about one's own epistemic states are generated. It may be that these mechanisms are by and large reliable, i.e. only tend to produce the belief that one knows that *p* when one actually does know that *p*, in which case knowing that one knows will be commonplace, or it may not. But this issue is wholly independent of whether one knows or believes anything about the reliability of one's belief that *p*. Someone entirely ignorant of whether their belief that *p* is or is not reliably formed, may have a reliably-formed belief that they know that *p*, or they may not. In short, knowing that you know that *p*, for an externalist, does not require knowing (or believing) that your first-order belief that *p* is reliably formed; rather it requires having a true and reliably-formed *higher-order* belief that you know that *p*.

This shows that there is no immediate inference from externalism to the rejection of the KK principle, contrary to what is often thought. The core externalist idea – that for a true belief to constitute knowledge an 'external' condition must be met – has no direct bearing on whether a subject knows that they know that *p*, when they know that *p*. Parallel considerations show that there is no immediate inference from internalism to the KK principle. Internalists hold that for a true belief to constitute knowledge, an 'internal' condition must be met, i.e. one to which the subject has epistemic access. Thus a typical internalist analysis of 'S knows that *p*' is '(i) S believes that *p*; (ii) *p* is true; (iii) S has good reason to believe *p*', where it is assumed that a subject can always tell whether she has good reasons for her beliefs. Armed with this analysis, it may seem tempting to argue from K*p* to KK*p*, as follows:

1. K*p*   (assumption)
2. K*p* ≡ [B*p* & *p* is true & good reasons for B*p*]   (internalism)
3. B*p*   (from 1, 2)
4. K[B*p*]   (from 3, assuming one knows

		what one's beliefs are)
5.	K[good reasons for B*p*]	(epistemic access to reasons)
6.	K[B*p* & *p* is true & good reasons for B*p*]	(from 1, 4, 5, assuming knowledge is closed under conjunction)
7.	KK*p*	(from 2,6)

Of course, the inference from (3) to (4) could be queried on the grounds that we do not always know what our beliefs are, and the closure assumption used to derive (6) could also be queried. But even granting these assumptions the argument fails, since the inference from (2) and (6) to (7) commits an intensional fallacy. If a subject knows that they believe *p*, that *p* is true, and that they have good reasons for believing *p*, it does *not* follow that they know that they know that *p*, even granting the internalist their definition of knowledge.

That one standard argument from externalism to the rejection of the KK principle fails leaves open, obviously, that there might be a better argument to the same effect. Indeed Tim Williamson (2002) has presented a quite different, sophisticated argument against the KK principle from externalist premises, that he dubs the 'anti-luminosity' argument. Interestingly, in a footnote Williamson (2005) briefly contrasts his anti-luminosity argument with "the simple-minded idea that externalist accounts of knowledge have us know without knowing that we know; the latter idea is often based on an illicitly internalist understanding of 'know' when the sentence in its scope includes epistemic vocabulary" (p.231 n.14). Here Williamson alludes to another flawed argument in the literature from externalism to the failure of the KK principle, based on 'illicit internalism'. It is worth asking how this argument relates to the flawed argument based on the intensional fallacy discussed above.

Williamson is referring, I think, to the following line of argument.[3] "Knowing that *p* requires having a reliably-formed belief that *p*. But people typically don't have any information about the reliability of their belief-forming mechanisms, so a person who knows that *p* won't have any reason to think that their belief that *p* is reliably formed, so won't know that they know that *p*." Williamson's complaint is that this argument rests on an internalist reading of the outer knowledge operator in KK*p*; to argue from someone not having any reason to believe *x* to their not knowing *x* is to assume internalism. A thoroughgoing externalist must allow that

---

[3] Williamson gives no citations, but I think he is correct that this 'illicit second-order internalism' is often found in the literature.

someone can know that they know that p whether or not they have reason to think that their belief that p is reliably formed.

Williamson is right that the line of argument sketched above rests on illicit second-order internalism. But importantly, the argument *also* commits the intensional fallacy. Fully spelled out, the argument runs:

1. Kp (assumption)
2. $\neg$ Reason to believe [Bp is reliable] (assumption)
3. $\neg K$[Bp is reliable] (from 2)
4. $Kp \equiv$ [Bp & p is true & Bp is reliable] (externalism)
5. $Kp \Rightarrow KKp$ (KK principle, assumed for *reductio*)
6. KKp (from 1, 5)
7. K[Bp & p is true & Bp is reliable] (from 4, 6)
8. K[Bp is reliable] (from 7, assuming knowledge distributes across conjunction)
9. $\neg[Kp \Rightarrow KKp]$ (from 1, 3, 4, 5, 8, by *RAA*)

This argument tries to get from externalism to the falsity of the KK principle, via the assumption that a person who knows that p (premise 1) will not have reason to believe that their belief that p is reliably formed (premise 2). But the argument is doubly fallacious: the inference from (2) to (3) is guilty of the 'illicit second-order internalism' of which Williamson complains, while the inference from (4) and (6) to (7) commits the intensional fallacy. Note that illicit internalism alone is not sufficient to yield the desired conclusion; the intensional fallacy is needed too.

This doubly fallacious argument against the KK principle differs from the singly fallacious argument with which we began in just one way. Instead of simply positing that a person who knows that p may not know that their belief that p is reliably formed (i.e. $\neg K$[Bp is reliable]), the doubly fallacious argument tries to deduce this from the assumption that the person lacks reason to believe that their belief that p is reliably formed. But this additional step is unnecessary as well as illegitimate. Since knowledge requires belief, a person who doesn't believe that their belief that p is reliably formed (e.g. because they lack the concept

'reliable belief-forming mechanism') will not know that their belief that *p* is reliably formed. Since it is presumably quite common for people to lack beliefs about the reliability of their belief-forming mechanisms, the step in the argument which requires illicit internalism is dispensable. However Williamson may be right that as a matter of fact, those who have taken externalism to imply the immediate failure of the KK principle, such as the authors quoted above, have been guilty of illicit second-order internalism.[4] However the deeper problem with their argument is that it rests on an intensional fallacy.

References

Bird, A. (1998) *Philosophy of Science*, London: Routledge.

Dretske, F. (2004) 'Externalism and Modest Contextualism', *Erkenntnis* 61, 173-186.

Hemp, D. (2006) 'The KK (Knowing That One Knows) Principle', *Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/kk-princ/.

Pagin, P. (2011) 'Assertion, Inference and Consequence', *Synthese* online first, http://dx.doi.org/10.1007/s11229-011-9906-8, pp. 1-17.

Schantz, R. (2004) 'Introduction', *The Externalist Challenge*, Berlin: de Gruyter, 1-36.

Shapiro, S. (2006) 'Externalism, Anti-realism and the KK Thesis', in D. Devidi and T. Kenyon (eds.) *A Logical Approach to Philosophy*, Dordrecht: Springer, 22-35

Williams, M. (1995) *Unnatural Doubts: Epistemological Realism and the Basis of Skepticism*, Princeton: Princeton University Press.

Williamson, T. (2002) *Knowledge and its Limits*, Oxford: Oxford University Press.

Williamson, T. (2005) 'Contextualism, Subjective-Sensitive Invariantism and Knowledge of Knowledge', *Philosophical Quarterly*, 55: 213–235.

---

[4] Of the authors quoted above, Dretske, Pagin and Schantz appear guilty of illicit second-order internalism.