

# 'KNOWLEDGE IS TRUE BELIEF' REBUTTED

SCOTT F. AIKIN

Vanderbilt University

## ABSTRACT

Crispin Sartwell has argued that knowledge is merely true belief. The two arguments for this thesis are (1) from counter-examples to third requirements for knowledge and (2) from a dilemma for justification-theorists. I will show that the purported counter-examples are inconclusive because they do not reflect an informal pragmatic element of knowledge-attribution. The dilemma is inconclusive, because one horn is easily graspable. Further, I will refine Lycan's argument that the thesis that knowledge is true belief is inconsistent with epistemic modesty.

**Keywords:** knowledge, justification, epistemic minimalism, Crispin Sartwell, William Lycan

## 1.

Crispin Sartwell, among others,<sup>1</sup> has argued that knowledge is merely true belief. Call this the TB thesis, or TB for short. I will argue here that Sartwell's two arguments for TB are inconclusive and, further, that the endorsement of TB is inconsistent with intellectual modesty and the attitude of epistemic fallibilism. This yields a case against TB.

I will address here the arguments for TB from Crispin Sartwell's "Knowledge is Merely True Belief" (1991) and "Why Knowledge is Merely True Belief" (1992), though I believe the arguments here are relevant considerations for other cases for TB. Sartwell's first argument for TB comes from a set of examples of knowledge-attribution, and his second argument proceeds as a dilemma for non-minimalist theories of knowledge. Sartwell frames his case by examples where it seems legitimate to reject the demand for justification, cases where the requirement is excessive. The cases can be paraphrased as follows:

**Flash of insight:** A is struggling with some math problem then has a 'flash of insight' as to the solution. Given that A's been

1 See, for example, Jakko Hintikka (1962), Alvin Goldman, who holds that there is *a sense* of knowledge that is true belief, (1999 and 2002), and David Martens (2005).

working on the problem, she can't see the proof, just the solution. *She knows the solution.*

**Religious Faith:** B has faith in a beneficent God. The issue of God's existence and beneficence is a rationally undecidable matter, but such a God exists. *B knows there is a loving God.*

**Mundane Faith:** C's son is charged with committing a grisly murder. The evidence against the son is piling up, but C holds to the belief in the son's innocence. The son is innocent. *C knows he is innocent.*

**Knew it All Along:** P is some proposition about which D could not have any information before time t1. D, however, believed that p well before t1 – ever since t(1-n). D has the truth about p now at t1, and may legitimately claim that *she knew it all along.* (Adapted from 1991, 160-1)

The point, Sartwell claims, is that in every case presented there are subjects who know because they merely have true beliefs. Justification is unnecessary. Sartwell concludes modestly:

Of course, even if we admit such cases, they would establish only that justification is not a necessary condition for knowledge; more would need to be said in order to do more than merely suggest that true belief is sufficient... I am not attempting a demonstration, but offering a plea for the suspension of burden of proof. (1991, 60)

The burden of proof, as Sartwell takes it now, is symmetric – the examples above are enough to undercut the presumptive status of the falsity of TB. The dialectical burdens, then, are shared on both sides of the debate. On the assumption that Sartwell correctly diagnoses the dialectical situation, what is necessary here are two arguments in response to Sartwell: one to address the undercutting that Sartwell takes his examples to do, and another to make the positive case against TB.

Sartwell's dilemma for those committed to the J-condition runs as follows. Epistemic justification's normativity is either instrumental for knowledge or not. If justification is instrumental (namely, conceived in terms of truth-conducivity or a means to attaining true beliefs), then subjects pursue justification with regard to their beliefs not because of justification's intrinsic worth, but *as a means* for the beliefs depending on those reasons to be true. Justification serves a pragmatic or instrumental purpose, then, for picking out cases of knowing or convincing others that you know. But it itself is not constitutive of knowledge. It is only a means for having knowledge or showing that you know. Sartwell reasons: "Epistemic justification is therefore in the final analysis only an instrumental value, not an intrinsic one" (1991, 161). Sartwell then frames his case that justification, so considered, cannot be part of knowledge:

[P]hilosophers (committed to the instrumental value of justification) *have*

committed themselves implicitly to the view that knowledge is merely true belief, and that justification is merely a *criterion* rather than a logically necessary condition for knowledge... Justification, on the present view is ... a means by which we achieve knowledge ... it provides a test of whether someone has knowledge, that is, whether her beliefs are true. (1992, 173-4, emphasis in original)

TB, then, follows. If, on the other hand, justification is not valued as merely instrumental to achieving true belief, then it must be valuable as some intrinsic good, perhaps as satisfying a demand of rationality or self-possession. Knowledge, then, is the achievement of *two* goals of cognitive life. If so, then there is always the possibility that in fulfilling one goal, one may not fulfill the other. That is, one may have justification and not truth, and *vice versa*. And so, justification, pursued on its own (assuming fallibilism is true for justification) may lead *away* from truth. Sartwell notes, “[N]ow that we recognize two primitive epistemic values, they may well conflict” (1992, 177). Since there are intelligible cases where the two can fail to be “realized simultaneously”, the J-requirement pits two cognitive norms (justification and truth) against one another (1992, 180). That gives us two potentially conflicting goals of inquiry, which makes knowledge an “incoherent notion” (1992, 180). TB follows.

I will first respond serially to Sartwell’s counter-examples. Second, I will argue that the second horn of Sartwell’s dilemma is graspable. And third, I will offer an argument that TB is an uncomfortable thought for anyone with fallibilist inclinations.

## 2.

Sartwell is right that A knows in **Flash of Insight**, but he is wrong that it is a counter-example to the J-condition. What makes the case compelling is that A is a mathematician (or at least a math student) with understanding of the problem and training with thinking about similar issues. A isn’t, say, a toddler ignorant of math altogether and merely dreams up the solution when looking at his older sister’s math notebook. This is important, because A’s expertise here is a causal condition for A to have the ‘aha’ moment. In such cases, it seems plausible to say A’s belief, from A’s own conscious perspective, is spontaneous, but was nevertheless a consequence of some set of *unconscious inferences*.<sup>2</sup> A, later, may be able to reconstruct how she saw the solution by talking about the things she was committed to and that some connection dawned for her. Further, the belief may be non-doxastically justified. A may *intuitively see* that some proposition is right, but not now have the means to prove it correct. If A’s belief is based on that intuition, it seems the belief could be justified. So far, the case with A is consistent with traditional theories of knowledge.

The cases with B, C, and D just don’t seem like cases of knowledge, so I don’t see how they are counter-examples. Here’s an analogy why. Let Carl go to Las Vegas. Carl *loves*

---

<sup>2</sup> Sartwell, importantly, has what seems an inferential requirement for belief-formation, as he is clear that guesses are not beliefs (1991, 159).

the number three. Carl plays Keno, and bets his house on three. He fully believes that three will win. He does so out of his love for or faith in three and three's power. Three ends up winning, but Carl didn't *know* that three would come up. Carl's reasons were *terrible*, and that seems to defeat his claim that he knows. That is, justification can show others that you know not just by showing them that the belief is true (as Sartwell correctly argues), but giving justifications can show that you know in that you *understand how* it is that what you believe is true. If Carl has no plausible understanding of how 3 would win, he doesn't know that it would, even if he was right. That seems intuitive enough.

We might, however, *let Carl get away with saying he knew* that three would come up. We would do so not because he would be right, but because it would be *conversationally costly* to correct him. Knowledge attributions like these have informalities because we don't want to get too technical or snippy with people when it doesn't matter one way or the other whether or not we allow them to claim they knew. With Carl and gambling, only the truth matters for the payout. The pit boss isn't going to quibble that we're stretching the meaning of 'knows' when we allow Carl to say he knew. He will just count out the chips. And the same (at least for the examples) goes for the B, the theist, C, the trusting parent, and D, who "knew it all along." When it's only the truth that counts, we don't haggle over the further messy details of knowledge-attribution. In these cases, we may be *pragmatically warranted* in attributing knowledge, despite the fact that it's *false* that the people in these cases actually know. Here, the thought is that, all things considered, it may be reasonable to let some people get away with claiming they know when they did, in fact, not know. The truth of such claims is most certainly important to us, but there may be cases where that value can be trumped by something of greater value – perhaps a friendship, maintaining good will, and in the case of the card dealer, keeping his job.

But change the case where it is important that we have justification, not just truth. Let's have S3, the trusting parent, claiming that the son is innocent. In one context, say as sympathetic friends, we may attribute knowledge to C. We may share C's confidence in the son, and by this, we could reasonably attribute knowledge to C. For us and for the parent, only his being innocent is important. But in another context, say as the parents of the murder victim or members of a jury, we would want also to know *why* C believes the son's innocent. Sartwell, here, makes the argument that these considerations make justification instrumental in legitimating knowledge-claims, not in constituting them. Surely, the instrumentality is right, but it doesn't follow that they aren't constitutive. Precisely because if C believes on the basis of *terrible reasons* (e.g., on the basis of some clearly irrelevant piece of evidence or on the basis of mere wishful thinking), we would not attribute knowledge to C *because C did not know*.<sup>3</sup> And often, if we aren't C's friends or family, we'll be more willing to pay the costs for saying what's true.

It has been widely noted that the conditions for warranted knowledge-attribution change with context, and here, standards for allowable knowledge-attribution change according to what's at stake. In cases where, once the belief is true, justification is

<sup>3</sup> Duncan Pritchard also pushes this *terrible reasons* line in 2004, 110.

beside the point from our interests, it is informally allowable for people to attribute knowledge to people without justifying reasons. But that does not make those attributions strictly correct. Go back to Carl. The Keno dealer, once it's clear that Carl has won, can reasonably let Carl claim that he *knew* three would win. That's because he has no real reason (beyond the technical reason) to correct him. But Carl bet the house. Surely, say, *Carl's wife* would want more than the fact that he won to call it a case of *knowledge*. The point here is that in the cases Sartwell proposes, we are tempted to conflate the pragmatics of knowledge-attribution with the semantics of knowledge conditions.<sup>4</sup>

So the story with S2 would go that so long as the God in question isn't an evidentialist and the conditions for successful belief in Him aren't evidential in nature, then it would seem reasonable to attribute knowledge to B. But if the evidence is also important, then B couldn't be reasonably said to know. Take D, who 'knew it all along' also. We generally allow people to attribute knowledge in hindsight to themselves; however, it is clear that a good deal of such knowledge-attributions are actually distortions of the epistemic situation. There is a demonstrable tendency for subjects to exhibit what is called 'hindsight bias' on behalf of their present epistemically better-off beliefs when assessing their previous beliefs: they systematically attribute these evidentially better off views to themselves in the past, so they, even when they had no idea at the time, assess themselves as having 'known it all along.'<sup>5</sup> Consequently, the varieties of knowledge-attribution and their inherent distortions require of us a measure of tolerance for analytically incorrect but contextually appropriate claims (in these cases, just having true belief) to knowledge.

The point here is that Sartwell's counter-examples are cases where, if all we care about is the truth of the beliefs in question, we don't bother about the details of justification for knowledge-attribution. C's sympathetic friends and Carl's Keno dealer don't correct these cases because it's costly to do so. In cases where the costs of wrong future outcome outweigh the conversational costs of correction (the jury and Carl's wife), it is reasonable to correct those attributions. Sartwell is right that there is a pragmatic element to this aspect of knowledge attribution, but he's wrong that it counts in his favor.

### 3.

The second argument for minimalism is Sartwell's dilemma. The first horn of the dilemma, I think, is right – that if we think justification is merely statistical or modal, then the J-condition is either merely internal to the T-condition or entirely detachable from knowledge. I think that horn is sharp, and I will leave defenses of externalism to externalists.<sup>6</sup> But Sartwell is wrong about the second horn. He argues that because

---

<sup>4</sup> See LeMorvan (2002,154-5) for a similar argument that the pragmatics-semantics distinction is not respected by the Sartwell counter-examples.

<sup>5</sup> See Frischhoff (2007) and Mazzoni and Vannucci (2007) for reviews of the various distortions possible in the general class of 'hindsight bias.' Thanks to Rachel L. Phillips for this point.

<sup>6</sup> See LeMorvan 2002 for an account of the instrumentality of justification as still intrinsic to knowledge and Hofman's response in 2005.

there are “two primitive epistemic values,” they “may well conflict” (1992, 177). He reasons as follows:

[I]f justification is itself proposed as an intrinsic goal (a demand for reason, for example), then knowledge is an incoherent notion. It gives us two goals for inquiry, which cannot always be realized simultaneously. (1992, 180)

But now the question is how knowledge is *incoherent* when two goals of cognitive life may not always be mutually instantiable? We want to have the truth and have it in the right way, and there may be cases when we have one or the other but not both. How does *that* make knowledge inconsistent, much less incoherent? A few analogies can help clarify this. Being a good sport, I think, requires the joint successes of playing well and playing by the rules. But, surely, one might play better by bending the rules sometimes. It may also be possible to play by the rules, but nevertheless play horribly. The detachability of these two goals of being a good sport from each other does not make the concept of being a good sport incoherent. Rather, that the two are not always mutually instantiable helps explain why there are so few good sports and correlately why being a good sport is something one has to really work hard to be. Another analogy: Living well requires that we enjoy ourselves and be moral, but sometimes it's a drag to do the right thing. The fact that two goals of a good life are separable and may not be always jointly instantiable does not make the goal of a good life incoherent.<sup>7</sup> This tenuousness, instead, helps make intelligible what the successes are in living well, and how hard it can be sometimes to balance these goods. As a corollary, with inquiry as the pursuit of knowledge, the idea that one can have justification but not truth is the central thought behind fallibilism: that a subject may be justified in a belief but not correct. How else might fallibilism be intelligible unless the two notions were *not* co-extensive? But when one knows, one has satisfied two demands of cognitive life: have the truth and be rational. There are two points in the offing here. First is that the fact that some notion may have two necessary conditions that are not co-extensive does not make that notion incoherent. The second is that the fact of non-coextensive necessary conditions for being a good sport, living well, and knowing, can help explain why these are real achievements. So the second horn of Sartwell's dilemma is not sharp at all.

4.

I will close here by refining what I take to be the most direct case against the TB thesis. Holding TB commits subjects to an attitude of infallibility with regard to their beliefs. William Lycan (1994) has responded to Sartwell's position by pointing to its being self-refuting. Though Lycan's argument is on the right track, it is incomplete.<sup>8</sup> Instead of showing that *TB* is self-refuting, the argument shows that a commitment to *TB* is inconsistent with intellectual modesty or fallibilism. Lycan's argument runs:

<sup>7</sup> In fact, a *tu quoque* seems reasonable here, since Sartwell's notion of justification would (if internal to the T-condition) not be able to account for justified but widespread false beliefs (Latus 2000, 38). That conception of justification seems positively incoherent.

<sup>8</sup> Lycan's argument has been criticized by Skidmore (1997, 124), LeMorvan (2002, 165) and Martens (2006, 16)

- (1) Sartwell believes TB.
- (2) TB is true (assumed for *reductio*).
- (3) If TB is true, then (if some S believes p and p is true, then S knows that p).
- (4) If (1) and (2), then Sartwell knows TB. (3)
- (5) Sartwell knows TB. (1, 2, 4)
- (6) If any subject S believes p and S is aware that S believes p, then S believes that S believes p truly.
- (7) Sartwell is aware that he believes TB.
- (8) Sartwell believes that he believes TB truly. (1, 6, 4)
- (9) If any subject S believes p and S knows that q is equivalent to p by analysis, then S believes q. (substitutivity of known analytic entailments in belief-contexts)
- (10) Sartwell believes he knows TB.

Lycan holds we should balk at (10), since it is highly unlikely that anyone *knows* highly controversial philosophical claims to be true, and it is unlikely that serious philosophers are so arrogant to think they are exceptions. The argument generalizes to yield:

- (11) For any proposition p that Sartwell believes and is aware he believes, Sartwell believes he knows p.

(11) is something we should definitely balk at, since it entails that we eschew all intellectual modesty. Any belief Sartwell has, no matter how fleeting or flighty is one that he, on TB and his belief the TB is true, must take himself to know. Every belief. That doesn't seem right. Lycan reasons, then, that we should reject (11). So we should reject (2) by *reductio*.

The problem is that the *reductio* is incomplete, because (2) is introduced into an already inconsistent set. What makes (11) false is the following premise:

- (12) Sartwell has at least one belief that he is aware he has but he does not believe is a case of knowledge.

The problem is that (1), (3), (6), (9), and (12) are already inconsistent. We can see this because (11) follows from (1), (3), (6) and (9) without the need of introducing (2). Note that the only inference that (2) plays a role in is to (5), but (5) is a dead end in the proof. The trouble, then, is that the proof can't be a *reductio* of (2), because the set of premisses it's introduced into is already inconsistent.

The Lycan self-refutation argument isn't successful. However a weaker conclusion is still right – namely that (1), (3), (6), (9), and (12) are inconsistent. By this, we have a

reason to reject (1). Neither Sartwell nor anyone else, for that matter, can believe TB and remain epistemically modest. Given that we are and ought to be epistemically modest, even about TB, we cannot be committed to TB. Unless the argument for TB is coupled with an argument against epistemic modesty, we cannot maintain a commitment to it. The argument extends to TB's falsity as follows. It is a reasonable presumption behind epistemology that getting clearer about the concept of knowledge would make us epistemically less vicious. Certainly, if we have the right conception of knowledge, we shouldn't be made by that conception more vicious. A further thought is that having the right conception of knowledge should improve and not make worse our knowledge-assessments. TB is a conception of knowledge, and believing it (with the first step in the reductio above) makes us epistemically vicious. Further, given that if we believe TB, we will attribute knowledge to ourselves more widely than we would if we required justification, TB will induce a wide variety of false knowledge-attributions (as first-personal knowledge attribution would be co-extensive with first-personal belief-attribution), even on its own analysis. TB, then, cannot be true of knowledge.

## REFERENCES

- Alston, W. 1986. Concepts of Epistemic Justification. In *Empirical Knowledge*, ed. Moser, P. Rowman and Littlefield. 23-54. Originally published in *The Monist*. 1985. 68: 57-89.
- BonJour, L. 1985. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard UP.
- Fischhoff, B. 2007. An Early History of Hindsight Research. *Social Cognition* 25(1): 10-13.
- Goldman, A. 1986. *Epistemology and Cognition*. Cambridge, MA: Harvard UP.
- Goldman, A. 1999. *Knowledge in a Social World*. Oxford University Press.
- Goldman, A. 2002. *Pathways to Knowledge*. Oxford University Press.
- Hintikka, J. 1996. *Knowledge and Belief*. Ithaca, NY: Cornell University Press.
- Hofman, F. 2005. Epistemic Means and Ends: In Defense of Some Sartwellian Insights. *Synthese* 146: 357-69.
- Latus, A. 2000. Our Epistemic Goal. *Australasian Journal of Philosophy* 78: 28-39.
- Lehrer, K. 1974. *Knowledge*. New York: Oxford UP.
- Le Morvan, P. 2002. Is Mere True Belief Knowledge? *Erkenntnis* 56: 151-68.
- Le Morvan, P. 2006. Goldman on Knowledge as True Belief. *Erkenntnis* 62: 145-55.
- Lycan, W. G. 1994. Sartwell's Minimalist Analysis of Knowing. *Philosophical Studies* 73: 1-3.
- Mazzoni, G. and Vannucci, M. 2007. Hindsight Bias, the Misinformation Effect, and False Autobiographical Memories. *Social Cognition* 27(1): 203-20.
- Maitzen, S. 1995. Our Errant Epistemic Aim. *Philosophy and Phenomenological Research* 55(4): 869-76.
- Martens, D. 2006. Confidence in Unwarranted Knowledge. *Erkenntnis* 65: 141-64.
- Moser, P. 1985. *Empirical Justification*. Dordrecht: Reidel.
- Pritchard, D. 2004. Epistemic Deflationism. *Southern Journal of Philosophy* 42: 103-34.



- Sartwell, C. 1991. Knowledge is Merely True Belief. *American Philosophical Quarterly* 28:2, 157-65.
- Sartwell, C. 1992. Why Knowledge is Merely True Belief. *The Journal of Philosophy* 89:4, 167-80.
- Skidmore, A. 1997. On Sartwell's Thesis that Knowledge is Merely True Belief. *Southwest Philosophy Review* 13: 123-7.

---

Received: October 15, 2009

Accepted: March 7, 2010

111 Furman Hall  
Vanderbilt University  
Nashville, TN 37240  
U.S.A.  
scott.aikin@Vanderbilt.Edu