Notes on Warfield's "Knowledge from Falsehood"

03/09/10

Branden Fitelson

1 Some Background on Inferential Knowledge

There is a "Received View" (RV) about inferential knowledge.[1] (RV) involves various tenets, including:

- **Foundationalism**. All inferential knowledge rests (ultimately) on a foundation of basic *knowledge* (*i.e.*, all inferential knowledge "traces back" *via some epistemic chain* back to *the foundation*). I will also assume that **Foundationalism** entails the following about the nature of epistemic/inferential chains: (*i*) that there are *no infinite epistemic chains*, and (*ii*) that there are *no circular epistemic chains*.

- **Preservation** (*a.k.a.*, **Closure**). If S knows that p, and S competently deduces q from p (while maintaining their belief in p), then S knows that q.

- **Conservation**. If S does *not* know that p, and S competently deduces q from p (while maintaining their belief in p), and the *only* epistemic grounds S has for believing q is p, then S does *not* know q.

Warfield is interested in *arguing against* (RV). Actually, Warfield rejects *all three* of the above tenets of (RV). In his paper "Knowledge from Falsehood", he's taking aim primarily at **Conservation**. In fact, Warfield claims to have come up with examples of inferential knowledge that is based on a *false* premise. I will reconstruct the dialectic slightly differently than he does. I will (ultimately) take the target to be **Conservation**.

2 Warfield's Example

Warfield gives various examples which he claims are examples of "knowledge from falsehood". I will focus on the following example, and various alternative renditions thereof:

> I have a 7pm meeting and extreme confidence in the accuracy of my fancy watch. Having lost track of the time and wanting to arrive on time for the meeting, I look carefully at my watch. I reason: 'It is exactly 3pm; therefore I am not late for my 7pm meeting'. I know my conclusion, but as it happens it's exactly 2:58pm, not 3:00pm.

Just to make the case a bit clearer, I will work with the following propositions:

(p) It is exactly 3:00pm.

(q) It is earlier than 7:00pm.

And, I will characterize the example in the following way:

- I do *not* know that p (since p is false).

- I competently deduce q from p (while maintaining my belief in p).

(†) My *sole* epistemic grounds/basis for believing that q is my belief that p.

Therefore, if I *do* know that q in this case, then this example is a *counterexample* to **Conservation**. Note: if Warfield is right, then this is more than *merely* a counterexample to **Conservation** — it's a case of *inferential knowledge on the basis of a **falsehood***. Ultimately, though, it is **Conservation** that is the real target. Next, I'll discuss some "resistance" strategies for (RV)-defenders. All resistance strategies will take issue with (†).

[1] See Chapter 6 of Audi's *Epistemology* (now on our website) for a nice introduction to The Received View.

3 Defending (RV) from Warfield's Example

Defenders of (RV) will want to resist Warfield's "knowledge from falsehood" (KFF) claim, and even my weaker "knowledge from non-knowledge" (KFNK) claim. Warfield discusses some attempts to defend (RV) against (KFF). I don't think these attempts are fully compelling. Here's what Warfield says about "the resistance":

> The suggestion is that one has knowledge despite the presence of an involved falsehood if there is a justified and (at least) dispositionally believed truth entailed [or epistemically supported] by the falsehood that serves as the premise in one's inferential argument. ...I evaluate (and reject) this proposal...Grant ...the assumption that in every counterexample case we will find dispositional beliefs entailed by the involved falsehood or epistemically supported by it. Both proposals fail catastrophically in Gettier cases. I seem to see a dog in the yard. On this occasion I form the belief that there is a dog in the yard and then reason as before to the conclusion that there is at least one animal in the yard. My belief is false (there is no dog, only the toy) and my conclusion though true, because of the squirrel behind the brush, is not known. The third and fourth resistance strategies get this clear 'no knowledge' case wrong. They both imply that I know that there's at least one animal in the yard. After all, there is a justified and dispositionally believed truth that is both evidentially supported by and entailed by my false belief that there is a dog in the yard: the truth is 'there is a dog or squirrel in the yard'. The third and fourth resistance strategies both rule that this belief epistemizes my conclusion. I therefore am judged to have knowledge in this case which goes against the standard and correct Gettier verdict. The sophisticated third and fourth resistance strategies fail and with them fails the resistance.

I think this passage is not probative, in light of the fact that Warfield's *actual opponents* (*e.g.*, Audi, Williamson, and others) defend **Conservation**, and so would (presumably) go for the following "resistance" strategy:

- **Resistance**. S has (inferential) knowledge of q despite the presence of an involved falsehood p (in S's inference that q) iff there is a truth p' such that: (a) S (at least) dispositionally believes p', (b) p' is either entailed or epistemically supported by p, and (c) S *is in a position to know* that p' is true.

Once condition (c) is added to the **Resistance**, Gettier examples of the kind Warfield discusses above are no longer sufficient to show that the resistance fails (take a moment to convince yourself of this). The problem with **Resistance** — *if there is any* — will *not* be that it *over*-generates knowledge attributions. As I will explain below, the only problem **Resistance** might have is that it *under*-generates knowledge attributions. At this point, I will leave the defense of (RV) from alleged cases of (KFF) behind. Next, I will focus (instead) on what I think the *real aim* ought to be here, which is to defend (RV) from alleged cases of (KFNK).

3.1 Resistance: Defending (RV) from (KFNK)

I will break my discussion of **Resistance** into two subsections, corresponding to these two versions:

(1) p *entails* p'.

(2) p does *not* entail p', but p *epistemically supports* p' (possibly, relative to one's total evidence).

3.1.1 Defeating the (1)-Resistance to (KFNK)

The standard rendition of (1) uses something like the following "alternative epistemicizer":

(p') It is *approximately* 3pm.

While p' might seem like a plausible candidate in Warfield's original example, I think it — and any other (1)-candidate — is doomed to fail in the following *refinement* of Warfield's example. This time, let's suppose (with the same setup as Warfield's original example) that I competently deduce a claim of the form:

($q_n^\star$) It is 3pm $\pm$ n seconds.

from

($p^\star$) It is 3pm $\pm$ 60 seconds.

As before, our "premise" $p^\star$ is *false*, because it is exactly 2:58. Moreover, and for the same reason, if $q^\star_n$ is going to be *true*, then we must have $n > 120$. Now, suppose that there exists a *threshold value* k such that:

- $q^\star_k$ is the logically strongest claim about the time that I am in a position to know in the context.

I will discuss the possibility that there does *not* exist such a k, below. But, for now, suppose there *is* such a k. And, suppose I competently deduce $q^\star_k$ from $p^\star$. Now, it can be shown that there can be no (1)-style "alternative epistemicizer" p' to go proxy for $p^\star$ *in such a case*. For simplicity (but without loss of generality), I will assume that all "alternative epistemicizers" must be of the form:

(p'_a) It is 3pm $\pm$ a seconds.

Now, there are only three possible cases to consider:

- **Case 1**: $a < k$. In this case, p'_a cannot serve, since — *by assumption* — I am *not* in a position to know *any* such p'_a. Thus, condition (c) of **Resistance** is violated by p'_a.
- **Case 2**: $a = k$. In this case, p'_a cannot serve, *since it is identical to the conclusion* $q^\star_k$, and **Foundationalism** *precludes circular epistemic chains.*
- **Case 3**: $a > k$. In this case, p'_a cannot serve, because it is *too weak* to (generally) ground inferential *knowledge* that $q^\star_k$. [Think of a concrete example here. Suppose that $k = 1000$ and $a = 1001$. Intuitively, one cannot come to know $q^\star_{1000}$ *solely on the basis of an inference from* p'_{1001}.[2]]

So, *if* there is such a "smallest k" of the kind presupposed above, *then* the (1)-rendition of **Resistance** is *bound to fail.* What if there is *no* such "minimal k"? I can only think of two possibilities here:

- **Infinite Chains**. One way in which there could fail to be a "smallest k" (such that $q^\star_k$ is the logically strongest claim about the time that I am in a position to know in the context) is if there is an *infinite chain* of stronger and stronger epistemicizers. Let $\ulcorner p \rightsquigarrow q \urcorner$ assert that $\ulcorner p$ epistemicizes $q \urcorner$. One way we could get non-existence of such a "minimal k" is if we had some *infinite chains* like the following:

  $$\ldots p'_{1000.0625} \rightsquigarrow p'_{1000.125} \rightsquigarrow p'_{1000.25} \rightsquigarrow p'_{1000.5} \rightsquigarrow p'_{1001}$$

  Of course, this won't help the (orthodox) (1)-**Resistance**, since **Foundationalism** *precludes infinitism.*[3]

- **Vagueness**. The only other way I can see that there could fail to be a "smallest k" (in the above sense) is if there were some *vagueness* about which $q^\star_k$'s we were in a position to know in such contexts. Interestingly, some (RV)-theorists (I mean Williamson) are *epistemicists* about vagueness, which (here) means they think there *always is* a fact of the matter about whether one is a position to know $p'_a/q^\star_k$. But, if we think this, then we're back to the arguments above, *and the (1)-**Resistance** fails.* So, only (RV)-theorists who are *non*-epistemicists about vagueness, *and* who think claims like "S is in a position to know $p'_a/q^\star_k$" *are vague* can wriggle out of the above (1)-**Resistance** refutation. Having said that, I don't see how taking this sort of vagueness stance is ever going to lead to a *victory* for the (1)-**Resistance**. After all, they need to show that *there always does exist* a type-(1)-alternative-epistemicizer p'_a, for any case in which knowledge that $q^\star_k$ is (definitely) obtained *via* inference from some (definitely) *non*-known p. I don't see how the existence of "gray areas" in which *there is no fact of the matter* as to whether or not S is in a position to know $p'_a/q^\star_k$ is going to *help* the (1)-**Resistance** meet that burden.

To sum up: I think my variation on Warfield's example shows (rather definitively) that the (1)-**Resistance** to (KFNK) is not going to work. So, I agree with Warfield on that score (but for independent reasons). In the next section, I consider a completely different tack on behalf of The Received View: (2)-**Resistance**.

---

[2]Note: I do *not* mean to presuppose that the inference is "deductive" (and, hence, somehow "logically mistaken" or "defective"). The problem here is just that the inference (whether it be "deductive" or "inductive") won't — *in general* — yield *knowledge* of $q^\star_k$.

[3]Peter Klein (Warfield's adviser at Rutgers) has defended *both* (KFNK) *and infinitism* (*i.e.*, the existence of infinite inferential chains). This is interesting, since my example reveals that making the case for (KFNK) is *harder* if infinitism is true!

3.1.2 Can the (2)-Resistance to (KFNK) be Defeated?

Interestingly, Warfield never really considers the (2)-**Resistance**. He only considers a case in which p *both* entails *and* epistemically supports p', which is an example of (1)-**Resistance**. Since I think it's clear the (1)-**Resistance** is bound to fail (and for reasons independent of Warfield's), this omission of Warfield's is unfortunate. Indeed, I suspect the strongest case for the **Resistance** will involve "proxies" satisfying (2).

Let's work with my example above, and let's suppose (for simplicity) that there *is* a smallest k such that $q_k^\star$ is the logically strongest claim about the time that I am in a position to know in the context. And, just to refresh your memory, we are talking about a case in which I competently deduce this "strongest claim":

($q_k^\star$) It is 3pm $\pm$ k seconds.

from (the *false* claim):

($p^\star$) It is 3pm $\pm$ 60 seconds.

We assume, also, that I *come to know* $q_k^\star$, and that this knowledge is *inferential*. Now, on behalf of the (2)-**Resistance**, consider the following candidate "alternative epistemicizer" of $q_k^\star$:

(p'') My watch *reads* "3:00pm", *and* my watch is accurate to within k seconds.

Note, first, that $p^\star$ does *not entail* p'', although (plausibly[4]) $p^\star$ *does epistemically support* p''. Next, in order for this to count as an instance of (2)-**Resistance** to (KFNK), we'll need p'' to be *something I am in a position to know in the context.* Thus, *the key question* is going to be the following:

- Can we modify our example, so as to render my belief in p'' *merely* a *justified true belief*, *while preserving the intuition* that p'' can serve as an epistemicizer of my inferential *knowledge* that $q_k^\star$?

Presumably, I know (*perceptually*) that my watch *reads* "3:00pm". So, the only room for tinkering here is with the following claim:

($\star$) my watch is accurate to within k seconds.

I will just assume that (in this case) my knowing that p'' depends *only* on my knowing that ($\star$). What if my belief that ($\star$) is *not* known? What if it is *merely justified and true*? This is tricky, but what about the following sort of scenario?

> I remember that the owner's manual that came with my watch said that it was accurate to within k seconds. As it happens, the box contained the owner's manual for the previous model of my watch (I did not notice this, and that owner's manual has since been destroyed), which just happens to have the same accuracy as my watch. [Maybe even add the following information to the example? Most other watch upgrades by the same manufacturer were occasioned by *increased* accuracy, and none involved a *decrease* in accuracy.]

I'm not really sure what to think of this case. It doesn't sound crazy (to me) to say that I *know* $q_k^\star$ in this case, even though I'm not (or may not be) in a position to know ($\star$). Of course, if one insists on always following a (2)-**Resistance** strategy, then one might now *retreat further* to the following "alternative":

(p''') My watch *reads* "3:00pm", *and* the owner's manual that came with my watch *said* that my watch is accurate to within k seconds.

In general, this is how an (RV)-theorist is going to try to explain inferential knowledge cases. If pressed, they will eventually produce (finite) epistemic chains that trace back to *basic sources of knowledge.*

One interesting thing to note about this progression is that it leads to more and more confidence in the *truth/knowledge* of the premise-beliefs, but it also leads to less and less confidence that the premise-beliefs (propositionally) *justify* (or *warrant*) the conclusion to a sufficiently strong degree (so as to undergird inferential *knowledge* of the original conclusion). This is not an atypical pattern in epistemological theorizing. We can make our "evidence" more *secure* by "retreating" in this way, but this can also lead to *weakened* evidence claims that may seem incapable of doing all the "*supporting*" that we need our evidence to do.

[4]It is plausible to suppose that [if this (2)-**Resistance** strategy is going to work] p'' will *raise the epistemic probability* of $p^\star$. And, since probability-raising is *symmetric*, we'll also have $p^\star$ raising the epistemic probability of p''. In this sense, we'll have *support.*