TW Chapter 3 - Roush

This chapter is meant to deepen and broaden an argument that goes something like this:

The internalist (about mental states) thinks that knowledge can't be a mental state because knowledge involves this external element of truth which isn't local to the subject. Any mental state can explain human action, and to do so it must be local. (It is in *virtue* of the locality of mental states that they can explain action.) Hence, knowledge can't be a mental state because it isn't local and for that reason can't explain human action, or doesn't add anything to the explanation of human action that was not there from other mental states (like belief or true belief).

It seems that according to the argument TW attributes to the internalist we see independently that knowledge adds nothing to the explanation of action that belief or true belief didn't give us, and then one explains this by saying that mental states and only mental states can explain action because action requires locality, and mental states are internal. The first mover in the argument of the supposed internalist seems to be that action can be explained only by something local (due to something like a classical causal picture) and that's both why mental states can explain action, and why knowledge can't (and so can't be a mental state).

In arguing that knowledge is broad, TW argued that knowledge can add something mental to belief or rational belief or… whatever idea short of knowledge that you might try to capture the mental in.

The internalist can come back and say, well, okay, knowledge is broad, but the thing that knowledge adds to the mental is an internal part of knowledge, which can be isolated from the external part in a decomposition into conjuncts. If knowledge is prime, then it can't be decomposed into internal and external conjuncts. This is why TW is arguing that knowledge is prime. In the section on causal explanation of action, he argues by intuition that knowledge *can* do work in explaining action that belief or true belief just cannot do.

Question: Why not conclude from the intuition that knowledge can add something essential to explanations of action that mental states alone cannot explain action? What explains action is a seamless (prime) web of relation between the mental and the physical or what remains of the environment once the prime mental is hived off? In other words, mental states can be prime and have external aspects without *knowledge* being a mental state. On this view, knowledge contains a prime conjunct [prove that if knowledge contains a prime conjunct then knowledge is prime, so the prime conjunct can *explain* the primeness of knowledge], and it's that prime conjunct that explains action. TW hasn't ruled out that possibility.

So two issues arise: is it plausible that a prime conjunct explains the ability of knowledge to explain action? Yes. Tracking is a prime conjunct and tracking explains why knowledge can explain action.

TW implicitly appeals here to his view (consistent with his FMSO argument) that truth is *not an independent variable* in knowledge. (If knowledge is prime only because of a prime *conjunct* then truth can be an independent variable for all we know from the primeness argument.) Why should we think truth is or is not an independent variable?

The argument from Gettier cases that truth is not an independent variable in knowledge: (Williamson does not appeal to it in the book but he endorses it.)

See Zagzebski (1994) paper, "The Inescapeability of the Gettier Problem". Basically, any justified true belief theory where the justification requirement is "beefed up" to get rid of Gettier examples will either have to be infallibilist or have to fail to escape Gettier cases. An infallibilist view implies that what qualifies you as knowing (the beefed up justification requirement) *guarantees* that your belief is true. (Thus, truth is not an independent variable in knowledge.) If your justification requirement does not imply the truth of the belief, then it is possible for it to be fulfilled by a Gettier case and so give a belief that counts as knowledge but does not look like knowledge.

What do we need for a Gettier case? We need the justification-type requirement to be fulfilled, the belief to fail *in this case* to be made true in virtue of the things the justification-type requirement picks out, and the belief to be made true by things the person doesn't know about, so can't use to justify.

Another way to see the point: To stop the Gettier possibility your justification requirement is going to have to be strong enough that there is no (misleading) evidence that would undermine your justification (such misleading evidence would be the first step in constructing a Gettier case). However, since confirmation is non-monotonic (can zig-zag all the way up) there never will be such a point until you get to truth or 100%.

Williamson's view is infallibilist, or at the very least shares with infallibilism the implication that truth is not an independent variable. Verify that it is not susceptible to Gettierization.

Let's verify the infallibilism. Fallibilism says that if you know then you still could have been wrong, for all you did to get to your status of knowing. What you did doesn't guarantee the truth of your belief. If knowledge is the most general factive stative mental state, then knowing does guarantee that what you did guarantees the truth of the belief. After all, that's the whole point of a factive attitude, that its holding guarantees the truth of the statement the attitude is toward. "Seeing that", for example, is one way of knowing on TW's view, and that guarantees the truth of the statement about the thing seen.

So, Williamson's use of primeness to argue against the causal-explanation-of-action-can-only-be-local folks depends on the assumption that truth is an independent variable, that knowledge is infallible, and that the primeness of knowledge that makes it capable of explaining action comes from the primeness of knowledge rather than the primeness of a conjunct in the analysis of knowledge.

The tracking view, mine or Nozick's, is now relevant because TW has a suite of claims that involved infallibilism, truth not being an independent variable (all resting, I think, on the FMSO argument and a response to Gettier), primeness, and the explanation of action. We've seen above that knowledge is prime according to tracking, and since primeness is what makes knowledge able to explain action, TW can't rule out the tracking view on that ground alone. (Also, the tracking view can explain *why* knowledge is prime and why knowledge is able to explain action, which TW can't). But the other issue about infallibilism is yet to be addressed for tracking. Tracking, either old or new style, is fallibilist (only >.95 probability is needed), and though it works against all the usual Gettier examples (you can verify) it seems that there is a recipe through which you can get something similar. The usual recipe says describe a case where the justification requirement is fulfilled but for all the person has done p could be false. Then make p true a different way that the person has no evidence about. For tracking we're going to have to say: imagine someone who fulfills the tracking requirements and for all their dispositions the statement they believe could be false. Then make it true another way. Try a guy who firmly believes that every person smiles if and only if they've won the lottery. He walks by a guy who is smiling and concludes he won the lottery. He is right about this. Moreover, it just so happens that this is a guy who wouldn't have been smiling had he not won the lottery (his wife just died), and would definitely smile whenever he did. So our subject fulfills the tracking conditions and has a true belief. He has knowledge on the tracking view and it looks weird. Is it a Gettier case though?

Suppose it is. Now we have to assess the advantages and disadvantages of TW's view systematically. For it may be that tolerating some accidental knowing is worth it because it gives us something else and something has to give. I think this is the case. TW gets rid of Gettier cases through an infallibilist move (albeit a fresh and clever one--FMSOs), and also a move that involves explanatory loss. Here's what I have in mind: supposedly, it is sufficient for knowledge of p to have a factive mental state toward p. But what is the factive mental state through which we have scientific knowledge, say, through which we could conclude that Einstein *knew* that simultaneity was relative to reference frame? He didn't *see* it surely. Indeed, it wasn't a mental state at all. It was a medley of a whole lot of things that put his beliefs in a responsive relation to the way the world is. One factive mental state also does not seem to be necessary for scientific knowledge.

Notes on Section 3.6 of *Knowledge and its Limits*
09/27/06 (B.F.)

In §3.6, TW sets out to explain how it is possible for the following three claims to be true simultaneously:

1. D is sufficient for both C and E.
2. E is not sufficient for C.
3. E is "more strongly correlated with" C than D is.

Here, C is the *explanandum* that "one will perform a certain action," D is "the very specific condition obtaining in the case at hand (for example, completely determining the agent's internal physical state and the physical state of the environment)," and E is "the condition that one knows p." The idea is that D and E are meant to be *competing explanans* regarding C. The ultimate aim of TW's arguments involving the explanatory power of knowledge *vs* belief (*etc.*) is to motivate the joint possibility of claims (1), (2), and:

4. E is a *better explanation of* C than D is.

While degree of correlation is not the same thing as degree of explanatoriness (of course), TW thinks it is an important "sign" or "indicator" of degree of explanatoriness. Thus, while the possibility of (1)–(3) is not sufficient to establish the possibility of (4), it is nonetheless important to TW [in his attempt to argue for the possibility of (4)] that (1)–(3) can be true simultaneously. Question: if (1)–(3) *couldn't* be true simultaneously, would that *disconfirm* the possibility of (4)? If not, then I'm not sure what the point of the section is.

Williamson makes (heavy and essential) use of the following probabilistic measure of "the degree to which C is correlated with D" (note that his definition of ρ on page 84 contains a typo in the numerator):

$$\rho(C,D) = \frac{\Pr(C \wedge D) - \Pr(C) \cdot \Pr(D)}{\sqrt{\Pr(C) \cdot (1 - \Pr(C)) \cdot \Pr(D) \cdot (1 - \Pr(D))}}$$

Here, the probabilities are meant to be *objective chances* (where we hold fixed some set of background conditions, which may include *inter alia* the agent's desires, and which are bound to vary from explanatory context to explanatory context). Williamson also says that these are *not* supposed to be *single case* probabilities (*e.g.*, propensities). He wants the conditions C, D, and E to be *general* (as opposed, say, to *token events*). He says that they "can obtain in many actual cases". OK, that sets the stage for my discussion.

First, note that it is certainly true (indeed, it is *mathematically* true) that *if* one measures "degree of correlation" using ρ, *then* (1)–(3) are jointly satisfiable, just as TW wants (and claims). This is established by the probability model TW reports in this section. I'll call that model $\mathcal{M}_1$. I include it here for completeness:

($\mathcal{M}_1$)

C	D	E	State Descriptions (s_i)	$\Pr(s_i)$
T	T	T	$C \wedge D \wedge E = s_1$	$\Pr(s_1) = 1/10$
T	T	F	$C \wedge D \wedge \sim E = s_2$	$\Pr(s_2) = 0$
T	F	T	$C \wedge \sim D \wedge E = s_3$	$\Pr(s_3) = 3/10$
T	F	F	$C \wedge \sim D \wedge \sim E = s_4$	$\Pr(s_4) = 1/10$
F	T	T	$\sim C \wedge D \wedge E = s_5$	$\Pr(s_5) = 0$
F	T	F	$\sim C \wedge D \wedge \sim E = s_6$	$\Pr(s_6) = 0$
F	F	T	$\sim C \wedge \sim D \wedge E = s_7$	$\Pr(s_7) = 1/10$
F	F	F	$\sim C \wedge \sim D \wedge \sim E = s_8$	$\Pr(s_8) = 4/10$

What I want to discuss are the following questions (and related ones):

- Where does ρ come from (*i.e.*, for what purpose what was it introduced)?
- Does ρ have any properties might make it *not* such a good proxy for "degree of explanatoriness"?
- What other probabilistic "measures of correlation" are there (and how do they compare with ρ)?
- What is the relationship between *explanatory correlation* and *evidential/predictive* correlation?
- Do other accounts of explanation support Williamson's claims and argumentative strategy here?

- Are appeals to explanatoriness (which seem *pragmatic*) probative wrt TW's *metaphysical* claims?

Where ρ comes from. ρ was originally introduced (by statisticians doing linear regression) as a measure of *linear* correlation among random variables that can take more than two values. The "conditions" (C, D, and E) that TW is using here are *dichotomous* or *two-valued* random variables. This is important, because in the case of RVs with more than two values, (linear) *non-correlation* ($\rho = 0$) *does not entail probabilistic independence*. Thus, in the regression setting, correlation is *distinct* from probabilistic dependence. However, when we look only at 2-valued RVs (as TW does), there is no longer any distinction between non-correlation ($\rho = 0$) and *independence*. As such, there is no obvious reason to use ρ in this context, since what we are talking about here is just *degree of probabilistic dependence* (it's somewhat misleading to call it "correlation"). This is one reason why almost nobody uses measures like ρ to measure dependence between *propositions*.

Some (Other) Properties of ρ as a Measure of *Probabilistic Dependence*. TW reports various properties of ρ in his Appendix. Most of these have to do with deriving necessary and sufficient conditions for the *comparative* relation $\rho(C,D) \leq \rho(C,E)$, see below. Here are two other properties of ρ that bear mentioning

- For all C, D, $\rho(C,D) = \rho(D,C)$ and $\rho(C,D) = \rho(\sim D, \sim C)$.

It is not at all clear why such symmetries should hold generally for a measure of probabilistic dependence. I think this is especially true if one intends to use these measures as a "guide" to "degree of explanatoriness". Intuitively, "degree of explanatoriness" is certainly not symmetric in these respects. There are other measures that satisfy many of the "intuitive" properties of "degree of correlation" that TW mentions *without* being symmetric in these ways. Moreover, even if we decided that we wanted our measure of "degree of correlation" (in this context) to be symmetric in these ways, this still would not single out ρ among the plethora of measures of probabilistic dependence. Most importantly, it wouldn't even ensure *comparative agreement* with ρ, which is what TW needs here (at least, with respect to the concrete example he gives).

Other Measures of Probabilistic Dependence. As it turns out, there are *many* measures δ of probabilistic dependence that have the basic properties Williamson mentions in connection with ρ (even the symmetry properties of ρ, if you like). However, these many measures disagree radically concerning *comparative* claims of the form $\delta(C,D) \leq \delta(C,E)$, even in the very example TW gives in this section. The key, then, to TW's discussion in this section are his necessary and sufficient conditions for $\rho(C,D) \leq \rho(C,E)$:

$$(\Pr(C \mid D) - \Pr(C)) \cdot (\Pr(C) - \Pr(C \mid \sim D)) \leq (\Pr(C \mid E) - \Pr(C)) \cdot (\Pr(C) - \Pr(C \mid \sim E))$$

Here, Williamson describes this inequality in a rather loaded and perplexing way, as follows:

(the degree to which D raises the probability of C) $\cdot$ (the degree to which $\sim D$ helps us predict $\sim C$)
$\leq$
(the degree to which E raises the probability of C) $\cdot$ (the degree to which $\sim E$ helps us predict $\sim C$)

But, how should one measure "*the* degree to which D raises the probability of C"? TW seems to be assuming there is *just one* such measure $[\Pr(C \mid D) - \Pr(C)]$. But, it is more accurate to say that there are *many* such measures, since there are many ways of saying that "D raises the probability of C", and there are many (corresponding) ways of generalizing these to quantitative measures (and, again, these many measures disagree radically). The same can be said for "*the* degree to which $\sim D$ helps us predict $\sim C$". The corresponding *mathematical* expressions denote different sorts of *probabilistic dependence* measures. And, we've been given no way of choosing between them for the purposes at hand. This is important because one can construct a relevance measure to get just about any comparative judgment one wants in any given example. Basically, what TW has done here is find *a* measure of probabilistic dependence that makes (1)–(3) jointly satisfiable. But, there are many measures of probabilistic dependence out there that make (1)–(3) jointly unsatisfiable (or unsatisfi*ed* in $\mathcal{M}_1$ specifically). What he should try to do here is to argue that *any* "appropriate" (in this context) measure of probabilistic dependence will make (1)–(3) possible. Moreover, it would be nice to see an independent argument for the intuition Williamson seems to have that (in the salient sense) E *is* "better correlated with" C than D is (in $\mathcal{M}_1$). That's not obvious to me, especially if we think about how strong D is as *evidence* regarding C in $\mathcal{M}_1$. It seems clear to me that D provides *better evidence* for C than E does

(in $\mathcal{M}_1$). That may not be a reason to doubt that *E better explains C* than *D* does, but what's at issue here is whether *D* is "better correlated with" *C* than *E* is (in $\mathcal{M}_1$). And, I'm afraid I don't have an independent grasp on "degree of correlation" that isn't parasitic on some other concept that we're trying to *explicate using* probabilistic dependence (like degree of explanatoriness or degree of confirmation, *etc.*). I'll return to the question of evidential support (cum probabilistic relevance) when we get to TW's chapter on evidence. There, he takes evidential support to be a(nother?) kind of "probabilistic correlation" between propositions.

Two Existing Approaches to Scientific Explanation. There is a vast literature on (probabilistic or otherwise) approaches to scientific explanation (and, of course, causation, but I will bracket the causal questions here). On the course website, I have added links to papers by Chris Hitchcock and Ban van Fraassen. Both of these papers take a *contrastive* view of explanation. And, both of these papers would, it seems to me, take a different view about the probabilistic (and philosophical) status of Williamson's examples than he does. Hitchcock and van Fraassen think explanatory claims are, generally, of the following *contrastive* form:

- *E* rather than *D* explains why *C* rather than C' is the case (relative to a set of contextual presuppositions).

Hitchcock suggests that (typically) one of the presuppositions in force is that the pairs of contrast conditions $\{E, D\}$ and $\{C, C'\}$ are not logically redundant (he even seems to think that they are usually presupposed to be *mututally exclusive*). That would already rule-out the claim "*E* rather than *D* explains why *C* rather than $\sim C$ is the case (relative to the presuppositions in the context TW has in mind, some of which are encoded by $\mathcal{M}_1$)" as infelicitous. Even if we ignore this part of Hitchcock's theory and we grant that such a claim can be felicitous (despite the fact that *D* and *E* are, in a sense, not *competing explanans*, since *D entails E* relative to $\mathcal{M}_1$), Hitchcock would still disagree with Williamson's conclusion about his $\mathcal{M}_1$ example. This is because Hitchcock thinks the appropriate *probabilistic* component of the assessment of the claim in question is:

- $\Pr(C \mid D) = 1 > \Pr(C \mid E) = \frac{4}{5}$.

This would lead us toward the conclusion that *D* is more explanatorily relevant to *C* than *E* is (contrary to what TW wants us to conclude in this example). Williamson thinks we need to look not just at $\Pr(C \mid D)$ *vs* $\Pr(C \mid E)$, but also $\Pr(C \mid \sim D)$ *vs* $\Pr(C \mid \sim E)$. There are various reasons why Hitchcock does not think $\Pr(C \mid \sim D)$ and $\Pr(C \mid \sim E)$ are relevant (or probative) here. One of the main motivations for a contrastive account of explanation is the belief that relations of positive or negative explanatory relevance only hold *relative to specific alternatives*. So, unless we are contrasting the explanatoriness of *D vs* $\sim D$ or *E vs* $\sim E$, there is no need to look at $\Pr(C \mid \sim D)$ or $\Pr(C \mid \sim E)$ here, and, indeed, looking at these quantities will usually be misleading in cases where we're contrasting the specific "alternatives" *D* and *E*. [This raises another issue about Williamson's claim. From a contrastive point of view, we should also be concerned with *specific* alternatives to *C*. So, we should also be looking at what other specific actions C' the agent might have performed in the context, and how *D* and *E* fare as competing explanations of *C vs* C'. I will let that go.]

van Fraassen also has a contrastive account of explanation. He offers a general *pragmatic* theory of why-questions. He maintains that explanation is *deeply* contextual and pragmatic. Indeed, van Fraassen seems to think that explanation (and explanatory relevance) are *so* pragmatic and context sensitive that, for *every* explanatory claim (of the general contrastive form above) there will exist *some* context relative to which that claim is true. To take TW's claim in particular, van Fraassen would simply grant that there are some contexts relative to which *D* is a better explanation of *C* than *E* is. But, he would be quick to point out that this doesn't imply that knowledge is "essential" to the explanation of behavior, since there will also be contexts relative to which *D* is a better explanation of *C* than *E* is — even holding fixed all the probabilistic facts encoded in $\mathcal{M}_1$ (or in any other model you like, for that matter!). Moreover, van Fraassen thinks that, while certain probabilistic quantities are often (even always?) involved in the evaluation of answers to why-questions, there is no *single* probabilistic comparison that is a "guide to explanatoriness" across all contexts. If van Fraassen is right about explanation (and I am sympathetic), it's unclear to me how probative appeals to explanation will be for Williamson. What, exactly, are the examples about explanation supposed to show, and how does this bear on the *metaphysical* stance on knowledge and mind Williamson wants us to take? [This is reminiscent of the question of the role of explanation in science and how it bears on scientific realism. On that set of issues, I recommend the first few chapters of van Fraassen's *The Scientific Image*.]