Williamson - Knowledge and its Limits

Brian T. Miller

October 25, 2018

0 - Introduction

Williamson's project is to make *knowledge* central to the project of epistemology. He defends several main theses:

1. Knowledge is a mental state
2. Our epistemic access to our own mental states is far more constrained that generally supposed
3. Evidence = Knowledge
4. Knowledge is the norm of assertion

Why is (1) controversial?

First reason to doubt (1): Failure of Analysis

The orthodox account: belief is more conceptually basic than Knowledge

> Knowledge that P $=_{df}$ belief that P & truth of P & [something that survives Gettier cases]

If the orthodoxy is true, then it's plausible that knowledge is composed of a mental state (belief) plus a non-mental state (whatever makes that belief true, the state of the extra-mental world) plus the Gettier defeater.

Why believe the orthodoxy?

Proposal 1: because believing P is a necessary but not sufficient condition for knowing P; the conjuncts are conceptually prior to the conjunction

Objection: no reason to believe that there is a non-circular, conjunctive analysis of knowledge.[1]

[1] compare: there probably isn't a non-circular, conjunctive analysis or 'red' of the form: red $=_{df}$ colored & X

Proposal 2: because we have analysis X (for some value of X) of knowledge in terms of belief, and the correctness of analysis X is analytic or knowable a priori

Objection: we have no such analysis that's immune to counterexamples

Proposal 3: concede that current analyses are inadequate, but claim that the inadequate analyses we do have are *almost* good enough, and they indicate that an adequate analysis is soon forthcoming

Objection: close approximations are not an indication that an adequate analysis is forthcoming[2]

[2] Example: 'parent' is closely approximated by 'ancestor of' & 'isn't an ancestor of an ancestor' but not quite (counterexample: father who is both ancestor of son and grandfather (and hence ancestor of ancestor) of son (via incest). So 'parent' can't be analyzed in terms of 'ancestor & X'.

Second reason to doubt (1): Content Externalism

Conceptual priority of belief to knowledge is partially rooted in an internalist conception of mind, on which:

- Content of belief — the proposition believed — is determined purely internally
- Knowledge requires that belief is true, and truth is (usually) determined externally
- So, knowledge can be analyzed into an internal (mental) and an external (non-mental) component

Objection to that conception: internalism about mental content is false, and externalism about mental content is true.

The disagreement about what determines what properties an object has. Broadly speaking, *Internalism* is the thesis that whether or not an object has property F is determined by the physical object at the present moment, and Externalism is the denial of Internalism.

Examples:

- X is my wedding ring, and Y is an exact physical replica of X. Does Y have the propery of being my wedding ring?
  - Internalists about the property of *being my wedding ring* think that the question is answered by looking only X itself. Since Y is my wedding ring, and X is an exact physical replica of Y, the X has the property of being my wedding ring too.
  - An externalist might disagree by claiming: whether X is my wedding ring depends partially on its historical properties, e.g. on whether it was given to me by my wife at my wedding. Since the replica wasn't (let's stipulate), then it doesn't have the property of being my wedding ring.[3]
- X is a mosquito bite. Y is an exact physical replica of X created by incredibly precise microsurgery. Is Y a mosquito bite?
  - Internalists about mosquito bites say 'yes', externalists about mosquito bites say 'no', presumably because they thing that

[3] I can't remember where this example comes from

the causal having the property depends in part on the causal origins of the physical object.[4]

- X is a picture of Winston Churchill that I traced in the sand with a stick. Y is a physically identical picture traced by ants crawling around in the sand (it's a fluke, not smart ants). Is Y a picture of Winston Churchill?
  - Internalists about pictorial content say 'yes', presumably because of how it looks, a property determined by the internal physical structure of the pile of sand. Externalists say 'no', presumably because a picture can't represent something unless the artist has had some sort of causal interaction with Winston Churchill.[5]

[4] This example is from the SEP article *Externalism about Mental Content*. Need a crash course on externalism about mental content? Check out the article — it's great.

[5] Example from Putnam's *Reason, Truth, and History* Ch 1.

Internalism about mental states claim a person's mental states supervene on their internal physical states. This implies that I have the exact same mental states as my exact physical duplicate, regardless of our histories or our environments. Externalists disagree.

Williamson's example: Tom has experiences of tigers, and at time *t* he believes that tigers growl. Schmom has experiences of schmigers, 'beasts of a similar appearance belonging to a different species'. At time *t* Schmom is an exact physical replica of Tom. Does Schmom believe that tigers growl, or does he believe that schmigers growl?

Putnam's classic example[6]: Oscar lives on Earth in 1750 (before people knew that water = H_2O). Toscar lives on Twin-Earth in 1750, which is exactly like Earth except that the clear liquid in rivers and streams is XYZ, not H_2O. Oscar and Toscar are exact physical duplicates. [7]

[6] from 'The Meaning of Meaning'

[7] Let's ignore the unfortunate fact that Oscar's body is 2/3 H_2O and Toscar's body is 2/3 XYZ, so they can't be exact physical replicas. This wasn't the best example to make the point, but that's how things played out.

Question about *linguistic* content: When Oscar says 'water', what is he referring to — H_2O or XYZ? Both? When he points to a glass of XYZ and says 'that's water', has he spoken truly?

Question about *mental* content: what are Oscar's aqueous thoughts about — H_2O? XYZ? Both? When he takes a sip of XYZ and comes to believe 'I'm drinking water', is his belief true?

Why does this matter?

- Internalism is threatening to Williamson's Knowledge as a Mental State thesis because it facilitates factoring Knowledge into an internal component (belief) and an external component (the world that determines whether the belief is true), along with the Gettier condition.

  - If belief content externalism is true then there's no purely internal component for that factorization because the environment outside of the believer has a role to play in determining the content of the agent's beliefs.

- Belief is unambiguously a mental state. So depending on the environment must be consistent with being a mental state.
- So can't say that knowledge isn't a mental state just because it depends on the environment.
- Same goes for any other putatively *factive mental state*: mental states that entail the truth of their contents any time someone is in that state. Factive mental states include knowing, perceiving, and remembering.

Knowledge and Skepticism

Skeptic's claim: you and your mental state duplicate might be in radically different environments. Your beliefs are mostly true and you know lots of things, but your duplicate is in a skeptical scenario in which most of their beliefs are false and she doesn't know much. In other words, the truth values of your beliefs vary depending on the environment, even once all mental states are fixed.

But, if knowledge is a mental state then this setup is impossible: if in the skeptical scenario your duplicate believes falsely that P then in that scenario she doesn't know that P (because Knowledge is factive). Since you share all the same mental states with your duplicate, if knowledge is a mental state and your duplicate doesn't know that P, then you don't know that P either. That contradicts the description of the case.

Knowledge and Evidence

What is evidence?

- If belief is prior to knowledge, then it's natural to think that evidence and justification are too. In that case it's unfruitful to try to understand evidence and justification in terms of knowledge.
- What if E=K: i.e. what if my evidence consists in all and only the propositions that I know?[8]
  - In that case it's knowledge that does the justifying

[8] NB: In addition to claiming that knowledge justifies belief, Williamson also claims that knowledge is the norm of assertion: I ought to assert P only if I know P.
This claim has received lots of attention in the literature, but we're not going to spend much (any?) time on it in this seminar.

- Still allows for justified belief that isn't knowledge (because known propositions might constitute misleading evidence for a false proposition)
- Provides a response to regress argument: the foundation is your knowledge
- Justification is no longer though of as a precondition of knowledge, but merely as a status of beliefs that fail to live up to the real standard, knowledge (i.e. justified belief that isn't knowledge = defective knowledge)

We'll spend several weeks on Williamson's claim that E=K. If he's right, then:

- All evidence is propositional
  - In that case what's the epistemic significance of expeirence?
- every bit of propositional evidence is true
  - So false propositions that I'm justified in believing can't justify other beliefs via inference?
- In order for a true proposition to be *my* evidence, I must *know* that it's true (as opposed to merely believe it with justification)

Anti-Luminosity

So far: knowledge is a 'mental state which constitutes the evidential standard for assertion and belief'

Single objection to all three parts of this picture: KK principle is false, so:

- If knowledge is a mental state then we don't have perfect access to our own mental states
- If E=K then we don't always know what our evidence is
- not always in a position to know whether your assertions live up to the knowledge norm

Each objection assumes: there are non-trivial mental states that really are accessible in the relevant way, and knowledge isn't accessible in that way.

Williamson's response: no condition is such that we are nearly always in a position to know whether we are in it (this is his famous 'Anti-Luminosity'). Here's a very rough sketch of his argument:

1. a condition C [ex: *being cold*] is luminous iff whenever it obtains, one is in a position to know that it obtains
2. Our powers of discrimination are limited, so for any W where luminous condition C obtains there will always be a W' which is subjectively indistinguishable from W.
3. So, anything we're in a position to know in W we're also in a position to know in W'
4. So if we're in luminous condition C in W, we're also in luminous condition C in W'
5. Rinse and repeat for W", W"'É
6. So if a luminous condition obtains anywhere then it obtains everywhere
7. Conclusion: almost no conditions are luminous

Upshot: whatever level of accessibility evidence must have, luminosity is too stringent, so it's no problem that knowledge isn't luminous.

'Once the standard for the epistemic accessibility of evidence is set at an attainable level, knowledge meets the standard.' (15)

1 - A State of Mind

1.1 Factive Attitudes

Central Claim: knowledge is a mental state.

Clarifications:

- We're talking here about propositional knowledge or knowledge-that rather than knowledge-how
- Knowing that p is a relation between an agent and proposition p
- Knowledge is factive: is S knows that P then p is true.

So a consequence of the claim is that there exist genuinely mental states that are factive. Other examples: remembering, regretting, lamenting

Contrast: belief is a non-factive relation between an agent and a proposition. Belief is also a paradigmatic mental state.

Reason to think that knowing isn't a mental state: knowing that p is a composite, consisting of a genuine mental state (belief that p) plus a non-mental component: the truth of p. Truth isn't mental, and no composite containing a non-mental component is itself mental.[9]

[9] Knowing isn't special in this regard: by parallel argument, no factive state is mental.

Important bit of burden-shifting (22):

Williamson: burden of proof should not rest with the opponents of the above view. Knowledge is very similar[10] to paradigmatic examples of mental states like belief or desire. Not clear why knowledge itself isn't one of those paradigmatic examples. It's possible that there could be theoretical reasons to abandon that view, but the burden of proof is on W's opponents to provide those reasons.

[10] W doesn't say *how* it's similar

Why this is important:

W's strategy is to identify putative differences between knowing and non-factive paradigm mental states and 'eliminate those differences'. Even if he's successful, this approach does not conclusively establish that knowing is a mental state, it merely undermines the case against. But with the burden of proof shifted to W's opponents, this amounts to a win for W.

1.2 Mental States, First-Person Accessibility, and Scepticism

KK principle is almost certainly false. Illustration:

> Consider, for example, the situation of a generally well-informed citizen N.N. who has not yet heard the news from the theatre where Lincoln has just been assassinated. Since Lincoln is dead, he is no longer President, so N.N. no longer knows that Lincoln is President (knowing is factive). However, N.N. is in no position to know that anything is amiss. He continues reasonably to believe that Lincoln is President; moreover, this seems to him to be just another item of general knowledge. N.N. continues reasonably to believe that he knows that Lincoln is President. Although N.N. does not know that Lincoln is President, he is in no position to know that he does not know that Lincoln is President. (23)

Transparency: for every mental state S, whenever one is suitably alert and conceptually sophisticated, one is in a position to know whether one is in S.

If transparency is true then knowledge isn't a mental state[11]

Counterexamples to Transparency involving paradigmatic non-factive mental states:[12]

> I believe that I do not **hope** for a particular result to a match; I am conscious of nothing but indifference; then my disappointment at one outcome reveals my hope for another. When I had that hope, I was in no position to know that I had it...
>
> It fails for the state of **believing** p, for the difference between believing p and merely fancying p depends in part on one's dispositions to practical reasoning and action manifested only in counterfactual circumstances,[13] and one is not always in a position to know what those dispositions are.
>
> Transparency is even doubtful for the state of **being in pain**; with too much self-pity one may mistake an itch for a pain, with too little one may mistake a pain for an itch. (24)

This isn't to say that we have *no* privileged access to our own mental states, just that it's imperfect.

Other possible disanalogies between belief and knowledge:

- perhaps knowing (but not believing) requires reflecting on reasons/ evidence.
  - but, if knowing requires reflecting on reasons, then so does believing rationally, which is a paradigm mental state
- perhaps knowing and rationally believing aren't mental states because they are normative concepts

[11] if Transparency is true and knowledge is a mental state then KK is true; but it isn't. So either T or KMS is false.

[12] The failure of Transparency leaves open the possibility that at least *some* mental states are always apparent. The anti-luminosity argument in ch. 4 is stronger, concluding that *no* non-trivial mental state is apparent/ luminous.

[13] Is that true? Does Williamson need it to be true?

- so is believing: "to have any mental attitude to a content one must in some sense grasp that content, and therefore have some minimal ability to deal rationally with it..."[14] (25)

[14] really?

- perhaps my belief that I know p is defeasible (e.g. when I learn that p is false), while my belief that I believe p isn't
  - my belief that I believe p is defeasible too: I might think that I believe one thing, then act in a way that indicates that I believe something else[15]

[15] What's the relationship between belief and behavior? Does this claim depend on some latent functionalism or behaviorism?

1.3 Knowledge and Analysis

Surprising set of claims:

- believing that it's raining is a mental state
- knowing that it's raining is a mental state
- believing truly that it's raining is not a mental state[16]

[16] Williamson takes it for granted that believing truly isn't a mental state, thinks his task is to explain why this isn't a problem for the claim that knowing isn't a mental state.

Putative problem: Knowing that p $\models$ believing truly that p $\models$ believing that p. But that violates the no-sandwich principle:

No-sandwich principle: There are no three states X, Y, and Z such that state X and state Z are of the same type of state, state Y isn't of that type, and being in X $\models$ being in Y $\models$ being in state Z

Reason to think that the no-sandwich principle is false: it's false when applied to properties. Example:

> being an equilateral triangle (a geometrical property) $\models$ being a triangle whose sides are indiscriminable in length to the naked human eye (a non-geometrical property) $\models$ being a triangle (a geometrical property)

Is the contrast between knowing and believing truly conceptual or metaphysical?

Conceptual claim: the concept *knows* is a mental concept, the concept *believes truly* is not[17]

[17] Williamson's convention is to italicize the names of concepts but not the names of states; I follow his convention.

Metaphysical claim: knowing is a mental state, believing truly is not

Williamson's metaphysics of states and concepts:

- necessarily coextensive states are identical
- necessarily coextensive concepts might not be identical
- possible to have more than one concept of a single state

- possible that one concept of state S is mental, but another concept of S isn't

> For example, since gold is necessarily the element with atomic number 79, the state of having a tooth made of gold is the state of having a tooth made of the element with atomic number 79, but the concept *has a tooth made of gold* is not the concept *has a tooth made of the element with atomic number 79*. Similarly, for any mental state S, the concept *is in S and such that gold is the element with atomic number 79* is necessarily coextensive with the concept *is in S*, so they are both concepts of S. [presumably, though, the latter but not the former is a mental concept] (29)

<u>Relationship between conceptual contrast and metaphysical contrast:</u>

- the existence of a single mental concept of state S is sufficient for S to be a mental state
  - So the existence of a mental concept of *knows* entails that the state of knowing is mental
  - But the existence of a non-mental concept *believes truly* is not sufficient to establish that the state of believing truly isn't a mental state
  - i.e. the conceptual contrast $\not\models$ the metaphysical contrast
- if state S is not mental then there can't be a mental concept of S
  - so, since the state of believing truly isn't mental, there isn't a mental concept *believes truly*
  - but, just because the state of knowing is mental (he's supposing), it doesn't follow that any particular concept of that state is mental, including *knows*
  - i.e. the metaphysical contrast $\not\models$ the conceptual contrast

So, neither contrast entails the other. But:

> ...it is hard to see why someone should accept one contrast without accepting the other. If the concept *believes truly* is non-mental, its imagined necessary coextensiveness with a mental concept would be a bizarre metaphysical coincidence. If the concept *knows* were a non-mental concept of a mental state, its necessary coextensiveness with a mental concept would be an equally bizarre metaphysical coincidence. In practice, sloppily ignoring the distinction between the metaphysical and conceptual contrasts is unlikely to do very much harm. (29)

<u>Here's as close as Williamson has come to explicitly characterizing mental concepts thus-far:</u>

> If the concept C is the conjunction[18] of the concepts C_1,...,C_n, then C is mental iff each C_i is mental. (29)

> ...non-mental concepts in the content clause of an attitude ascription do not make the concept expressed non-mental; the concept *believes that there are numbers* can be mental even if the concept number is not. (29)

> ...the concept *believed truly* is the conjunction of the concepts *believed* and *true*. The conjunct *true* is not mental, for it makes no reference to a subject. Therefore, the concept *believed truly* is non-mental. (30)

So we can now see why Williamson is so eager to reject any identity between the concepts *truth + belief + X* and *knowledge*: if a conjunctive concept includes *true* as a conjunct it isn't mental and neither is any concept identical to it[19]

Even if *knowledge* is a mental concept of the state of knowing, there might still be some necessarily coextensive non-mental concept of the mental state of knowing *true-belief + X*.

But there's good inductive evidence[20] that there is no such analysis.

General skepticism the programme of conceptual analysis:

Observation: Philosophically interesting concepts tend not to have analyses (e.g. *causes* or *means*). Why expect *knows* to be any different?

> 'Bachelor' is a peculiarity, not a prototype. (31)

Motivation for the project of analysis was something like Russell's epistemological programme: we're 'acquainted' with a small number of proposition-constituents, and everything that we understand is built up of those constituents.

With this as a background theory, analysis is a reasonable project. But for epistemologists who do not accept it, analysis is not a reasonable project.

1.4 Knowing as the Most General Factive Mental State

In this section, W offers a 'modest positive account' of knowing and then considers objections to it.

Interesting dialectical point:

> The [modest positive account] sketched below will appear thin by comparison with standard analyses. That may not be a vice. Indeed, its thinness will clarify the importance of the concept as more complex accounts do not. (33)

[18] same goes for concepts constructed using connectives other than conjunction

[19] and since any adequate conjunctive analysis of *knowledge* would include *true*, W is committed to rejecting all such analyses. Note that an analysis here is understood not just a necessarily coextensive concept, but as an identical one.

[20] in the form all of the failed analyses that followed Getter (1963)

The account:

> The proposal is that knowing is the most general factive stative attitude, that which one has to a proposition if one has any factive stative attitude to it at all. (34)

What's a factive stative attitude?

attitude it's a variety of propositional attitude

factive if you have a factive attitude towards p then p is true

stative stative attitudes are synchronic: they don't play out over time[21]

[21] contrast: processes, like forgetting, which are factive attitudes that aren't stative

Factive stative attitudes are realized in language by *factive mental state operators* (FMSO). For any FMSO Φ,

- Φ operates syntactically like a verb
- Φ is semantically unanalyzable[22]
  - stipulate that 'bluly' is synonymous with 'believes truly'.
  - Meaning of 'believes truly' is determined by the meanings of its parts[23], so 'bluly' isn't a FMSO
  - But, some FMSO's are syntactically analyzable

  (11) She felt that the bone was broken.

  (12) She could feel that the bone was broken.

    * 'could feel' in (11) is syntactically analyzable: it's composed to two words
    * but, it's *meaning* isn't composed of the meanings of those words; it's not *semantically* analyzable
      - Suppose the meaning of 'could feel' is determined by meaning of 'could' (in this sense meaning 'has the ability to') and feel, so 'could feel' means 'has the ability to feel'.
      - but that's not what 'could feel' means in (11), which is somethign like 'knows by sense of touch'
      - but 'knows by sense of touch' is (i) factive and (ii) perceptual, and 'has the ability to feel' is neither.
- subject of Φ is generally an animate object, and the object of Φ
  - bearing a factive attitude toward a proposition requires grasping that proposition, which requires possessing some concept for each of the component parts of the proposition

[22] i.e. there's not conjunction of other expressions that's synonymous with Φ

[23] i.e. its meaning is determined *compositionally*

  - Example: can't know that Olga is playing chess unless you have a concept of chess

- 'S Φ's that A' is factive: the inference from 'S Φ's that A' to 'A' is valid, not a mere cancellable presupposition
- FMSO's are stative:

(5) She is proving that there are infinitely many primes.

(6) The shoes are hurting her.

(7*) She is knowing that there are infinitely many primes.

(8*) She is believing that there are infinitely many primes.

(9*) The shoes are fitting her.

Proposal: if Φ is a FMSO then 'S Φ's that A' entails that 'S knows that A'

- i.e. knowledge is the most general factive mental state. Examples:
  - Seeing that it's raining is a way of knowing that it's raining[24]
  - Remembering that tomorrow is Friday is a way of knowing that tomorrow is Friday

[24] NB: seeing *that* it's raining isn't the same thing as seeing the rain

Summary of FMSO discussion:

(18) If Φ is an FMSO, from 'S Φ's that A' one may infer 'A'.[25]

(19) 'Know' is an FMSO.

(20) If Φ is an FMSO, from 'S Φ's that A' one may infer 'S knows that A.

[25] NB that W has switched (in both (18) and (19)) from talking about the *entailments* of statements involving FMSO's to talking about what *inferences* they support. There's a big difference between the two, so (18) and (20) don't really summarize what came before as much as they assert claims that are related but distinct.

Clarification. Compare:

claim knowing is the most general stative propositional attitude such that, for all propositions p, necessarily if one has it to p then p is true.

- simpler: knowing is the MGSPA such that: knowing p guarantees the the truth of p
- this attributes a property to knowing

*claim** for all propositions p, knowing p is the most general mental state such that necessarily if one is in it then p is true.

- simpler: every proposition p is such that: knowing that p is the most general mental state that guarantees the truth of p
- this attributes a property to each and every proposition

Counterexample to claim*: believing that 2+2=4. Believing is a more general mental state than knowing, and believing that 2+2=4 guarantees the truth of 2+2=4,[26] so there's a proposition that fails to have the property asserted in claim*,[27] so it's not the case that all propositions have that property

[26] because everything does

[27] that knowing that proposition is the most general mental state such that necessarily if one is in it then that proposition is true.

1.5 *Knowing and Believing*

Does knowing that A entail believing that A?

W's account might be taken as a reason to deny the entailment because it 'provides no basis for a conceptual connection between believing and knowing'.

But that's too quick: W denies that *believes* is part of a conjunctive analysis of *knows*, but that's not the only type of necessary connection

Alternative proposal:

> ...reverse the direction of analysis, and validate [the thesis that *knows* entails *believes*] by an analysis of believes in terms of knows. The simplest suggestion is that the concept believes is analysable as a disjunction of knows with other concepts. The word 'opine' will be used here as a term of art for the rest of the disjunction. On this analysis, one believes *p* if and only if one either knows *p* or opines *p*. Given that opining *p* is incompatible with knowing *p*, it follows that one opines *p* if and only if one believes *p* without knowing *p*. (44)

Problem: if *believes* $=_{df}$ *knows* $\vee$ *opines*, then it must be possible to grasp *opines* without previously grasping *believes*[28]

[28] recall that on the general picture of analysis that W is considering, we grasp complex concepts by grasping their component parts

opines is (in this context) a made-up technical term. How should we understand it?

First understanding:

opine $=_{df}$ *believes* & $\neg$*know* (44)

Problem: in that case the analysis of *believes* is circular:

believes $=_{df}$ *knows* $\vee$ (*believes* & $\neg$*know*)

Second understanding:

one opines that *p* $=_{df}$ one is in a state which is, for all one knows, knowing *p* (45)

So the analysis becomes:

believes $=_{df}$ *knows* $\vee$ *is in a state which is, for all one knows, knowing p*

Problem: if you can't grasp p then you're in a state which is, for all you know, knowing p. So by this analysis you believe p. But if you can't grasp p then you don't believe p.

Third understanding:

> one opines $p =_{df}$ one has an attitude to the proposition p which is, for all one knows, knowing (45)

Problem: the task here is to find a disjunctive analysis of *believes*, and this isn't disjunctive

Fourth understanding:

> one opines $p =_{df}$ one is under the illusion that $p \vee$ one is irrationally certain that $p \vee$ ... [no doubt many disjuncts to follow] (45)

Problem: how could we possibly fill out this list without already knowing that it's a list of ways to believe without knowing, which of course presupposes possession of the concept *believes*[29]

[29] NB: this objection is rooted in a claim that there's something we can't plausibly *do*: we can't populate the list of ways to merely believe without already understanding that it's a list of ways to believe without knowing. Since the proposal is a conceptual analysis, if it's correct then our grasp of the long disjunction must precede our grasp of *believing*; W is claiming that this is implausible, so the analysis fails.

Zoom out for a minute. Above we've been searching for an a priori analysis of *believes* in terms of a disjunction of *knows* and something else. How about instead we search for a non-a priori necessary connection between the states that determine whether the concepts actually apply.

Metaphysical proposal:

> one believes p if and only if one is in either the state of knowing p or the state of opining p (46)

This doesn't imply a claim about what what we can do, it just says that to be in the state of believing is to be in the state of knowing or the state of opining. Knowing and opining, then, are either unified (i.e. metaphysically unanalyzable) states, or they're composed of other states that are unified. Once the metaphysical analysis is complete we have believing explained in terms of a long disjunction of unified states, and being in any one of those states is sufficient for believing.

Problem: the presupposition of an attempted analysis of believing in terms of something else is that believing is not a unified mental state, it's a composite. But why think that believing is a unified state, but that opining (or the component sates that comprise opining) isn't? The proposal is unmotivated.[30]

[30] This is a strange line of argument. The motivation for this section is to explain why knowing entails believing even though we've rejected the Knowledge $=_{df}$ true-belief + X analysis. Goal is to instead provide a Belief $=_{df}$ Knowledge $\vee$ X analysis. Maybe the objection is that it's not *independently* motivated?

Intermediate conclusion: 'A strictly disjunctive account of belief is not correct at either the conceptual or the metaphysical level.' (46)

So why does knowing entail believing?

First non-disjunctive proposal:

> ...one believes *p* if and only if one has an attitude to the proposition *p* which one cannot discriminate from knowing, in other words, an attitude to *p* which is, for all one knows, knowing. (46)

W: this is a 'reasonable approximation' of an analysis[31], but it both over- and under-generates:

[31] Is a reasonable approximation of an analysis a type of analysis, or is it a plastic fish?

false positive: a creature who lacks the concept of knowing can't distinguish desiring that *p* from knowing that *p*, so by this 'analysis' desiring entails believing (which it doesn't)

false negative: believing *p* as a leap of faith is a type of believing that *p*, but it's subjectively distinguishable from knowing that *p*, so on this analysis it isn't a type believing

Second non-disjunctive proposal:

> to believe p is to treat p as if one knew p-that is, to treat p in ways similar to the ways in which subjects treat propositions which they know (46-7)

Important way they treat propositions which they know: they use them as premises when reasoning.

Important: this proposal falls short of a 'full-blown exact conceptual analysis of belief'; it's 'rough', a 'crude generalization'.

If this is correct then knowing is central to believing because it 'sets the standard of appropriateness for belief'. (47) This idea will be developed more W's chapters on evidence.

2 - Broadness

2.1 *Internalism and Externalism*

<u>Three varieties of internalism (and externalism)</u>

First variety: *Epistemic* Internalism is a thesis about what sorts of factors determine the justificatory status of a belief[32]

[32] Or a proposition, depending on whether we're talking about doxastic justification or propositional justification.

> *Epistemic Internalism:* the justificatory status of S's beliefs is completely determined by factors internal to S
>
> *Epistemic Externalism* is the negation of Epistemic Internalism; it's the thesis that: a least one of the factors that determines the justificatory status of S's beliefs is external to S external to S

What does it mean for a factor to be epistemically 'internal' to S? Two senses:

> *Access Internalism* is the thesis that the justificatory status of S's beliefs is determined by factors that are accessible to S[33]
>
> *Access Externalism* is the negation of Access Internalism

[33] typically 'accessibility' is understood as requiring that, if factor F obtains, then S is in a position to know via introspection that S obtains

> *Mentalist Internalism* is the thesis that the justificatory status of S's belief is completely determined by S's mental states[34]
>
> *Mentalist Externalism* is the negation of Mentalist Internalism.

[34] Strong versions of Mentalist Internalism require that the contributing factors be restricted to S's *present* mental states

Second variety: *Mental State* Internalism is a thesis about what factors determine the mental states that agents are in.

> *Mental State Internalism* is the thesis that an agent's mental states are completely determined by their physical states
>
> *Mental State Externalism* is the negation of Mental State Internalism

Third Variety: *Mental Content* Internalism is a thesis about what factors determine what our mental states are about (i.e. what factors determine the content of our mental states)

> *Content Internalism* is the thesis that the content of S's mental states is completely determined by S's physical states.
>
> *Content Externalism* is the negation of Content Internalism

Throughout the book Williamson advocates for mental state externalism and mental content externalism, but by the internal logic of his project, epistemically speaking he's actually a mentalist internalist.[35]

[35] Much more on this later.

Mental state internalism is inconsistent with the thesis that knowing is a mental state: because knowledge is factive, S's knowing that it's raining is partially determined by the fact that it's raining. Facts about the weather aren't facts about S's physical states, so facts about what the S knows aren't determined by S's physical states. Mental state internalism is the thesis that S's mental states are completely determined by S's physical states, so if mental state internalism is true then knowing isn't a mental state.

Mental state internalists often pursue a reductionist programme for knowledge: break down S's knowing[36] that *p* into components that depend only on S's physical states (belief that *p*, further factors concerning evidence and justification) and those that don't (truth of *p*).

[36] W is talking about the state of knowing rather than the concept *knows*; the reductionist programme is a metaphysical programme.

Mental state externalists often pursue a similarly reductionist programme, but file some 'further factors'[37] under the components of knowing that aren't completely determined by S's physical states.

[37] e.g. putative partial determinants of justification, such as the causal relationship between the fact believed and the fact itself, or the reliability of the belief-forming faculty, etc.

The reductionist programme has failed [no further argument at this point], indicating that this programme is misconceived.

To the extent that internalism suggests the reductionist programme, this failure indicates that internalism too is misconcieved.

2.2 *Broad and narrow conditions*

Some terminology:

A **case** is a possible total state of a system, the system consisting of an agent at a time paired with an external environment, which may of course contain other subjects.[38]

[38] A case is like a possible world, but with a distinguished subject and time: a 'centred world'

A **condition** obtains or fails to obtain in each case. Conditions are specified by 'that' clauses.

A condition C **entails** a condition D if and only if for every case α, if C obtains in α then D obtains in α.[39]

[39] The conditions C and D are identical if and only if for every case α, C obtains in α if and only if D obtains in α.

A **case** α is **internally like** a case β if and only if the total internal physical state of the agent in α is exactly the same as the total internal physical state of the agent in β

A condition C is **narrow** if and only if for all cases α and β, if α is internally like β then C obtains in α if and only if C obtains in β.[40]

[40] In other terminology, narrow conditions supervene on or are determined by internal physical state: no difference in whether they obtain without a difference in that state.

C is **broad** if and only if it is not narrow.

A **state** S is **narrow** if and only if the condition that one is in S is narrow; otherwise S is broad.

[Mental State] **Internalism** is the claim that all purely mental states are narrow; externalism is the denial of internalism.

Claim (from TW): We often attribute broad mental states to ourselves and to each other

Data:

- 'S sees Naples': there could be two internally alike cases α, β s.t. in α S is seeing Naples and in β S is hallucinating, and hence not seeing, Naples
- 'S believes that there are tigers': there could be two internally alike cases α, β where in α S's belief is about tigers, but in β there are no tigers, only schmigers[41], and hence in β S has had no causal contact with tigers and for that reason can't have mental states with tiger-contents

[41] schmigers look just like tigers but they're a distinct species

Important burden-shifting:

This language appears to attribute broad mental states, but Internalists claim that all mental states are narrow. So internalists must argue that the broad-state language 'fails to reflect the structure of the underlying facts'; they must 'isolat[e] a level of description of contentful attitudes that is both narrow and genuinely mental' (54)

In other words, internalists must argue that the broad language in mental state attributions really attributes only narrow conditions to agents, which requires them to describe, e.g. the narrow condition(s) that S is in when the sentence 'S sees Naples' is true.

The burden of proof is on the internalists to pull this off, and until they do we have a good reason to reject internalism. (54)

2.3 Mental differences between knowing and believing

Williamson's rigorous reconstruction of the Internalist's argument that that knowing isn't a mental state:

1. Suppose for reductio that: for every proposition p, there is a mental state S such that in every case α, one is in S if and only if one knows p.[42]
2. So, for all propositions p and cases α and β if one is in exactly the same mental state in α as in β then in α one knows p if and only if in β one knows p.[43]

[42] i.e., knowing is a mental state

[43] from (1)

3. For all cases α and β, if α is internally like β then one is in exactly the same mental state in α as in β.[44]

4. So, for all propositions *p* and cases α and β if α is internally like β then in α one knows *p* if and only if in β one knows *p*.[45]

[44] i.e. mental state internalism is true

[45] from (2) and (3)

But (4) is false: knowing that *p* requires the truth of *p*, and typically that's not determined by one's internal states

So, either (1) or (3) must be rejected. (3) is a core commitment of internalism, so internalists reject (1), inferring that knowing is not a mental state.

The Internalist Reductionist Programme:

Internalists also seek to factorize knowing into internal and external components. Williamson characterizes the internalists position as the thesis that 'knowing *p* adds nothing mental to believing *p*'. How to state this thesis rigorously?

First attempt:

5. For all propositions *p* and cases α, if in α one believes *p* then in some case β one is in exactly the same mental state as in α and one knows *p*.

> For if (5) is false, one can believe *p* while in a total mental state T incompatible with knowing *p*; but then the information that one knows *p* adds something mental to the information that one believes *p*, for it implies that one is in a total mental state other than T. Thus if knowing p adds nothing mental to believing *p*, then (5) holds. Conversely, if (5) holds, then knowing *p* imposes no constraints on one's mental state beyond those already imposed by believing *p*, so knowing *p* adds nothing mental to believing *p*. (56)

But this formulation is problematic *for reasons independent of internalism.*

Problem with the first attempt: the fact that *p* is false in α might be sufficient to ensure that it's false in β:

- Example: subject has a false belief about their own mental state in α.[46]
- Example: subject believes a necessary falsehood.[47]

[46] Since *p* is a proposition about the subject's mental state, and her mental states are the same in α and β, *p* is true in both or neither.

[47] Anything at all is sufficient to ensure the falsity of any necessarily false proposition.

These types of counterexamples both involve false beliefs, so maybe we should limit ourselves to cases of believing truly:

Second attempt:

6. For all propositions *p* and cases α, if in α one believes *p* truly then in some case β one is in exactly the same mental state as in α and one knows *p*.

First problem with the second attempt: maybe subject doesn't know because she believes irrationally

Problem with the first problem: epistemic externalists don't think rationality facts supervene on the subject's mental states. So an epistemic externalist who's a mental state internalist might be able to avoid this charge.

Second problem for the second attempt: the goal here is to isolate a purely mental component of knowing. Believing truly isn't purely mental.

Third attempt:

7. For all propositions *p* and cases α, if in α one rationally believes *p* then in some case β one is in exactly the same mental state as in α and one knows *p*.

NB: this formulation is meant to appeal to mental state internalists who are content externalists

Problem with the third attempt: content externalists can't hold that believing rationally is an internal mental state because 'content externalism makes rational belief an externalist mental attitude'. (58)

Example: I hallucinate a barking dog, come to believe 'that dog is barking', and on that basis infer 'a dog is barking'. According to the content externalist, 'that dog' doesn't refer to anything, so 'that dog is barking' doesn't express a proposition, so my inference to 'a dog is barking' is unsupported. Hence my rational belief is a broad mental state. But then there's no in-principle problem for states that are broad and mental, so what's the problem with knowledge as a mental state?

NB: TW's presentation on 58 is very unclear, and I'm not sure that I've reconstructed it correctly. In particular, he doesn't seem to be offering a counterexample to (7) like he did to (5) and (6)

- in his example the agent doesn't believe rationally, so the antecedent of the conditional is false, so the conditional is trivially true

Fourth attempt:

8. For all propositions p and cases α, if in α one believes p then in some case β one is in exactly the same mental state as in α and one does not know p.

   For if (8) is false, someone can believe p while in a total mental state T incompatible with not knowing p; but then not knowing p adds something mental to believing p, as the former but not the latter is sufficient, given that one believes p, for being in a total mental state other than T. Thus if not knowing p adds nothing mental to believing p, then (8) holds.

   Problems with the fourth attempt:

- If p is about my own mental state — e.g. 'I'm in pain' — then internalists will say that believing p implies knowing p.
- If p is a necessary truth — e.g. '79 + 89 = 168' — then internalists will say that grasping a proof of p is sufficient for both belief and knowledge.

Conclusion: the burden of proof is on those who claim that there's a purely mental component of knowing — Internalists — and that burden has not been met.

2.4 *The causal efficacy of knowledge*

Alternative motivation for internalism:

1. genuine states are causally efficacious
2. mental states are causally efficacious only if they're narrow
3. factive states (like knowing) are not narrow
4. so, factive states (like knowing) are not causally efficacious
5. so, factive mental states (like knowing) are not genuine states[48]

[48] presumably TW means here that factive mental states (like knowing) are not genuine states in the sense that they're artificial constructions of other states

Standard content externalist response: attributions of broad content do play an essential role in causal explanations *when the explanandum is itself characterized in broad terms*. Example:

> Consider a causal explanation as simple as 'He dug up the treasure because he knew that it was buried under the tree and he wanted to get rich'. Note that the explanandum ('He dug up the treasure') makes reference to objects in the environment (the treasure) as well as to the subject's immediate physical movements. The internalist cannot substitute 'believe' for 'know' in the explanation without loss, for the revised explanans, unlike the original, does not entail that the treasure was where he believed it to be; the connection between explanans and

> explanandum is therefore weakened. **The explanans does less to raise the probability of the explanandum.** (61-2; emphasis added)

Internalists might want to substitute 'believes truly' for 'believes' in the above passage, but that sometimes involves a loss of explanatory power:

Example:

Explanandum: a burglar enters a house and ransacks it for a long time.

Explanans 1: he knew that there was a diamond in the house

Explanans 2: he believed truly that there was a diamond in the house

Methodological point: the bold sentence from above suggests that, all else equal, the strongest explanation is the one that raises the probability of the explanandum more than any competing explanans

Another way to put the thought: the better explanation is the one that's more strongly correlated with the evidence.

Burglar might believe truly that there's a diamond in the house on the basis of an inference from a false premise. Example: the premise is 'the diamond is under the bed', when really it's in the drawer.

In that case you'd expect the burglar to give up looking once he sees that the diamond isn't under the bed.

Contrast: *knowing* that there's a diamond in the house on the basis of an inference from a false premise is impossible, so the bed/ drawer case is impossible. Hence:

P(ransack | knowledge) > P(ransack | true belief)

So the evidence (= the fact that he ransacked the house for a long time) supports the *knows* hypothesis more strongly than the *believes truly* hypothesis.

Big-picture objection to this section:

There's more to the strength of an explanation than just probability raising/ strength of correlation.

Know what's *really* strongly correlated with the burglar ransacking the house all night? The burglar ransacking the house all night. Perfect correlation. Hence:

P(ransack | ransack) $\gg$ P(ransack | knowledge) > P(ransack | true belief)

That's a super-crappy explanation, and Internalist might accuse TW of pulling a similar trick.

Presumably internalists are looking for a *mental state* that explains behavior, and TW gives them *knowledge*, which from their perspective is a conjunction of a mental state and a non-mental state. But if that's allowed then why not choose a conjunction of the mental state: **believes there's a diamond in the house** and the non-mental state: **ransacked the house all night**?

P(ransack | believes & ransack) $\gg$ P(ransack | knowledge)

Presumably TW's response is: because **knowing that there's a diamond is in the house** is a mental state but **believing there's a diamond in the house & ransacking the house all night isn't**. But that's precisely the point of contention between TW and the internalists that this section was supposed to help us to adjudicate.

Who is begging the question here?

3 - Primeness

Main point of chapter: argue that the internalist's project of factoring states into internal and external components is impossible

3.1 – Prime and Composite Conditions

More terminology:

A **case** is a possible total state of a system, the system consisting of an agent at a time paired with an external environment, which may of course contain other subjects.[49]

A **condition** obtains or fails to obtain in each case. Conditions are specified by 'that' clauses.

A condition C **entails** a condition D if and only if for every case α, if C obtains in α then D obtains in α. [50]

A case α is **internally like** a case β if and only if the total internal physical state of the agent in α is exactly the same as the total internal physical state of the agent in β

A condition C is **narrow** if and only if for all cases α and β, if α is internally like β then C obtains in α if and only if C obtains in β.[51]

C is **broad** if and only if it is not narrow.[52]

* A condition C is **environmental** if and only if for all cases α and β, if α is externally like β, then C obtains in α if and only if C obtains in β[53]

* A condition C is **composite** if and only if it is the conjunction of some narrow condition D with some environmental condition E[54]

* C is **prime** if and only if it is not composite

A **state** S is **narrow** if and only if the condition that one is in S is narrow; otherwise S is broad.

Observation: All narrow conditions are composite

- any narrow condition can be conjoined with an environmental condition that always obtains, e.g. that nothing is traveling faster than light

TW seems to thing that this is important and embarrassing for the internalist, but it's just a trivial consequence of his stipulative definitions. Note also that:

[49] A case is like a possible world, but with a distinguished subject and time: a 'centred world'

[50] The conditions C and D are identical if and only if for every case α, C obtains in α if and only if D obtains in α.

[51] In other terminology, narrow conditions supervene on or are determined by internal physical state: no difference in whether they obtain without a difference in that state.

[52] This is merely the denial of supervenience claim of narrowness; it's satisfied any time there exist a pair of internally alike cases α and β s.t. C obtains in one but not both

[53] In other terminology, environmental conditions supervene on or are determined by the physical state of the external environment

[54] As a special case, a narrow mental condition is trivially composite, for it is the conjunction of itself with the environmental condition that holds in all cases whatsoever

Observation: All environmental conditions are composite

- any environmental condition can be conjoined with a narrow condition that always obtains, e.g. that the agent is self-identical
- by definition of 'case', each case includes an agent, and all agents are self-identical, so the condition 'that the agent this case is centered on is self-identical' is always true

Again, this is a trivial consequence of TW's stipulative definitions, so it's not philosophically interesting. What's interesting is whether a particular condition is composite in a non-trivial way.

3.2 – Arguments for Primeness

Once again, TW's target is internalism. He thinks that internalists are likely to embrace the following line of reasoning:

> The causing of my present action is here and now. Only narrow conditions supervene on the here and now; so narrow conditions must play a privileged role in the causal explanation of action. If a causal explanation of action cites a broad mental condition, an underlying narrow condition must do the real work. **We can isolate that narrow condition by subtracting from the broad mental condition the environmental accretions that make it broad.** We can recover the original broad mental condition from the narrow condition by adding back those accretions. (65 – emphasis added)

In this section he attacks the sentence in bold. If it's false, then internalism is hopeless.

TW's argument is really complicated. The basic idea is this: internalists and externalists agree that knowing is a broad condition, but they disagree about whether knowing is the conjunction of a narrow condition and an environmental condition, i.e. about whether knowing is prime or composite.

Suppose knowing that some dogs bark is a conjunction of a narrow condition **believing with justification that some dog barks** and an environmental condition **some dog barks**.

There are lots of ways one could satisfy the narrow condition: you can hear a dog barking, or be told of a dog that barks, etc. There are also lots the world might be such that the environmental condition that some dog barks obtains: Fido barks, or Spot, or Spike.

<u>TW's crucial assumption:</u> if **knowing that some dog barks** is composite, then it should be the case that any time you somehow

satisfy both the narrow and environmental condition, you know that some dog barks. If not, then **knowing that some dog barks** isn't composite: it's prime.

In more detail, here's how to show that a broad condition C is prime:

Broad conditions entail both narrow and environmental conditions.

First, define **virtual-C** as the strongest narrow condition entailed by C, i.e. the narrow condition that obtains any time C obtains

Definition: let virtual-C be the condition which obtains in a case α if and only if C obtains in some case internally like α

Observation 1: C entails virtual-C

Proof: Suppose that C obtains in α itself. The 'internally like' relation is reflexive[55], so αis internally like α, so C obtains in some case internally like α, so the right side of the biconditional is true, so by the definition of virtual-C the left side must be true as well: virtual-C obtains in α.

[55] R is reflexive $=_{df} \forall_x(R_{x,x})$

Observation 2: virtual-C is narrow

Proof: Suppose that virtual-C is *not* narrow, i.e. suppose that there is a a pair of internally alike cases s.t. virtual-C obtains in one but not both. Suppose also that virtual-C obtains in α. Then C obtains in some β internally like α (by the definition of virtual-C, left to right direction). By Observation 1, virtual-C obtains in β. So α and β aren't the pair of cases we're looking for. Now consider some case γ which is internally like β. By the symmetry[56] of the internally like relation, if γ is internally like β then β is internally like γ. By the right-to-left direction of the definition of virtual-C, since C obtains in a case internally like γ (because it obtains in β), virtual-C obtains in γ. So γ and β aren't the pair of cases we're looking for. There's nothing special here about γ , so parallel reasoning establishes that virtual-C obtains in every case internally like β (and hence internally like α (by transitivity[57])). So the pair of cases we're looking for doesn't exist, so virtual-C is a narrow condition.

[56] R is symmetric $=_{df} \forall_{x,y}(R_{x,y} \rightarrow R_{y,x})$

[57] R is transitive $=_{df} \forall_{x,y,z}(R_{x,y} \& R_{y,z} \rightarrow R_{x,z})$

Observation 3: virtual-C entails every narrow condition that C entails

Proof: if C entails a narrow condition D, and virtual-C obtains in a case α, then C obtains in some case β internally like α, so D obtains in β (since C entails D), so D obtains in α (since D is narrow); hence virtual-C entails D

Lesson: virtual-C is the strongest narrow condition which C entails

We can also define:

Definition let outward-C be the condition which obtains in a case α if and only if C obtains in some case externally like α.

By parallel reasoning we can also establish that:

Observation 1*: C entails outward-C

Observation 2*: outward-C is environmental

Observation 3*: outward-C entails every environmental condition that C entails

In sum: just as virtual-C is the strongest narrow condition which C entails, so outward-C is the strongest environmental condition which C entails.

Why this matters:

- All sides agree that knowing is a broad condition
- Internalists also claim that knowing is a composite condition: it's a conjunction of narrow and environmental conditions. This puts the burden of proof on the internalist to produce the narrow condition
- The schematic form of this factoring will be to take knowing as our C and then identify the relevant virtual-C and outward-C: the strongest narrow condition and the strongest environmental condition that are each entailed by C
- If C is any conjunction of narrow and environmental conditions at all, then it is the conjunction of virtual-C and environmental-C.
  - **Proof:** Let C be the conjunction D & E of a narrow condition D and an environmental condition E. Since C entails D, virtual-C entails D; similarly, outward-C entails E. Thus the conjunction of virtual-C and outward-C entails D & E, that is, C.
- But the mere fact C entails conjunction of virtual-C and outward-C *does not show* that C is composite: all broad conditions do that, regardless of whether they're prime or composite

So: to show that C is composite, construct a case in which virtual-C and outward-C obtain but C does not obtain. Here's the recipe:

1. Identify three cases α, β, and γ, s.t.

(a) C obtains in both α and β[58]

(b) γ is internally like but externally unlike α

- by Observation 1 above, since C obtains in α, virtual-C obtains in α as well
- by Observation 2 above, since virtual-C obtains in α and α is internally like γ, virtual-C obtains in γ

(c) γ is externally like but internally unlike β

- by Observation 1* above, since C obtains in β, outward-C obtains in β as well
- by Observation 2* above, since outward-C obtains in β and β is externally like γ, outward-C obtains in γ

2. For any α, β, and γ that satisfy these conditions, both virtual-C and outward-C obtain in γ
3. Note that if C is composite the C *must* obtain in γ as well[59]
4. Note that if C is prime than C *might not* obtain in γ

So, if you want to show C is not composite (i.e. that C is prime), satisfy the above conditions and be sure that C doesn't obtain in γ

One of TW's examples:

> Let α be a case in which one knows by testimony that the election was rigged; Smith tells one that the election was rigged, he is trustworthy, and one trusts him; Brown also tells one that the election was rigged, but he is not trustworthy, and one does not trust him. Let β be a case which differs from α by reversing the roles of Smith and Brown; in β, one knows by testimony that the election was rigged; Brown tells one that the election was rigged, he is trustworthy, and one trusts him; Smith also tells one that the election was rigged, but he is not trustworthy, and one does not trust him. Now consider a case γ internally like α and externally like β. In γ, one does not trust Brown, because one does not trust him in α, and γ is internally like α. Equally, in γ, Smith is not trustworthy, because he is not trustworthy in β and $\gamma$is externally like β. Thus, in γ, neither Smith nor Brown is both trustworthy and trusted. We can legitimately assume that in none of the three cases does one have any other way of knowing that the election was rigged. Consequently, in γ, one does not know that the election was rigged. Yet, in α and β one does know that the election was rigged. Thus the condition that one knows that the election was rigged is prime. Since the example does not turn on the specific content of the knowledge, it can be modified to show for almost any proposition *p* that the condition that one knows *p* is prime. (72)

Crucial features of the case:

[58] "...we may assume that such a case is possible because otherwise that interdependence of the internal and the external would itself undermine the idea that they can be separated (see section 3.3 for more on this assumption)" (67)
NB:
i. This assumption is essential to TW's argument (see point 3 below)
ii. It's not obviously true: let's be sure to look carefully at §3.3

[59] After all, C just is the conjunction of virtual-C and outward-C, and both conjuncts obtain

Case	Narrow condition	Env. condition	Knows?
α	**Believe on Smith's testimony**	Smith trustworthy	Yes
β	Believe on Brown's testimony	**Brown Trustworthy**	Yes
γ	**Believe on Smith's testimony**	**Brown Trustworthy**	No

In each case:

- C = one knows that the election was rigged
- virtual-C = one believes rationally that the election was rigged
- outward-C = one has received trustworthy testimony that the election was rigged

Virtual-C and outward-C are satisfied in γ, but in γ one fails to know

So, knowing is prime.

3.3 – Free Recombination

Free recombination: given cases α and β there is a case γ internally like α and externally like γ[60]

[60] NB that this is the essential assumption flagged in footnote 58

But there are independent reasons to doubt that there is always a possible case combining the internal conditions of α and the external conditions of β

Example: there might be 'nomic constraints' between narrow and environmental conditions that prevent certain recombinations

TW's discussion here is slightly complicated (especially his discussion of Fodor), but we can ignore all that because his simple dialectical point is obviously correct:

> . . . counterexamples to free recombination are really a problem for the attempt to separate the internal and external" (73)

This for two reasons:

First: it's the internalist who thinks that internal and external conditions are neatly separable, and the failure of free recombination suggests that they aren't so separable after all (due to nomic connections or whatever).

Second: suppose Free Recombination fails *as a principle* because there are counterexamples arising from nomic constraints that exist between the internal and the external.

Who cares? The present debate is about whether the narrow and environmental conditions entailed by factive mental states — and knowing in particular — can freely recombine. And I don't see any reason to think that they can't.[61]

Similar considerations arise when considering doubts about whether there's a sensible distinction to be drawn between the internal and the external

[61] In fact, lots of work in epistemology (including much in the internalist tradition) supposes that they can: NED scenarios, Cartesian evil deceiver scenarios. (The cartoon version of those scenarios you see in movies like The Matrix and that most philosophers take for granted, not the third source of skeptical doubt that Descartes actually describes in M1)

3.4 – The Explanatory Value of Prime Conditions

Point of this section: to argue that prime conditions have explanatory value for the same reason that broad conditions do[62]

[62] see §2.4

We need broad conditions to explain action over timespans extending beyond the immediate future:

> One is thirsty; how likely is one to be drinking soon? Likely enough, if one sees water. Much less likely, if what one sees is a mirage. (75)

Unpacking the example: take our explanandum to be that S is drinking water and the candidate explanations to be that S sees water and that S hallucinates water.

Drinking water is more closely correlated with seeing water than it is with non-veridically hallucinating water: Drinking water and seeing water entail that there's some water nearby, but hallucinating water does not. So seeing water is the better explanation.[63]

[63] See §2.4

NB: in this example, the explanatory advantage of S seeing water comes from the fact that it implies the *environmental condition* that there's water nearby; it depends on the broadness of the condition that S sees water. In contrast, that S hallucinates water is narrow: it entails no environmental conditions[64], so it doesn't entail that there's water nearby.

[64] That's too strong actually, but only for reasons that don't affect the argument under discussion so let's let it pass

So, composite conditions can have explanatory advantages over narrow conditions.

Recall that there are two kinds of broad conditions: prime and composite.

TW: prime conditions often have explanatory advantages over composite conditions:

> The relevance of seeing water now to drinking soon is not exhausted by the agent's internal state and the presence of water.[65] Before one can drink the water, one must get oneself to it. Typically, one will steer one's way by keeping the water in sight and making a complex

[65] i.e. by the mere conjunction of a narrow condition and an environmental condition

> series of adjustments to one's position in a feedback loop. The present **coincidence** of one's internal physical state with the state in which one would be if one saw the water from that perspective is not enough; the **coincidence** must continue until one reaches the water. The kind of causal relation in which one stands to something when one sees it often enables one to keep it in sight. By contrast, if the matching of internal state and external environment is mere coincidence, then there is no reason why it should continue. (76-7, emphasis added)

If the relationship between S's internal physical state and the state of the environment more is a mere coincidence, then the resulting *broad* condition is merely *composite*: it's merely the conjunction of the internal condition of S's seeming to see water and the external condition that there's water nearby.

If that relationship is more than a mere coincidence — if the broad condition in question is more than a mere conjunction of S's seeming to see water and there being water nearby — then the resulting *broad condition* is *prime*.

So, of the varieties of broad conditions, prime conditions are more explanatorily powerful than composite conditions

Interesting application of this idea:

Question: for someone who wants to go to Larissa, why is knowing the way more valuable than believing truly the way?

TW's suggestion: to reach Larissa, you need to maintain your true belief along the way. True belief is more likely to be abandoned in the face of contrary evidence than knowledge, and if you stop believing then you probably won't reach Larissa.

TW generalizes the point:

> If your cognitive faculties are in good order, the probability of your believing p tomorrow is greater conditional on your knowing p today than on your merely believing p truly today. . . Consequently, the probability of your believing p tomorrow is greater conditional on your knowing p today than on your believing p truly today. (79)

I'm inclined to agree, but then he continues:

> . . . profoundly dogmatic beliefs which are impervious to future evidence and do not constitute knowledge may be even more likely to persist than beliefs that are rationally sensitive to future evidence and do constitute knowledge, **but then the subject's cognitive faculties are not in good order**. Since the difference between your present knowledge and your present true beliefs matters for predicting your future beliefs, it matters for predicting your future actions, because they will depend on your future beliefs. (79 emphasis added)

Yes, agents with profoundly dogmatic beliefs are cognitively deficient. But the point of this section is to argue that prime conditions like knowing are more explanatorily powerful than composite conditions like believing dogmatically.[66] How is the cognitive deficiency of the agent relevant to that question?

[66] recall that TW thinks that all narrow conditions are trivially composite, since they are coextensional with that narrow condition and a logical truth

Big picture observation: TW thinks knowing is explanatorily powerful because in part because it entails certain facts about the environment (e.g. that there's water nearby) as well as certain facts about how the agent will respond to defeaters for what they know. But why can't we just specify those independently as a conjunction of environmental and narrow condition, i.e. as a composite condition?

Exegetical hypothesis: though TW starts out the section talking about the relative explanatory value of prime and composite conditions, his real point is that there's no plausible candidate of virtual-C that can match knowing in explanatory value. True, dogmatic belief might be more strongly correlated with certain behaviors in isolated circumstances (those in which the agent is cognitively deficient) but those circumstances are rare, and so the explanation is often unavailable.

This reading is consistent with TW's further claim that:

> What the argument does suggest is that when a condition stated in non-circular terms (belief, truth, justification, causation, . . .) fails to be necessary and sufficient for knowledge, that divergence will yield a divergence in implications for future action; the task of stating non-circularly a condition equivalent to knowledge with respect to implications for future action is no easier than the task of stating non-circularly a condition necessary and sufficient for knowledge.

But now I'm confused about TW's comments on the road to Larissa. That's a puzzle about what makes knowing *more valuable* than believing truly, and TW's response is that knowing is valuable because it's robust in the face of defeaters[67]. But dogmatic true belief is even more robust in the face of defeaters. So dogmatic true belief is more strongly correlated with reaching Larissa.

[67] which would be misleading defeaters in this case, as the belief in question is veridical

3.5 – The Value of Generality

<u>An objection to the explanatory value of prime conditions:</u>

Consider the explanatory value of prime condition C in case α

Let D be the strongest narrow condition in α

Let E be the strongest environmental condition in α

Then D&E specifies the total physical state of α

So (plausibly) D&E entails C

So any explanatory value in C is redundant once we've already specified D and E

TW's response:

D&E is much less general than C: there will be lots of cases in which C obtains but C&D doesn't[68]

Generality is a theoretical virtue for explanations

Examples:

- ... "one can explain why someone died by saying that he was run over by a bus; the explanation becomes worse, not better, if one specifies that the bus was red, for its colour had nothing to do with his death." (81)
- "If all metals have a certain property, one will be unhappy with attempts to explain why gold has it which cite properties of gold not shared by other metals." (81)

More formally: according to TW, one (of many) theoretical virtue for an explanans is correlation with the explanandum

Spelling this out precisely is tricky,[69] but here's one theoretical virtue that TW is committed to: the closer that E approaches being a necessary condition[70] for O, the better E explains O. Better still: E's exhibiting a high degree of necessity for O is a good-making feature of E as an explanation of O.

A hyper-specific explanans achieves a low degree of necessity, because it leaves out lots of other, distinct explanans for the explanandum:

> Consider again the cases that demonstrated primeness: α and β were mutually symmetric; the best explanation of the agent's subsequent actions might well generalize across α and β citing a condition that obtains in both. But if it also obtained in γ, the explanation would be weakened, for then the cited condition would not rule out a range of cases in which the agent's subsequent actions in α and β are much less likely.[71] (82)

A hyper-general explanans would also be problematic, this time because it exhibits a low degree of sufficiency: that 2+2+4 is maximally general (it obtains in every case) and is sufficient only for other necessary conditions.

[68] But not vice versa: since D&E completely specifies the physical state of the universe, it obtains only in α

[69] After all, if you're just looking for a perfect correlation between explanans and explanandum then just take the disjunction of all the strongest narrow and environmental conditions (e.g. C&D) that are present in any case that also includes the thing you want to explain. So your explanans looks like this: $(C\&D) \vee (C'\&D') \vee (C''\&D'')\ldots$ But that's a crappy explanation, so there are lots of other relevant considerations (as TW acknowledges).

[70] Sounds weird, right? You want to yell: 'being a necessary condition doesn't come in degrees! Ack!' Unfortunate locution, but let's let it slide...

[71] In other words, because the explanans in α and β— C&D — also obtains in γ, and because in γ the explanandum does not obtain, C&D exhibits a lower degree of necessity relative to a prime condition that obtains in α and β, but not in γ.

What we're looking for here is a sweet spot balancing generality with specificity. But the nature of that sweet spot will depend on matters beyond degree of necessity and degree of sufficiency. For example, TW says:

> ... someone was crying because she was bereaved, it does not improve the explanation to say that she was crying because she was bereaved or chopping onions. (83)

As an explanation for crying, 'bereaved or chopping onions' has a higher degree of necessity than 'bereaved'[72] without sacrificing any on the degree of sufficiency front. So the former explanation is more strongly correlated with crying than the latter, but it's a worse explanation. This indicates that there's more to strength of an explanation than mere correlation.[73]

[72] because it leaves out fewer cases of crying for reasons not referenced in the explanation

[73] Again, TW is fully aware of this and is striving to be sensitive to it in the text. Two things: (i) let's be sensitive to it to, and (ii) is TW successful in discussing the importance of correlation in isolation of other, competing explanatory virtues without ignoring important ways that they are inseparable?

Big picture: this isn't an in-principle feature of reductive accounts: it's at least possible for a composite condition to be a fully necessary for some explanandum.

> Successful reductions involve no loss of generality. Something common to all genuine instances of the given phenomenon is identified in lower-level terms. The present argument does not undermine the explanatory value of those reductions. But that value is not shared by explanations which use no such generalization about the phenomenon, and merely provide — or rather gesture towards — a maximally specific description in lower-level terms of the particular case at hand. (82)

Recall that the section opens by considering a C and D that are the strongest narrow and environmental conditions that obtain in α. This is an example of a 'maximally specific description in lower-level terms'.

The broader point is this: reductive accounts aren't necessarily problematic, but *the type of reductive account imagined at the beginning of the section* is hopeless. Hence the explanatory power of prime conditions is not redundant upon the conjunction of the maximally specific narrow and environmental conditions in a case.

3.6 – Explanations and Correlation Coefficients

This section is technical and inessential for our purposes. The general point is to lay out a formal framework to illustrate what was described informally in §3.5.

3.7 – *Primeness and the Causal Order*

New problem: maybe the high correlation between knowing and acting isn't the result of knowing causing certain actions, but rather than there's a third cause of both knowing and acting. In that case the focus of our causal explanation should be on this third cause, not on knowing.

Moreover, we can define just such a condition:

> Given deterministic laws, we might define a present condition D perfectly correlated with the condition C that one will perform the action, by stipulating that D obtains in α case a if and only if the total present state of the system (agent and environment) in α and the deterministic laws entail that one will perform the action. C and D obtain in exactly the same nomically possible cases. Thus, if the laws have probability 1, C and D are perfectly correlated. . . (88)

TW: but who cares? We're not looking for a condition that "unifies the cases in which D obtains by what happens later", we're looking for a condition that unifies those cases by what's happening *now*.

> We seek a correlation between a condition given by a concept that unifies the cases in which it obtains in terms of the present state of the system and a condition given by a concept that unifies the cases in which it obtains in terms of the future state of the system; we are willing to sacrifice some degree of correlation in order to achieve such unification. (89)

For example, we seek a correlation between the current condition of knowing that there's water nearby and the future condition of drinking. Knowing that there's water nearby might not correlate as well as condition D (above). ğ But that's a price we're willing to pay for an explanation that unifies the right sorts of conditions, and D is not the right sort of condition.

TW's NB: the point of this section is to defend the causal efficacy of prime conditions. We've been mostly talking about *mental* prime conditions, but that's incidental: that the ship is anchored to the seabed is a non-mental prime condition, and it clearly has causal efficacy (it prevents the ship from drifting away).

3.8 – *Non-conjunctive Decompositions*

TW argued in §3.2 that knowing isn't a conjunction of a narrow and an environmental condition, and hence it's prime.

Since ‘prime’ is just defined as ’not composite’, this is a valid argument.

But that definition also leaves space for prime conditions that are non-conjunctive functions of narrow and environmental conditions. For example, D∨E is prime

Better example: a condition might be a long disjunction of conjunctions of narrow and environmental conditions: $(C\&D) \vee (C'\&D') \vee (C''\&D'')\ldots$ The fact that each disjunct is a composite condition doesn’t imply that the disjunction is composite (by the definition of ‘composite’, it isn’t).

NB: TW doesn’t mention it, but for each prime condition like knowing there will be another equivalent condition of this form.[74] Because conditions are individuated coarsely, these two conditions are numerically identical – they’re the same condition.

[74] The form of long disjunctions of conjunctions is often called ‘disjunctive normal’ form

TW: the lesson is that arguments for the *primeness* of knowing and other factive mental states shouldn’t be taken as arguments of unanalyzability of those factive mental states.

Let’s step and remind ourselves why any of this matters: the putative composite nature of knowing was thought to support the internalist picture of mind.

TW: the sort of logical constructions we’re considering now (constructions in disjunctive normal form) don’t support that project anyway:

> . . . the arguments for primeness are needed to fix the role of the mental in the causal explanation of action. For even if a mental condition C were a disjunction $(D_1\&E_1) \vee D_1\&E_1)\vee\ldots$ of conjunctions of non-trivial narrow conditions D_i with non-trivial matching environmental conditions E_i, it would not follow that C could be replaced in causal explanations by corresponding narrow and environmental conditions; a composite condition can be so replaced. Given free recombination, the strongest narrow and environmental conditions entailed by the disjunction are $D_1 \vee D_2\vee\ldots$ and $E_1 \vee E_2\vee\ldots$ respectively. But if $(D_1\&E_1) \vee (D_1\&E_1)\vee\ldots$ is prime, then it is not entailed by its composite consequence $(D_1 \vee D_2\vee\ldots)$ & $(E_1 \vee E_2\vee\ldots)$. Only the former requires one’s internal state to match the state of the external environment. When the causal explanation depends on the primeness of $(D_1\&E_1) \vee (D_1\&E_1)\vee\ldots$, as section 3.4 argued that it often will, the extractable narrow condition $D_1 \vee D_2\vee\ldots$ [i.e. virtual-C] typically plays no explanatory role; it is a sort of epiphenomenon. [But the internalist thinks that the explanatory role of virtual-C is essential.] What would give the narrow condition an explanatory role is compositeness, not analysability; the arguments for primeness therefore tell against such an explanatory role for the narrow condition. (91-2)

4 - Anti-luminosity

4.1 – *Cognitive Homes*

New reason to doubt that knowing is a mental state:

We have 'guaranteed epistemic access' to our own mental states;[75] we don't have guaranteed epistemic access to whether we know; so, knowing isn't a mental state.

[75] Our current mental states that is, when we are attending to the question

TW has argued already (§1.2) that many other paradigm mental states fail to be epistemically accessible in this way

But, it might be thought that there is a core of mental states that are epistemically accessible; call these the agent's 'cognitive home'

Point of this chapter: argue that agents have no cognitive homes filled with states to which we have some kind of special epistemic access (spelled out below);[76] we are cognitively homeless.

[76] At this point TW just tells us that, although we can make mistakes about whether we're in such a state, those mistakes are always rectifiable

4.2 – *Luminosity*

Purpose of this section is to clarify the property of *luminosity*

Conditions

Luminosity is a property of conditions.

Conditions are like *propositions* in that they're expressed by sentential clauses (e.g. that one is hungry, that one knows that it's snowing). Conditions obtain in cases, which are centered[77]

[77] centered on agents and times possible worlds

But, conditions are unlike propositions because parameters — person, place, other circumstances — are left unspecified. That one knows that it's snowing can be a condition instantiated by lots of different people at lots of different times and places. In this sense cases are more like *properties* than propositions.

Conditions are coarsely individuated by the cases in which they obtain: they are identical if they obtain in exactly the same cases.

Being in a position to know

If one is in a position to know p, and one has done what one is in a position to do to decide whether *p* is true, then one does know *p*.

Intuitively, I'm in a position to know that *p* iff there exists no obstacle blocking my path to knowing *p*.

Being in a position to know is factive: if one is in a position to know p, then p is true.

To be in a position to know p, it is neither necessary to know p nor sufficient to be physically and psychologically capable of knowing p.

Definition:

> A condition C is luminous $=_{df}$ in every case α, if in α C obtains, then in α one is in a position to know that C obtains.[78]

Illustration: being in pain is luminous iff anytime someone is in pain, they're in a position to know it

Examples of conditions that are sometimes claimed to be luminous: being in pain, using two words that have the same meaning, being appeared to as if A.

4.3 – An Argument Against Luminosity

TW proceeds by taking a (what we might have thought to be a) paradigm case of a luminous condition — that one feels cold — and arguing that it is not in fact luminous. Later (in §4.4) he'll generalize that argument to other putatively luminous conditions.

Quick summary of argument:

The argument is a reductio: suppose that the condition that one feels cold is luminous. Then anytime you satisfy that condition and you're reflecting on whether you satisfy it, you know that you do. TW describes a case of which we judge that the person feels cold but doesn't know it, thereby demonstrating that the supposition must be rejected, i.e. that the condition that one feels cold is not luminous.

TW's description of such a case:

> Consider a morning on which one feels freezing cold at dawn, very slowly warms up, and feels hot by noon. One changes from feeling cold to not feeling cold, and from being in a position to know that one feels cold to not being in a position to know that one feels cold. If the condition that one feels cold is luminous, these changes are exactly simultaneous.[79] Suppose that one's feelings of heat and cold change so slowly during this process that one is not aware of any change in them over one millisecond. Suppose also that throughout the process one thoroughly considers how cold or hot one feels. One's confidence that one feels cold gradually decreases. One's initial answers to the question 'Do you feel cold?' are firmly positive; then hesitations and qualifications creep in, until one gives neutral answers such as 'It's

[78] As TW notes, due to the factivity of being in a position to know that C obtains, the this should really be a biconditional.

[79] Why? If one feels cold then the condition that one feels cold obtains. If that condition is luminous, then at any moment it obtains iff one is in a position to know that it obtains, i.e. they covary simultaneously. This is important for the argument.

> hard to say'; then one begins to dissent, with gradually decreasing hesitations and qualifications; one's final answers are firmly negative.

The argument depends upon two controversial premises.

First, and where α_i is a case α at moment t_i:

(1_i) If in α_i one knows that one feels cold, then in α_{i+1} one feels cold.

NB that the 'i' in (1_i) is a variable ranging over times between dawn and noon. We can replace that variable with actual times, like dawn, which gives us:

(1_{dawn}) If in α_i one knows that one feels cold, then in α_{dawn+1} one feels cold.

This just says that in α[80], if one knows that one feels cold at dawn then one feels cold at a moment shortly after[81] dawn.

[80] or any other case; α is a variable ranging over cases

[81] TW divides up these moments by milliseconds, but that's inessential for the argument: we can take whatever interval makes (1_i) most plausible

TW offers a preliminary defense of (1_i) in this section, then a more detailed defense in §4.4. Here's the preliminary defense:

> Let $t_o, t_1, ..., t_n$ be a series of times at one millisecond intervals from dawn to noon. Let α_i be the case at $t_i (0 \leq i \leq n)$. Consider a time t_i between t_o and t_n, and suppose that at t_i one knows that one feels cold. Thus one is at least reasonably confident that one feels cold, for otherwise one would not know. **Moreover, this confidence must be reliably based, for otherwise one would still not know that one feels cold.**[82,83] Now at t_{i+1} one is almost equally confident that one feels cold, by the description of the case. So if one does not feel cold at t_{i+1}, then one's confidence at t_i that one feels cold is not reliably based, for one's almost equal confidence on a similar basis a millisecond earlier that one felt cold was mistaken. In picturesque terms, that large proportion of one's confidence at t_i that one still has at t_{i+1} is misplaced. Even if one's confidence at t_i was just enough to count as belief, while one's confidence at t_{i+1} falls just short of belief, what constituted that belief at t_i was largely misplaced confidence; the belief fell short of knowledge. One's confidence at t_i was reliably based in the way required for knowledge only if one feels cold at t_{i+1}. In the terminology of cases, we have [premise (1_i)]. (97)

[82] NB! Epistemic internalists don't think that reliability is required for knowledge, so the mental state internalist who is also an epistemic internalist will reject this premise

[83] The reliability constraint that TW is reference here is usually called 'safety': If S were to believe that p, p would not be false. This in turn is often glossed as: if S were to believe that p, then p is true in all nearby (very similar) possible worlds. TW goes into more detail about how he understands the safety requirement on knowing in §5.3

Now here's the second controversial premise:

(2_i) If in α_i one feels cold, then in α_i one knows that one feels cold.[84]

[84] as before, i is a variable ranging over times, so we can sensibly talk about (2_{dawn}) as an instance of (2_i)

Why accept (2_i)? Because it follows from the supposition that the condition of feeling cold is luminous.[85]

[85] TW's ultimate goal is to reject this supposition.

We've supposed that the condition of one feeling cold is luminous. So if that condition obtains in α_i one is in a position to know that it obtains in α_i.[86] Given the stipulation that one 'thoroughly considers how cold or hot one feels' at every $t)i$, it follows that one knows that one feels cold at α_i.

[86] from the definition of 'luminous'

Informal statement of the argument:

We've stipulated that at dawn one feels cold. So by (2_{dawn}), at dawn one knows that one feels cold. So by (1_{dawn}), one feels cold at dawn+1, the moment right after dawn. So by (2_{dawn+1}), at dawn+1 one knows that one feels cold. So by (1_{dawn+1}), one feels cold at dawn+2...

We repeat this argument until we reach noon, thereby proving that if (1_i) and (2_i) are true, and if you start out cold at dawn, then you'll be cold at noon. But we've stipulated that you're warm at noon. So one of the premises must be rejected. TW thinks we should reject (2_i), but since that follows from the supposition that the condition of feeling cold is luminous (together with the stipulation that one is reflecting on one's coldness throughout the morning), that means that we should reject the luminosity of the condition of feeling cold.

TW's formal statement of the argument:[87]

[87] this page of the notes is cut and pasted from TW's p. 98

(1_i) If in α_i one knows that one feels cold, then in α_{i+1} one feels cold.

(2_i) If in α_i one feels cold, then in α_i one knows that one feels cold.

Now suppose:

(3_i) In α_i one feels cold.

By modus ponens, (2_i) and (3_i) yield this:

(4_i) In α_i one knows that one feels cold.

By modus ponens, (1_i) and (4_i) yield this:

(3_{i+1}) In α_{i+1} one feels cold.

The following is certainly true, for α_o is at dawn, when one feels freezing cold:

(3_0) In α_o one feels cold.

By repeating the argument from (3_i) to (3_{i+1}) n times, for ascending values of i from 0 to $n-1$, we reach this from (3_0):

(3_n) In α_n one feels cold.

But (3_n) is certainly false, for α_n is at noon, when one feels hot.[88] Thus the premises (1_0), ..., (1_{n-1}), (2_0), ..., (2_{n-1}), and (3_0)[89] entail a false conclusion. Consequently, not all of (1_0), ..., (1_{n-1}), (2_0), ..., (2_{n-1}), and (3_0) are true. But it has been argued that (1_0),..., (1_{n-1}) and (3_0) are true. Thus not all of (2_0), . . . , (2_{n-1}) are true. By construction of the example,[90] one knows that one feels cold whenever one is in a position to know that one feels cold, so (2_0), . . . , (2_{n-1}) are true if the condition that one feels cold is luminous. Consequently, that condition is not luminous. Feeling cold does not imply being in a position to know that one feels cold.

[88] by stipulation

[89] these are the premises generated by replacing the variable i with names of moments (0,1,2,...; n names noon) in α

[90] Because we're supposing that one's being cold is luminous and also that one is continually reflecting on whether they feel cold

4.4 – Reliability

The point of this section is to further defend premise (1_i).

Recall that TW's main argument is a reductio with only two controversial premises: (1_i), and (2_i), which follows from the supposition that the condition of feeling cold is luminous. Rejecting either premise is sufficient to avoid the absurd conclusion, so it's crucial that TW establish that it's more reasonable to reject (2_i) than to reject (1_i).

In defending (1_i) TW says:

> [If one knows that one feels cold then] one is at least reasonably confident that one feels cold, for otherwise one would not know. Moreover, this confidence must be reliably based, for otherwise one would still not know that one feels cold. Now at t_{i+1} one is almost equally confident that one feels cold, by the description of the case. So if one does not feel cold at t_{i+1}, then one's confidence at t_i that one feels cold is not reliably based, for one's almost equal confidence on a similar basis a millisecond earlier that one felt cold was mistaken. (97)

But what does it mean for one's confidence to be reliably based?

Break this into two questions:

(i) when we talk about confidence that comes in degrees, what exactly are we talking about?

(ii) what's the relevant sense of reliability, and how is it to be measured?

RE: (i)

TW: the notion of 'degree of confidence' that's relevant to knowledge is not the notion of subjective degrees of belief — credences — as measured by betting behavior.[91] The fact that one had a high credence in a p when p turns out to be false is not sufficient to undermine one's reliability (in the relevant sense). Argument:

> Suppose that draws of a ball from a bag have been made. The draws are numbered from 0 to 100. You have not been told the results; your information is just that on each draw i, the bag contained i red balls and $100 - i$ black balls. You reasonably assign a subjective probability of $i/100$ to the proposition that draw i was red (produced a red ball), and bet accordingly. You know that draw 100 was red, since the bag then contained only red balls, even if the proposition that draw 99 was red — to which you assign a subjective probability of 99/100 — is false. That does not justify a charge of unreliability against you. Intuitively, for any i less than 100, **your bets** do not commit you to believing outright that draw i was red.[92] Your outright belief may be just that the probability on your evidence that draw i was red is $i/100$, which is true. On draw 100, unlike the others, you can form the belief on non-probabilistic grounds that it was red. **What incurs the charge of unreliability is believing a false proposition outright, not assigning it a high subjective probability.** (99, emphasis added)

On draw 99 one has a credence of .99 that the ball is red. One also (presumably) has the *outright belief* that the chances of a red ball are 99/100.[93] That outright belief is perfectly consistent with the fact that a black ball is drawn.

OK, so it's the accuracy of *full beliefs* that determines reliability.

[91] These are the degrees of confidence that (many) Bayesians seek to model, and the thesis that the Bayesian formalism constrains rational belief is often rooted in arguments purporting to show that violating those constraints makes one vulnerable to making a series of bets that are guaranteed to lose money (Dutch Books).

[92] Is that right? What if we dramatically increase the numbers – does that create a rational obligation to believe? Even if 'my bets' don't commit me, does my very high credence? This gets as difficult questions of how to get partial-belief models to play nicely with full-belief models.

[93] Note that this is a deductive consequence of the information that the agent possesses

What are full beliefs?

> What is the difference between believing *p* outright and assigning *p* a high subjective probability? Intuitively, one believes *p* outright when one is willing to use *p* as a premise in practical reasoning. Thus one may assign *p* a high subjective probability without believing *p* outright, if the corresponding premise in one's practical reasoning is just that *p* is highly probable on one's evidence, not *p* itself.[94] Outright belief still comes in degrees, for one may be willing to use *p* as a premise in practical reasoning only when the stakes are sufficiently low. Nevertheless, one's degree of outright belief in *p* is not in general to be equated with one's subjective probability for *p*; one's subjective probability can vary while one's degree of outright belief remains zero. Since using *p* as a premise in practical reasoning is relying on *p*, we can think of one's degree of outright belief in *p* as the degree to which one relies on *p*. Outright belief in a false proposition makes for unreliability because it is reliance on a falsehood. **The degrees of confidence mentioned in the argument for (1_i) should therefore be understood as degrees of outright belief.** (99, emphasis added)

[94] I have not idea what TW is going for here – he knows very well that there are very good models for practical reasoning from partial belief (e.g. probabilistic versions of decision theory and game theory)

Upshot: when TW writes of my degree of confidence that I feel cold, he's talking about the degree to which I rely on an outright belief that I feel cold in my practical reasoning (e.g. my reasoning about whether to put on a sweater).

RE: (ii) what's the relevant sense of reliability, and how is it to be measured?

Common objection to the appeal to reliability in epistemology:

Generality Problem: If one believes *p* truly in a case α, in which other cases must one avoid false belief in order to count as reliable enough to know *p* in α?

Pace Conee and Feldman, even if there were no satisfactory answer to the generality problem we can still sensibly appeal to reliability facts. It means that we can't define 'reliable', not that the concept *reliable* is incoherent.[95]

[95] This is really all that TW has to say on the matter, but I'm inclined to agree. As an extremely tangential aside, isn't the also the obviously correct response to Quine's main argument in *Two Dogmas*?

In short, TW is claiming:

> If one believes *p* truly in a case α, one must avoid false belief in other cases sufficiently similar to α in order to count as reliable enough to know *p* in α. (100)

If we individuate the moments in α by milliseconds, the case α_i will be *extremely similar* to case α_{i+1},[96] one pair so by the above passage, if I know that *p* in α_i then it must be true in the extremely similar α_{i+1}.

[96] We could of course imagine a pair of worlds α_i and α_{i+1} that are radically different in spite of the latter being just like the former after one more millisecond has elapsed: perhaps the apocalypse occurs in the interval. That's beside the point: TW isn't claiming that every pair of cases separated only by a millisecond are extremely similar, only that at least one such pair is possible; he's just constructing a counterexample to a universal generalization (the luminosity principle).

Hence premise (1_i) is true.

4.5 – Sorites Arguments

Objection: isn't this a sorites argument?

> If with 0 hairs on one's head one is bald, and, for every natural number i, with i hairs on one's head one is bald only if with $i+1$ hairs on one's head one is bald, then for any natural number n, however large, it follows that with n hairs on one's head one is bald. (102)

TW: the flaw in sorites arguments is that they illicitly exploit the vagueness of a predicate (e.g. 'is bald'). Thus once we replace the vague predicate with a sharpened, non-vague predicate, at least one premise of the sorites argument is false.[97]

[97] it's not entirely clear to me whether TW actually thinks that this is the problem with the sorites argument, or if he's just considering a possible objection to the anti-luminosity argument

Example: replace 'is bald' with 'is bald*' $=_{df}$ 'has fewer than 60,000 hairs on the head'. Then the sorites argument is unsound because " with i hairs on one's head one is bald* only if with $i+1$ hairs on one's head one is bald*" is false when $i = 59,999$: someone with 59,999 hairs is bald* but someone with 59,999+1 hairs is not bald*.

<u>First proposal:</u> the anti-luminosity argument fails once we sharpen the vague predicates

Note that (1_i) and (2_i) together imply:

$(1_i)\&(2_i)$ If in α_i one knows that one feels cold, then in α_{i+1} one knows that one feels cold

And that looks a lot like the premise of the sorties argument that we just rejected.

But every party to the debate thinks that $(1_i)\&(2_i)$ should be rejected, the disagreement is which conjunct we should reject. TW's goal is to defend 1_i; does replacing the vague predicates in $(1_i)\&(2_i)$ mean that we should reject (1_i) rather than (2_i)?

The vague predicates in the anti-luminosity argument are 'knows' and 'feels cold'. Suppose that we sharpen 'is cold', identifying it with some measurable[98] physiological condition of the agent. Suppose that we sharpen 'knows', so that borderline cases now count as not-knowing.[99]

[98] in principle, but not by the cold person

[99] TW is skeptical that perfect sharpening is possible, but that's inessential to the argument

Then TW says:

> Such sharpening has the opposite effect to that predicted by the assimilation of the argument against luminosity to sorites reasoning; (1_i) becomes more not less plausible. (104)

Why? It's not really clear in the text. TW does say:

> The considerations about reliability remain as cogent as before, for they were based on our limited powers of discrimination amongst our own sensations, not on the vagueness of 'feels cold'.

Fair enough, but that doesn't mean that sharpening the predicates makes the case for (1_i) any *more* plausible...

Second attempt[100]: recall that (1_i) says

(1_i) If in α_i one knows that one feels cold, then in α_{i+1} one feels cold.

[100] TW gestures at this ever so briefly when he says "If anything, this [sharpening of vague predicates] strengthens the argument for (1_i), by building more into its antecedent."(104)

General point: this is a universally quantified conditional. Counterexamples to universally quantified conditionals always take the same form: a case in which the antecedent is true and the consequent is false.

Back to the argument.

Sharpening 'knows' to *exclude* borderline cases *strengthens* the antecedent of the conditional[101] in the sense that fewer cases will satisfy it.

[101] strengthens the antecedent relative to a reading on which 'knows' *includes* at least some borderline cases

Since all cases that are counterexamples to the conditional must make the antecedent true, strengthening the antecedent results in fewer cases that can be counterexamples to the conditional.

Similarly, sharpening 'is cold' to *include* borderline cases *weakens* the consequent, in the sense that more cases will satisfy it.

Weakening the antecedent results in fewer cases that can be counterexamples.

So, sharpening the vague predicates 'knows' and 'is cold' in this way only makes it harder to find a counterexample to (1_i), so in that sense at least it makes (1_i) more plausible.

Objection: this line of reasoning depended entirely on *how* we sharpened the predicates. If we sharpen in the other direction we weaken the antecedent and strengthen the consequent, making (1_i) less, not more, plausible.

Second Proposal: the vagueness of 'knows' and 'is cold' is essential to the truth of the luminosity principle, and 'any sharpening that falsifies (2_i) [violates] the intended meanings of the vague terms' because the vague meanings of those predicates makes (2_i) analytic. (105)

So where does the anti-luminosity argument go wrong? For some value of *i* and with unsharpened predicates, (1_i) will be less than perfectly true.

What does 'perfectly true' mean? In degree-theoretic semantics, truth comes in degrees. Sentences involving vague predicates are true to the degree that the subject instantiates the vague predicate. For example, 'the apple is red' is more true than 'the orange is red'. A sentence is perfectly true iff it is true to the maximal degree.

We can also define truth functional connectives using the notion of degrees of truth. A perfectly true conditional is one whose consequent is at least as true as its antecedent.

So, when the defender of luminosity asserts that (1_i) is less than perfectly true, what they are saying is that for some α and some i, 'in α_{i+1} one feels cold' (= the consequent of (1_i)) is at least a little bit less true than 'in α_i one knows that one feels cold' (= the antecedent of (1_i)).[102]

[102] TW doesn't provide any reason to believe that this is the case on behalf of his interlocutor, he's just considering whether this approach would provide the basis of a response if it could be defended

Problem: proponent of this argument must reject:

$(1P_i)$ If it is perfectly true that in α_i one knows that one feels cold, then it is perfectly true that in α_{i+1} one feels cold.

Why? Suppose that (2_i) is perfectly true:

$(2P_i)$ If it is perfectly true that in α_i one feels cold, then it is perfectly true that in α_i one knows that one feels cold.

But this sets us up for an analogue to the anti-luminosity argument, one concluding that one is perfectly cold at noon (which is impossible).

So the defender of luminosity who pushes this argument must reject $(1P_i)$. Presumably they will claim that:

> ...for some number i, it can be perfectly true that in α_i one knows that one feels cold, but slightly less than perfectly true that in α_{i+1} one feels cold. [i.e. $(1P_i)$ is false.] Can there be such an i? If it is less than perfectly true that in α_{i+1} one feels cold, then there is a strict standard by which it is false in α_{i+1} that one feels cold; so, by that standard, in α_{i+1} one is fairly confident of what is false, that one feels cold. If so, it is less than perfectly true that in α_i one knows that one feels cold, **if the reliability considerations are to be assigned any positive weight at all.** (105, emphasis added)

In other words, TW's opponent must claim that there are two cases separated only by a millisecond such that one feels cold in α_i and knows that they feel cold, but does not feel cold in α_{i+1}. Since those cases are incredibly similar, TW thinks that the fact that you believe in α_i when that belief is false in α_{i+1} indicates a glaring lack of reliability. Hence if you really think that this pair of cases is possible, then you should think that reliability is irrelevant to knowing.

Remember, however, that epistemic internalists do think that reliability is irrelevant to knowing – that condition is neither a mental state, nor a condition that's accessible to the agent!

4.6 – Generalizations

If the anti-luminosity argument from §4.3 is sound, then the condition that one feels cold is not luminous. Clearly analogous arguments can be formulated to show of other conditions that they are not luminous. What are the limits of that argument schema? Does the luminosity of any condition survive?

NB: this section is specifically concerned with the applicability of the anti-luminosity argument, not with whether any conditions are actually luminous. As TW points out, even conditions immune to the argument from §4.3 can be shown to be non-luminous using other arguments.

The argument generalizes to each of the putative paradigm cases of luminous conditions: that one is in pain, that two words have the same meaning, that one is appeared to thusly.

Objection: what about response-dependent conditions[103]? Being appeared to thusly is a paradigm response dependent condition it is claimed)...

[103] that is, conditions whose obtaining in some way depend constitutively upon the one's disposition to judge that they obtain

There are (at least) two ways to understand response-dependence:

Weak response-dependence: whether response-dependent condition C obtains has *some* constitutive dependence on whether one is disposed to judge that it obtains

Strong response-dependence: whether response-dependent condition C obtains depends entirely on whether one is disposed to judge that it obtains

Weakly response-dependent conditions needn't be luminous, so this is irrelevant

Strongly response-dependent conditions are luminous. But the argument in §4.3 shows that even the paradigm cases of putatively response-dependent propoerties – being in pain, being appeared to thusly – are not luminous, so they aren't strongly response-dependent.

But, the argument does depend on some specific features of the condition that one feels cold, so the argument only generalizes to conditions that share those specific features:

(1) The condition that one feels cold obtains in some cases but not others

The case is set up so that one feels cold at dawn and does not feel cold at noon. The argument is a reductio, where the putative absurdity is that that one feels cold at noon (a conclusion forced by the supposed luminosity of the condition of feeling cold) and that one does not feel cold at noon (a stipulation of the case). If feeling cold always obtains, or never obtains, then this absurdity cannot be derived. Such conditions are not vulnerable to TW's anti-luminosity argument.

NB: Conditions that never obtain are vacuously luminous

- The definition is of 'luminous' takes the form of an indicative conditional whose antecedent is false anytime the condition does not obtain

NB: Conditions that always obtain might be luminous, might not, depending on their guise of presentation[104]

(2) The condition that one feels cold is not only contingent: it's non-eternal

Eternal conditions are such that: if they obtain, they always obtain. So eternal conditions aren't vulnerable to the anti-luminosity argument

Example: the condition of feeling cold on New Year's Eve 1999.

NB: though eternal conditions are not vulnerable to the anti-luminosity argument, many can be shown to be non-luminous in other ways. E.g., even if C is eternal, it's possible to change from being in a position to know that C obtains to not being in a position to know: I was once in a position to know if that condition obtained in my case, but now I've forgotten, so I am not now in a position to know. So it obtains now but I'm not now in a position to know that it obtains. So it's not luminous.

(3) Assumes that one is considering the relevant condition under the relevant guise throughout the process

Here's TW's argument:

> [L]et C be the condition that one is entertaining the proposition that it is raining, and let G be the guise under which C has just been pre-

[104] This is a strange thing for TW to concede. Luminosity is a property of conditions, and conditions are individuated coarsely: if two conditions obtain in exactly the same set of cases then they are identical. Guises are (presumably) individuated more finely: being 2+2 years old is a different guise from being 2^2 years old. How does the guise under which a condition is presented affect whether the condition so-presented is luminous? If I'm considering a necessary condition under a guise s.t. I'm not in a position to know that that condition obtains, then it's not luminous. If the mere possibility of thinking about that condition under another guise that would put me in a position to know the condition is itself sufficient to put me in a position to know that the condition obtains, then the condition is luminous. So again, how are the guises relevant?

> sented here. To consider C under G is to consider as such the condition that one is entertaining the proposition that it is raining; in so doing, one thereby entertains the proposition that it is raining, so C obtains. Thus one cannot gradually pass from cases in which C obtains to cases in which C does not obtain while considering C under G throughout the process. Although one can gradually pass from cases in which C obtains to cases in which C does not obtain, one does not consider C under G in the late stages of the process. For all the argument shows, C is luminous: if one is entertaining the proposition that it is raining, then one is in a position to know that one is entertaining the proposition that it is raining. (108-9)

TW seems to be selling his own argument short here.

His anti-luminosity argument doesn't require the possibility that one can 'gradually pass from cases in which C obtains to cases in which C does not obtain while considering C under G throughout the process' because it doesn't actually require one to be considering C at the moment at which C fails to obtain. What he needs is for the supposition that C is luminous to produce a the following: at t_i C obtains and one knows it, and at t_{i+1} C does not obtain.[105] So clearly one must be considering whether C obtains at t_i — otherwise one couldn't know that C obtains — but there's no reason to require that one must consider whether C obtains at t_{i+1}. So the mere fact that considering C entails C (assuming that it is a fact) is irrelevant.

[105] recall that this is what the reliability of knowing is supposed to make impossible

TW's conclusion of this section:

> Luminous conditions are curiosities. Far from forming a cognitive home, they are remote from our ordinary interests. The conditions with which we engage in our everyday life are, from the start, non-luminous. (109)

4.7 – Scientific Tests

The point of this section is to consider the implications of a test for whether one feels cold (e.g. an FMRI scan).

If we could identify some measurable physiological condition V that is correlated with feeling cold, then we might be able to identify a moment at which one is reflecting on whether they feel cold, fails to know that they do[106], and yet they feel cold. In other words, we'd have empirical evidence of a counterexample to the luminosity of the condition of feeling cold.

[106] though this part too is complicated by the failure of the K¬K principle

Problem: how would you determine the physiological variable V is correlated with feeling cold if you don't have some independent means of determining in when you feel cold?

The prospects for this type of test are bleak

4.8 – Assertibility Conditions

Realist semantics: the meaning of a sentence is its truth conditions, i.e. the set of possible worlds in which that sentence is true. Realists claim that there could be meaningful sentences (sentences with truth conditions) even if no one is in a position to know what those truth conditions consist in.

Dummett's[107] objection: meaning and use are intimately connected, so when someone uses (says) sentence S correctly they must know the the truth conditions of S. It's not enough to be disposed to say the biconditional: 'S is true iff *p*', the speaker must actually *know* that the sentence expressing the biconditional true. But to do that you must know what the sentence 'S* is true iff (S is true iff *p*)' means, which means you must know S**... Regress looms. Dummett's alternative proposal: using language doesn't require knowledge truth conditions, but in knowledge of assertibility conditions: the conditions in which speakers are warranted in using sentences assertively. So given the close connection between meaning and use, meaning is determined by assertibility conditions, not truth conditions.

[107] Random aside: Michael Dummett was the Wykeham Professor of Logic at Oxford, a chair that TW now holds

According to Dummett, knowing the assertibility conditions of S requires: when sentence S's assertibility condition obtains, then we are in a position to know that it obtains. In other words, it requires that the condition that S's assertibility condition obtains is luminous. But by an analogue to the argument in §4.3, it isn't luminous.

Example: it gradually transitions from raining to not-raining. At first 'it's raining' is assertible, and eventually it isn't.

8 - Scepticism

8.1 – Plan

Point of this chapter: examine the significance of the anti-luminosity for skeptical arguments.

Why this is interesting: rational thinkers respect their evidence. The anti-luminosity argument shows that we aren't always in a position to know what our evidence is. How can we respect our evidence if we don't know what our evidence is?

8.2 – Scepticism and the non-Symmetry of Epistemic Accessibility

TW's interlocutors are skeptics who contrast good cases and bad cases.

Good Case: things appear generally as they ordinarily do, and are that way; one believes some proposition p (for example, that one has hands), and p is true; by ordinary standards, one knows p.

Bad Case: things still appear generally as they ordinarily do, but are some other way; one still believes p, but p is false; by any standards, one fails to know p, for only true propositions are known.

The sceptic argues that because one believes p falsely in the bad case, one does not know p (even though p is true) in the good case.

> For the sceptic, the two cases are symmetrical: just as it is consistent with everything one knows in the bad case that one is in the good case, so it is consistent with everything one knows in the good case that one is in the bad case. One simply cannot tell which case one is in. For the sceptic's opponent, the two cases are not symmetrical: although it is consistent with everything one knows in the bad case that one is in the good case, it is not consistent with everything one knows in the good case that one is in the bad case. (165-6)

Example: In the good case, I have hands and I know that I have hands. In the bad case, I'm a handless BIV, but I (falsely) believe that I have hands.

The things that I know in the bad case — nothing, in this example — are consistent with my being in the good case. But the things that I know in the good case — that I have hands — are not consistent with my being in the bad case.

8.3 – Difference of Evidence in Good and Bad Cases

Skeptic: In the bad case my evidence is 'insufficient for the truth' of *p*, and my evidence in the good case is the same as in the bad case, so in the good case my evidence my evidence is also 'insufficient for the truth' of *p*.[108]

But, if the evidence is not the same in both cases then 'false belief in the bad case would be a far less pressing threat to knowledge in the good case'.

The skeptic can't just assume that the evidence is the same in both cases: (some) externalists think that it's impossible to have the same evidence when the external world is so different.[109] So, the skeptic must describe the good and bad cases in neutral terms and then argue that the evidence is the same. But how?

Here's a first attempt an an argument:[110]

Suppose for reductio that (i) the evidence is different in the good and bad cases, and (ii) we were always in a position to know what our evidence is. In that case, if one were in the bad case then one could reason as follows: I notice that I have 'bad case evidence'; if I were in the good case I would have 'good case evidence'; so I must be in the bad case.

In this way one could come to know that one is in the bad case and not in the good case.

But, the anti-skeptic concedes that it is consistent with everything that one knows in the bad case that one is in the good case; contradiction! So, reject the supposition that the evidence is different in the good case and the bad case.

Problem: We started the reductio with two suppositions: (i) the evidence is different in the good and bad cases, and (ii) we were always in a position to know what our evidence is. Rejecting either supposition would resolve the absurdity. Why reject (i) instead of (ii)?

[108] This is a strange way to put the issue. Presumably knowing that *p* does not require that my evidence actually entail *p*. Presumably, fallibilism about knowledge is true, TW seems to be attributing *infallibilism* to the skeptic.

[109] Example: you assert a safety condition on evidence, and you claim that in cases nearby to the bad case you still don't have hands and it doesn't appear to you that you have hands. In that case it's just false that: if it had been the case that I didn't have hands then I wouldn't have believed it.

[110] TW spells out the argument in much greater detail in §8.4

8.4 – An Argument for Sameness of Evidence

In this section TW recapitulates the argument from §8.3, but much more carefully.

Background assumptions:

- One is rational, possesses the relevant concepts, and is currently reflecting on one's evidence s.t. if one is in a position to know that p then one knows that p
- One knows what one's evidence is, in the sense that if one's evidence has property π then one knows that one's evidence has property π
- Values of π are restricted to 'appropriate' properties, though TW doesn't really say what counts as appropriate. In any event, if π is appropriate, then so is not-π.

The Argument:[111]

(1) For any appropriate property π, in any case in which one's evidence has π, one knows that one's evidence has π.[112]

(2) For any appropriate property π, if in the good case one's evidence lacks π, then in the bad case one knows that in the good case one's evidence lacks π.[113]

(3) It is consistent with what one knows in the bad case that one is in the good case.[114]

(4) In the bad case one's evidence has π.[115]

(5) Suppose for reductio: In the good case one's evidence lacks π.

(6) In the bad case one knows that in the good case one's evidence lacks π. (from (2) and (5))

(7) In the bad case one knows that one's evidence has π. (from (1) and (4))

From 'In the good case one's evidence lacks π' and 'One's evidence has π' one can deduce 'One is not in the good case'. By (6) and (7), in the bad case one knows each premise of that deduction; hence:

(8) It is inconsistent with what one knows in the bad case that one is in the good case.

Now (8), which rests on assumptions (1), (2), (4), and (5), contradicts (3). Thus on assumptions (1)-(4) we can deny (5) by reductio ad absurdum:

(9) In the good case one's evidence has π.

We can conditionalize (9) on assumption (4):[116]

(10) If in the bad case one's evidence has π, then in the good case one's evidence has π.

[111] What follows is mostly taken verbatim from p. 171-2, though I've cleaned it up a bit for clarity and length

[112] i.e. one always knows what one's evidence is

[113] follows from the assumption that one can know what one's evidence would be in other cases

[114] from the description of the cases

[115] from the description of the cases

[116] everyone clear why?

Here (10) rests on assumptions (1)-(3).[117] Since the appropriate properties were assumed to be closed under complementation, we can run through the argument (1)-(10) with 'not-π' in place of 'π', yielding:

(11) If in the bad case one's evidence has not-π, then in the good case one's evidence has not-π.

Contraposition on (11) yields the converse of (10).[118] Therefore, generalizing on 'π' in (10) and (11), we have:

(12) One's evidence in the good case has the same appropriate properties as one's evidence in the bad case.

Assuming that evidence is individuated by it's appropriate properties, this is tantamount to the claim that one's evidence is the same in the good and bad cases.

8.5 – The Phenomenal Conception of Evidence

The demand that the evidence be the same in the good and bad cases rules out certain accounts of evidence.[119]

- evidence can't consist in all the true propositions
  - the truth values of lots of propositions differ between good and bad cases, so according to this account the evidence differs
- evidence can't consist in perceptual states with externally individuated contents
  - if the content of perceptual states is determined in part by facts about the world external to the agent, and if those facts differ between good and bad cases, then the perceptual states of the agent will differ between the two cases, and hence so will the evidence.
- evidence can't consist in retinal stimulations or brain states
  - 'in some sceptical scenarios they [the retinal stimulations or brain states] are unknowably different too'.

The basic point here: the way the skeptical scenarios are supposed to work is that you have the same evidence in both the good and the bad case, but what you believe is true in the good case and false in the bad case.[120] The lesson is supposed to be that there's something defective with your evidence, or with the putative knowledge that we obtain on the basis of that knowledge.

[117] everyone clear why?

[118] i.e.: if in the good case one's evidence has π, then in the bad case one's evidence has π.

[119] OR it rules out certain pairs of cases as candidates for skeptical scenario. The dialectic is this: TW thinks the skeptic needs to pick a pair of cases in which the evidence is the same, but you know in one case and not in the other. Once the cases are fixed, the sameness-of-evidence thesis requires that evidence be understood as something that's the same between the *those two cases*. Alternative methodology that the skeptic might employ: start with a particular account of evidence, then find a pair of cases on which the evidence is the same *according to that account of evidence*, but where the agent knows in the good case and doesn't know in the bad case.

[120] Better: you *know* in the good case but you don't know in the bad case.

But this picture requires that the evidence that one has in a case be substantially independent of other facts about the world. When evidence depends on the world, then we can't vary facts about the world between good and bad cases (as the skeptic requires) without thereby varying facts about one's evidence (which the skeptic prohibits)

The skeptic's need for evidence to vary independently of facts about the world pushes them towards a **phenomenal conception of evidence**, on which evidence includes present perceptual experience, present memory experiences, etc.[121]

[121] It's not a coincidence that Descartes assumes a phenomenal conception of evidence in the first meditation

8.6 – Sameness of Evidence and the Sorites

TW: we should reject (12), that one's evidence is the same in the good and bad cases.

The argument from (1) to (12) is valid, so we need to reject a premise.

TW offers an argument against (1), that one is always in a position to know what one's evidence is:[122]

[122] the following is essentially copied from p. 175-5, with some changes for clarity and length

The set-up is the same as in the anti-luminosity argument: imagine a series of times $t_i, ..., t_n$ one millisecond apart and α_i, ..., α_n the relevant series of cases. One's experiences change gradually over that time as one watches a sunrise.

(2_i) For any appropriate property π, if in α_{i-1} one's evidence lacks π, then in α_i one knows that in α_{i-1} one's evidence lacks π.[123]

[123] TW stipluates that 'α_i' and 'α_{i-1}' are abbreviations of descriptions of cases, and that these descriptions include all of the properties that one's evidence would have if one were in that case. hence no matter what case you're in, you are in a position to know what properties your evidence would have in α_{i-1} (or any other case).

Now consider the description of what is in fact the case one was in a millisecond ago. Given one's limited powers of discrimination, one does not know propositions from which one can deduce that that description does not apply to one's own case:

(3_i) It is consistent with what one knows in α_i that one is in α_{i-1}.

...(3_i) is obvious in roughly the way in which it is obvious that it is consistent with what I know by sight when I am in fact looking at a distant tree i millimetres high that I am looking at a tree only $i - 1$ millimetres high. From premises which I know on the basis of sight to the conclusion that I am not looking at a tree only $i - 1$ millimetres high, there is no hope of constructing a valid deduction, not even one which I am somehow not in a position to carry out. Similarly, from premises which I know in α_i to the conclusion that I am not in

α_{i-1}, there is no hope of constructing a valid deduction, not even one which I am somehow not in a position to carry out.

The argument proceeds as before. Restrict 'π' to appropriate properties and assume:

(4_i) In α_i one's evidence has π.

(5_i) Suppose for reductio: in α_{i-1} one's evidence lacks π.

Premises (2_i) and (5_i) entail:

(6_i) In α_i one knows that in α_{i-1} one's evidence lacks π.

Premises (1) and (4_i) entail:

(7_i) In α_i one knows that one's evidence has π.

From 'In α_{i-1} one's evidence lacks π' and 'One's evidence has π' one can deduce 'One is not in α_{i-1}' By (6_i) and (7_i), in α_i one knows each premise of that deduction; hence:

(8_i) It is inconsistent with what one knows in α_i that one is in α_{i-1}.

Now (8_i), which rests on assumptions (1), (2_i), (4_i), and (5_i), contradicts (3_i). Thus on assumptions (1) and (2_i)-(4_i) we can deny (5_i) by reductio ad absurdum:

(9_i) In α_{i-1} one's evidence has π.

We can conditionalize (9_i) on assumption (4_i):

(10_i) If in α_i one's evidence has π, then in α_{i-1} one's evidence has π.

Here (10_i) rests on assumptions (1), (2_i), and (3_i). Since the appropriate properties were assumed to be closed under complementation, we can run through the argument (1)-(10_i) with 'not-π' in place of 'π', yielding:

(11_i) If in α_i one's evidence has not-π, then in α_{i-1} one's evidence has not-π.

Contraposition on (11_i) yields the converse of (12_i). Thus, generalizing on 'π' in (10_i) and (11_i), we have:

(12_i) One's evidence in α_{i-1} has the same appropriate properties as one's evidence in α_i.

Proposition (12_i) rests on assumptions (1), (2_i), and (3_i). But the relation between the cases in (12_i) is transitive; if one's evidence in case β has the same appropriate properties as one's evidence in case γ and one's evidence in case γ has the same appropriate properties as one's evidence in case δ, then one's evidence in β has the same appropriate properties as one's evidence in δ, for what is in question is exact sameness in all properties from a fixed class. Although (3_i) claims only that α_{i-1} and α_i are indiscriminable, and indiscriminability is a non-transitive relation, we have deduced from it and the other premises the transitive relation of exact sameness of evidence in the appropriate respects. Thus (12_i), ..., (12_n) together yield:

(13) One's evidence in α_0 has the same appropriate properties as one's evidence in α_n.

The conclusion (13) rests on assumptions (1), (2_1),..., (2_n), (3_1),..., (3_n). But (13) is obviously false. One's evidence at the end of the process is grossly different from one's evidence at the beginning; it differs in many of its appropriate properties. Since (2_1),..., (2_n), (3_1),..., (3_n) are true, for reasons already given, (1) is false.

Sound argument?

BTM: TW has told us that the skeptic should be inclined to the phenomenal conception of evidence, so to be charitable to the skeptic let's assume that that's right.

TW wants to show that if we are always in a position to know what evidence we have, then we must have the same evidence looking toward the eastern horizon before dawn as we do at noon. That's an absurd result, so it licenses him to reject the 'always-in-a-position-to-know-your-evidence' thesis.

The basic move is to show that it implies that we have the same evidence in α_i and in α_{i-1}, and then to generalize that result for α_{i-1} and in α_{i-2}, etc. In other words, he needs to argue that α_i and in α_{i-1} aren't special. (this should sound familiar)

But why think that?

Suppose we take inspiration from Steup and claim the following model of phenomenal experience: experience isn't continuous or fine grained, it's coarsely chunked. That means that rather than my phenomenology changing a tiny bit every millisecond as the sun raises, it doesn't change at all for a few minutes and then updates all at once. In other words, changes in the actual position of the sun relative to the horizon are only approximately tracked by changes in my phenomenology.

If that's right, then I may well have the same phenomenology α_i and in α_{i-1}, *as long as those moments are both contained within a single* *phenomenological chunk*, but my phenomenology will be different if they fall on either side of a chunk-border. In other words, TW might be right about α_i and in α_{i-1} in particular, but wrong that that lesson generalizes to all pairs of cases.

TW needs it to be the case that the phenomenology is fine grained, but our ability to discriminate it is not. Given the phenomenal conception of evidence, this is tantamount to the claim that our evidence is fine grained but our ability to discriminate our evidence[124] is not, i.e. that we aren't always in a position to know what our evidence is.

[124] Better: our ability to discriminate what properties our evidence has

But on the Steupian alternative, the phenomenology is coarse grained and our discriminatory abilities are similarly coarse grained. Given the phenomenal conception of evidence, this is tantamount to the claim that both our evidence and our ability to discriminate our evidence is fine grained. And in that case there's no problem with the claim that we always know what our evidence is.

So where did the argument go wrong? It's valid, so where's the false premise?

Reconsider (3_i), specifically understanding 'α_i' and 'α_{i-1}' as *variables* ranging over cases[125] (but with the further understanding that they're roughly the same case separated by one millisecond (speaking loosely here)):

[125] rather than *names* of cases. Presumably that's what TW had in mind all along; I mention it only for emphasis

(3_i) It is consistent with what one knows in α_i that one is in α_{i-1}.

If 'α_i' and 'α_{i-1}' are variables, then (3_i) should hold of any possible values of those variables taken from the case described, i.e. it should hold for any pair of cases separated only by one millisecond in the sunrise-scenario

Here's TW's defense of (3_i):

> ...(3_i) is obvious in roughly the way in which it is obvious that it is consistent with what I know by sight when I am in fact looking at a distant tree i millimetres high that I am looking at a tree only $i-1$ millimetres high. From premises which I know on the basis of sight to the conclusion that I am not looking at a tree only $i-1$ millimetres high, there is no hope of constructing a valid deduction, not even one which I am somehow not in a position to carry out. Similarly, from premises which I know in α_i to the conclusion that I am not in α_{i-1}, there is no hope of constructing a valid deduction, not even one which I am somehow not in a position to carry out. (176-7)

Is it in fact 'obvious that it is consistent with what I know by sight when I am in fact looking at a distant tree i millimetres high that I am looking at a tree only $i-1$ millimetres high'? Given our Steupish theory of visual phenomenology, that depends: is the difference in tree-height a difference that falls within a phenomenological chunk, or do the differing cases straddle two chunks?

If they straddle two chunks, and if the phenomenological conception of evidence is true, then we have difference evidence in the two cases. Since I know what evidence I would have in α_i and in α_{i-1}, **if in fact I always know what my evidence is** then (3_i) is just false, since the evidence I would have in α_i is different than and inconsistent with the evidence I would have in α_{i-1}.

So in order TW's defense of (3_i) to be plausible we need to assume that I'm not always in a position to know what my evidence is. But that's the conclusion of the argument! So either (3_i) is false, or the argument for (13) is question-begging.

[end BTM]

8.7 – The Non-Transparency of Rationality

Problem:

> If rationality requires one to respect one's evidence, then it is irrational not to respect one's evidence. But how can failing to respect one's evidence be irrational when one is not in a position to know that one is failing to respect one's evidence? More generally, how can ϕ-ing be irrational when one is not in a position to know that one is ϕ-ing? (179)

Big[126] picture question: why be rational?

[126] really big

TW: Here's the standard picture of rationality: the ultimate goal is to believe truly. But that's hard: we don't have direct access to the truth.

Rationality is a method for accessing truth – although we don't have direct access to truth, we do have direct access to whether we're being rational. If that's right then although it's impossible to follow the method 'believe truly!', we can follow the method 'be rational!' because we're always in a position to know what rationality requires of us. But...

> If the argument of section 8.6 is correct, this picture of rationality is mistaken. Just as one cannot always know what one's evidence is, so one cannot always know what rationality requires of one. Just like evidence, the requirements of rationality can differ between indiscriminable situations. Rationality may be a matter of doing the best one can with what one has, but one cannot always know what one has, or whether one has done the best one can with it. If something is a method only if one is always in a position to know whether one

> is complying with it, then there are no methods for learning from experience. (179)

So what's TW's alternative to the standard picture?

> We can use something as a method in contexts in which one is usually in a position to know whether one is complying with it, even if in other contexts one is not usually in a position to know whether one is complying with it. **In that sense, we can use even believing truly as a method in contexts in which one is usually in a position to know what is true: for example, when forming beliefs in normal conditions about the spatial arrangement of medium-sized objects in one's immediate environment.**[127] In more difficult contexts, believing truly becomes an aim and we fall back on the method of believing rationally. Rationality becomes a sub-goal on the way to truth. That does not require one always to be in a position to know what rationality requires of one; it requires merely that one often knows what rationality requires when one does not know what truth requires. Nothing has been said here to undermine that requirement. In still more problematic contexts, paradoxes throw our very standards of rationality into doubt, and we fall back still further on what workable methods we can find. Cognition is irremediably opportunistic. (179-80, emphasis added)

[127] this is foreshadowing TW's account of evidence (that E=K) on the next chapter. Is it really plausible that we are sometimes in a position to believe truly, without the mediation of evidence?

Much of what follows in this section is foreshadowing of chapters 9 and 10, in which TW lays out his positive theory of evidence.

But here are some important points that TW thinks he's established:

- TW's argument against the accessibility of evidence and of one's rational status does not rely on any particular account of evidence.
- The skeptic needs it to be the case that one's evidence is the same in the good case and in the bad case
- Skeptics can't just stipulate that the evidence is the same, as many externalists will claim that, given what we're told about the external worlds in the good case and in the bad case, it's impossible to have the same evidence in each
- The skeptic's best argument for that the evidence is the same[128] relies on the premise that we always have access to our evidence
- TW has argued that that's false
- Importantly, TW thinks that his argument doesn't rely on any particular account of evidence

[128] that argument is stated informally in §8.3 and formally in §8.4

8.2 – Scepticism without Sameness of Evidence

The skeptical arguments heretofore considered require the metaphysical possibility of a bad case in which one believes that *p* and yet *p* is false, and one believes that *p* on the same evidence as in the good case (where one believes that *p* and *p* is true). If one's evidence is different in the two cases, then one would be in a position to reason 'my evidence in the good case would be X, but my actual evidence is X′, so I'm not in the good case'. That is, one could reason that way if one was always in a position to know what one's evidence is.

In these cases, the source of skeptical doubt is the *evidence*: the point is to use the bad case to show that the evidence is consistent with false belief, which creates doubt about the evidence (that *same* evidence) even in the good case.

TW has argued that evidence in the good and bad cases is not the same,[129] so the skeptical arguments fail.

[129] That's too strong: he's argued against one reason to believe that the evidence *is* the same, but he hasn't gone so far as to argue that it's not the same.

Evidence worries aren't the only source of skeptical argument. Some skeptics worry about the *methods* by which we form our beliefs.

Importantly, the source of doubt might be with the method even in cases in which one has *different* evidence in the good and bad cases:

> **Example 1:** one forms beliefs as follows: if the coin lands heads, believe *p*, and if it lands tails believe not-*p*. In this way one comes to believe in the good case that *p*. *p* happens to be a necessary truth, so it's true in every case. But in the bad case the coin lands tails, so one believes not-*p*.
>
> One's evidence in the bad case includes that the coin landed tails, and in the good case it includes that the coin landed heads, so the evidence is different in the two cases.
>
> Skeptic: in the good case one doesn't know that *p* because one's method for forming beliefs — flipping a coin — is unreliable.[130]

[130] It's unreliability is shown by the fact that it leads to false belief in the bad case.

Even when *p* is not a necessary truth, we can get skeptical worries from cases where the evidence is not the same:

> **Example 2:** *q* is a contingent truth. One forms beliefs on the basis of how things appear, but then notices that this method is unreliable when one is dreaming. In the good case, one's evidence is coherent and on that basis one believes that *q* and *q* is true. In the bad case one is dreaming and one's evidence is incoherent, and one believes that *q* but *q* is false.
>
> Now suppose that in order to form rational beliefs on the basis of how things appear, one must know that one has coherent evidence.

> Skeptic: given one's method for forming beliefs, one checks whether one's evidence is coherent by checking whether it appears that one's evidence is coherent. But that method is unreliable, since even when one is dreaming and one's evidence *is in fact* incoherent, it still *appears to be* coherent. So even in the good case one doesn't know that *q* on the basis of appearances.

Note that in both cases, the source of skeptical doubt is the *method* of forming beliefs, not the *evidence* on the basis of which those beliefs are formed.

TW: skeptics who are worried about methods are making two assumptions:

Assumption 1: knowing requires that one forms beliefs by following a suitable (reliable?) rule or method

Assumption 2: one is always in a position to know what rule or method one is following[131]

[131] 'For if one was not in a position to know what rules one was using, and one's rationality depended on the rationality of one's rules, how could one be required to be rational?' (183)

Problem: On what theory of rules/ methods are these assumptions plausible? In particular, how are we to individuate two rules R and R* so that, whenever one is following R, one is always in a position to know that one is following R (and not R*)?

TW: The skeptic needs rules to beindividuated *phenomenally*. Otherwise, '...one could to all appearances be using it while not in fact be doing so; in which case one would not be in a position to know that one was not using that rule.' (183)

But that sets up a familiar argument:

Argument (1) - (12) showed that, on the assumption that we're always in a position to know what our *evidence* is, we must have the same evidence in the good case and in the bad case. Since the belief formed on that evidence failed to amount to knowledge in the bad case, the belief formed on that evidence fails to amount to knowledge even in the good case.

By parallel argument, we can show that, on the assumption that we're always in a position to know what *rule* we're following, we must be following the same rule in the good case and in the bad case. Since the belief formed according to that method failed to amount to knowledge in the bad case, the belief formed on according to that rule fails to amount to knowledge even in the good case.

TW: For familiar reasons, this fails:

> The sceptic's conception of a rule collapses. By an argument parallel to (1) - (13) (via ($2i_i$) - (12_i)), only trivial rules meet the epistemic re-

> quirement. For a series of indiscriminable differences links a case in which one uses a given rule to a case in which one uses a quite different rule. For example, one initially believes p for reason R while giving no weight to reason gradually one gives less weight to R and more to until finally one believes p for reason R* while giving no weight to R. Rand R* differ so much in kind that believing for reason R and believing for reason R* amount to using different rules. An argument just like that of section 8.6 refutes the assumption that in every case one is in a position to know what rule one is using. (183)

Really interesting parting thought from TW:

By assuming that self-knowledge is so easy to come by we make it harder to obtain knowledge of the external world. Skeptical arguments are generated by claiming that (i) we can't tell the difference between our evidence/ method in the good and bad cases, and assuming that (ii) this implies that our evidence/ method *really is the same*. The assumption in (ii) is tantamount to a claim about our capacity for self-knowledge. Once it's rejected, skeptical scenarios cannot be constructed.

9 - Evidence

9.1 – Knowledge as Justifying Belief

Recent epistemology has deemphasized knowledge in favor of justified belief: the role of the epistemologist is to spell out the conditions under which one is justified in believing that *p*, etc.

TW has argued that this is a mistake, as knowing is more explanatorily powerful than justified, true belief (see Ch. 3)

Suppose TW's arguments about the independent value of knowledge are wrong, and that knowledge is valuable (if at all) solely from its relationship to the justification of belief. Then what's the nature of that relationship?

Historically: knowledge is taken as the more mysterious, less fundamental relata, is analyzed in terms of justification (plus truth and belief).

Slogan: Knowlege is the thing that gets justified.

TW has rejected this JTB analysis of knowledge, so that can't be right

Other possibility: Justification is the more mysterious relata, so we analyze justification in terms of knowledge

Slogan: knowledge is the thing that does the justifying

TW wants to explore/ advocate for the 'other possibility': knowledge is evidence (E=K), and evidence is what justifies belief, so knowledge is what justifies belief.

Clarifications about E=K:

- it's not a conceptual claim;[132] it's not an assertion of identity between the concept of evidence and the concept of knowledge.
- The claim might not be a priori
- the claim is just that E and K are coextensive in every metaphysically possible case
- speculation: the regress problem has been historically hard to resolve because foundationalism is true, but epistemologists were looking to the wrong foundations: all of one's knowledge

[132] TW is a bit ambiguous on this point: 'The proposed account uses the concept of knowledge in partial elucidation of the concepts of evidence and justification.'

9.2 – *Bodies of Evidence*

When is e evidence for the hypothesis h, for a subject S? Two conditions:

1. e should speak in favor of h
   (a) e should raise the probability of h. TW takes this to mean that the probability of h conditional on e should be higher than the unconditional probability of h; in symbols, $P(h \mid e) > P(h)$[133,134]
2. e should have some kind of creditable standing.

   What kind of probability is P?

   It's not a priori

- whether e raises the probability of h depends on background information[135]
- e shouldn't be 'built into the background information' either, though, since in that case $P(e) = 1$, so $\frac{P(h|e)}{P(h)} = 1$ so it's not the case that $P(h \mid e) > P(h)$, so by (1) above it's not the case that e is evidence for h

  What's the function of (2)?

> e may raise the probability of h in the sense that $P(h \mid e) > P(h)$ even if S knows that e is false or has no idea whether e is true; but then, for S, e would not be evidence for h. That is why we need the second condition, that e should have a creditable standing. A natural idea is that S has a body of evidence, for use in the assessment of hypotheses; that evidence should include e. The probability distribution P is informed by some but not all of S's evidence. (187)

TW's proposed definition of evidence:

EV e is evidence for h for S if and only if S's evidence includes e and $P(h \mid e) > P(h)$.

Consequence:

> e is evidence for h only if e is evidence for itself. For if $P(h \mid e) > P(h)$, then $P(e)$ is neither 0 (otherwise $P(h \mid e)$ is ill defined) nor 1 (otherwise $P(h \mid e) = P(h)$). Hence $P(e \mid e)$ is well defined with the value 1, which is greater than $P(e)$, so e is evidence for e,[136] by EV with 'e' substituted for 'h' (187)

Objection: isn't is viscously circular for e to be evidence for itself?

TW: circular, but not *viscously* circular

[133] NB: conditional probabilities are not conditionals: $P(h \mid e) \neq p \rightarrow q$. TW is following the standard view, on which conditional probabilities are defined in terms of unconditional probabilities: $P(h \mid e) = {}^{P(h\&e)}/_{P(e)}$ when $P(e) > 0$, and undefined otherwise.

[134] NB: this way of thinking requires that evidence e be a proposition: since $P(h \mid e)$ is only defined when both h and e are propositions. See below.

[135] TW doesn't actually say why this implies that P shouldn't be a priori, but presumably it's because P's sensitivity to background information makes $P(h \mid e) > P(h)$ contingent on that background information, and since it's not a priori whether that background information obtains, it's not a priori that $P(h \mid e) > P(h)$ (even when it's true)

[136] BTM: suppose e is a logical truth that I know. Then by E=K, e is part of my evidence. Problem: probabilistic coherence requires that all logical truths have probability = 1, if $P(e) = 1$ then there is no h s.t. $P(h \mid e) > P(h)$. In other words, on TW's account, e is part of my evidence, but by condition (1) above, e isn't evidence *for anything*. What could it mean for something to be evidence that isn't evidence for something?

- Assuming E=K, in order for *e* to be evidence in the first place it must be known, and knowing that *e* might not be easy. So although it's trivially easy to have evidence for *e* once I have evidence for *e*, obtaining that evidence in the first place might not be trivial
- on this theory *h* is evidence for itself, but when asked why I believe *h*, I would never reply 'because *h*'. But that's no problem: we don't do so because it's conversationally inappropriate, not because it's not true. It is true.
  - to illustrate the point, TW cites Grice's example:

    > In answer to the question 'Who lives in the same house as Mary?' it would be conversationally inappropriate to cite Mary herself; nevertheless, it is true that Mary lives in the same house as Mary (Grice 1989). The question 'What is the evidence for *h*?' is often a challenge to the epistemic standing of *h* and related propositions. In some contexts the challenge is local, restricted to propositions derived in some way from *h*. In other contexts the challenge is global, extending to all propositions with the same kind of pedigree as *h*. In answering the question, one is expected not to cite propositions under challenge, since their status as evidence has been challenged. Thus when the question 'What is the evidence for *e*?' is meant as a challenge to the epistemic standing of *e*, one is expected not to cite *e* in response. (187-8, emphasis added)
- Objection: maybe it's inappropriate to treat *e* as evidence for itself because *e* is not evidence for itself. [This part is important, so I'll quote TW at length]:

  > The idea would be that the question 'What is the evidence for e?', meant as a challenge, creates a context in which e falls outside the extension of 'S's evidence'. But that seems too drastic. For example, suppose that a doctor asks you, 'Do you feel a tingling sensation?' and you answer, 'No.' **If you were asked 'What is your evidence for the proposition that you do not feel a tingling sensation?', you might be at a loss to answer, for the question seems to expect some further evidence for the proposition, and you might look in vain for such further evidence.** Nevertheless, when we assess the status of your claim that you did not feel a tingling sensation on your evidence, we do not exclude that proposition from your evidence. Its presence justified your claim...The point is just that challenging e by itself is not enough to exclude e from the extension of 'evidence'. (188; emphasis added)

BTM:

TW is relying on a suppressed premise here: that all evidence is propositional. Suppose that's right: then in order to find evidence for the proposition 'I'm feeling a tingling sensation' you'd need to find *some other proposition*. His point is that there is no other proposition that you could cite in this case.

But why think that the evidence must be a proposition? Typically, when I come to believe that I'm feeling a tingling sensation, my evidence is *the tingling sensation itself*. We would normally say that tingling sensations, and experiences more broadly, are not propositions, although many (all?) of them have propositional contents or accuracy conditions or something like that.

If that's right then I have evidence for the proposition 'I'm feeling a tingling sensation' which is distinct from that proposition, so I won't 'be at a loss' about how to answer the question 'what is your evidence that you feel a tingling sensation'. Furthermore, we generally don't ask for evidence that one is having a particular sensation (e.g. a tingle), so it would be inappropriate to further enquire what my evidence is that I feel a tingle.

TW will address the question of whether all evidence is propositional in §9.5 below.

[end BTM]

Important to distinguish between two nearby questions:

1. what is it for evidence *e* to support hypothesis *h*?
2. what is the nature of *e*? What counts as evidence?

EV is an attempt to answer question (1), and it might need revision.

TW's real concern is with (2).

It's an important question: if we can't get clear about what evidence is, then it's impossible to get clear about what evidence one has in a given scenario, so it's impossible to get clear on what one's evidence supports.[137]

We also need a theory of the nature of evidence in order to address questions of the underdetermination of theory by data (i.e. evidence): we can't evaluate supervenience claims unless we know what's in the supervenience base.

[137] Of course since TW thinks that E=K and that KK is false, he already thinks that one is not always in a position to now what one's evidence supports. Still, if he could establish that E=K, then with an answer to (1) we would be in a position to know what one's evidence would support in counterfactual situations, *when the descriptions of those situations includes a specification of what one knows*.

9.3 – Access to Evidence

Point of this section: 'explore our limited access to our evidence in the light of the equation E=K'

That access isn't perfect: if E=K, then E*p* = K*p*. If we had perfect access to our evidence, then E*p* implies KE*p*, which by E=K implies KK*p*. But the KK principle is false, so we don't have perfect access to our evidence.

But, we saw in chapter 8 that on no theory of evidence is it plausible to claim that we have perfect access to evidence, so this isn't a special problem for E=K.

Potential problem: if we don't know what our evidence is, then a regress looms: if I believe *h* on the basis of *e*, but my imperfect access to my own evidence *e* requires some further evidence that I have evidence *e′* for *e*, but my imperfect access to *e′* requires...

Distinction:

Call one's belief in *p explicitly* evidence-based if it is influenced by prior beliefs about the evidence for *p*.

Call one's belief in *p implicitly* evidence-based if it is appropriately causally sensitive[138] to the evidence for *p*.

In order for the regress to get going it would have to be the case that (i) the belief in question is explicitly evidenced-based, and (ii) the evidence for that evidence is explicitly evidence-based, and (iii) the evidence for that evidence for that evidence is explicitly evidence based...

But, that an *explicitly* evidence-based belief might be supported by evidence that is not itself *explicitly* evidence-based is consistent with E=K, so E=K has no particular problem here.

There's not even an apparent problem for *implicitly* evidence-based beliefs, since mere causal sensitivity of belief that *p* to evidence for *p* need not involve further beliefs.

Additional problem for explicitly evidence-based beliefs: Hume tells us to follow the rule:

Rule 1: Proportion your belief in *p* to your evidence for *p*

[138] S believes *p sensitively* iff (i) S believes *p*, and (ii) if *p* were false (in some nearby possible world), then S wouldn't believe *p*.

Combining Hume's rule with E=K, we get:

Rule 2: Proportion your belief in *p* to the support that *p* receives from your knowledge

As a practical matter, we do best following that rule when we follow:

Rule 3: Proportion your belief in *p* to the support that *p* receives from what you believe to be your knowledge

So our belief in *p* is *explicitly* rule-based. But presumably now we are required to apply Rule 3 to our evidence, and if the rule is applicable then it looks like we have a regress of explicit evidence-basedness after all.

TW: Rule 2 is a 'standard of correctness, not a description of action', so it's possible to fail to follow the rule correctly without thereby following some other rule. False beliefs (e.g. about your evidence) make it hard to follow rules, but that's not a problem for the rule as a standard of action.[139]

But overall this isn't so troubling, since we're 'often in a position to know whether we know *p*'.

When we want to check whether we know *p*, we look for reasons that bear on the truth of *p*,[140] rather than merely introspecting about whether we believe the evidence that *p*, as one would expect if we were really trying to follow Rule 3.

[139] NB that TW here is essentially discussing the motivation at the heart of the skeptic's argument from chapter 8: that ought implies can, and hence if we ought to believe what our evidence supports, then we must have the capacity to do so, and that requires access to our evidence.

[140] presumably my knowledge that reliable sources say that *p*, and my knowledge that I saw *p* myself, etc

9.4 – An Argument [for E=K]

TW's argument for E=K is this:

(P1) All evidence is propositional

(P2) All propositional evidence is knowledge

(P3) All knowledge is evidence

(C) All and only knowledge is evidence

The conclusion is equivalent the conjunction of the premises.

The subsequent sections defend the three premises.

9.5 – *Evidence as Propositional*

TW's claim: all evidence is propositional

That's false as a *descriptive* claim about what people regard as evidence: law courts regard bloody daggers as evidence; Quine regarded stimulations of sensory receptors as evidence; Huemer regards seemings as evidence, etc.

That's uninteresting as a stipulative definition of a technical term.

Other options?

'We can single out theoretical functions central to the ordinary concept *evidence*, and ask what serves them.'

Signpost: TW first makes a positive case that evidence is propositional, and then defends the position from objections. We (and he) start with the positive case:

Consider what theoretical function evidence serves in various cases:

Inference to the Best Explanation: the hypothesis that best explains the evidence is confirmed. So evidence is the kind of thing that can be explained by a hypothesis.

Explanations take the form: '[evidence] because [hypothesis]', and in order to be grammatical we need to put that-clauses in both sets of brackets. That clauses express propositions, so evidence must be propositional.

NB: can't explain objects (e.g. Albania): we can make sense of the injunction '*Explain Albanina!*' only if we reinterpret it as as a demand to explain some fact about Albania.

TW: we can't explain sensations either: the sensation in my throat isn't the evidence that's explained by my getting a cold (the hypothesis), *that I have that tickle in my throat* is what's being explained. But *that I have a tickle in my throat* is a proposition.

BTM: is that fair? Must it really be *evidence* that's explained by the hypothesis? One could just as easily take the canonical form of an explanation to be '[putative fact] because [hypothesis]', and in that case TW's conclusion would require an additional argument, such as:

1. the things to be explained are all propositions

2. the things to be explained are all evidence
3. so, all evidence is propositional.

But (2) is highly dubious. One thing you might explain is some fact about Albania's economic performance relative to that of its neighbors. Is that fact really part of your evidence? It seems more natural to say that it's a conclusion that you've reached on the basis of some other evidence (observations? testimony?), and *that* evidence might not be propositional.

Zoom out a bit. TW's view blurs the distinction between evidence and the conclusions that it supports: if I come to know h on the basis of my evidence e, then h is part of my evidence (by E=K). One might reasonably find that to be an implausible consequence of the view – a kind of pan-foundationalism about knowledge – rather than a reasonable (implicit) assumption to be used to motivate the view in the first place!

Zooming out quite a bit more, it's instructive to consider whether TW is arguing against his opponents from neutral premises vs attempting to presenting a coherent alternative to be judged on its own merits.

end BTM

So far these observations are meant to support the hypothesis that evidence is propositional. But these cases also tell us what it means to have a proposition as part of your evidence: one can use a hypothesis to explain some evidence only if you grasp (?) the evidence. So only propositions one grasps are counted among one's evidence.[141,142]

[141] What exactly does it mean to 'grasp' a proposition? Presumably TW will tell us below.

[142] only propositions one grasps are counted among one's evidence, or only *and all* propositions one grasps are among one's evidence?

Probabilistic Reasoning:

Sometimes it's useful to compare how strongly a body of evidence e confirms hypothesis h compared to how strongly e confirms some other hypothesis h*

One way to do that is to compare $P(h \mid e)$ and $P(h^* \mid e)$. But there's a drawback: if $P(h) > P(h^*)$ then it might be the case that $P(h \mid e) > P(h^* \mid e)$ for reasons independent of the evidential impact of e upon h and h^*.

Here's another way: One way to write Bayes's Theorem is this:

$P(h \mid e) = \frac{P(e|h)}{P(e)} P(h)$

Focus first on the left side of the equation. Bayesians think that when I obtain evidence e, I should conditionalize according to the following rule:

Conditionalization: $P_e(h) = P(h \mid e)$

Here $P_e(h)$ is just the Bayesian's way of expressing the credence I ought to have in h once I've obtained evidence e and updated (conditionalized) appropriately.

Now look at the right side of Bayes's Theorem. The fraction $\frac{P(e|h)}{P(e)}$ gives us the *Bayesian Multiplier*: a value that we can multiply by $P(h)$ (my credence in h before I obtained evidence e) to yield $P_e(h)$ (my credence in h after I obtain evidence e and conditionalize appropriately).

General upshot: looking to Bayesian Multipliers provides a way of comparing e's confirmation of h to its confirmation of h^*: compare $\frac{P(e|h)}{P(e)}$ to $\frac{P(e|h^*)}{P(e)}$. This avoids the previous problem.

TW's upshot: in order to calculate the Bayesian Multiplier for h I need to have a credence in evidence e.

It's just this last point that TW cares about: for his purposes, what's important is that according to what's probably our best theory of formal epistemology, we need to assign credences in evidence. **But the only things that can be assigned credences are propositions.** So evidence consists of propositions.

Possibility that evidence might be inconsistent with the hypothesis

We often reject hypotheses for being inconsistent with the evidence. But if h is inconsistent with e, then it must be possible to deduce $\neg h$ from e, and the premises of deductions are propositions.[143] So evidence must be propositional.

[143] BTM: is this really necessary? One might instead think that h is inconsistent with my evidence when my evidence supports $\neg h$, leaving open the question of the propositionality of my evidence.

Signpost: This completes TW's positive case that evidence is propositional. Now we turn to his defense against objections. The primary objection is evidence-as-propositions-which-we-grasp account can't vindicate the intuition that perceptual experiences (which are not propositions) are a form of evidence.

TW begins:

> Experiences provide evidence; they do not consist of propositions. So much is obvious. But to provide something is not to consist of it. The question is whether experiences provide evidence just by conferring the status of evidence on propositions. On that view, consistent with E = K, the evidence for an hypothesis *h* consists of propositions $e_1, ..., e_n$, which count as evidence for one only because one is undergoing a perceptual experience ε. As a limiting case, *h* might be e_i. The threatening alternative is that ε can itself be evidence for *h*, without the mediation of any such $e_1, ..., e_n$. Both views permit ε to have a non-propositional, non-conceptual content, but only the latter permits that content to function directly as evidence. (197)

TW contrasts two views of the epistemic role of experiences. Since he tells us that '[t]he question is whether experiences provide evidence just by conferring the status of evidence on propositions', presumably the two views are characterized by the way they answer 'the question'.

Annoyingly, TW complicates things by considering whether the evidence for *h* consists of propositions $e_1, ..., e_n$, which are themselves evidence due to ε. But he allows that *h* might be one of the e_i's, in which case we're just considering whether e_i is evidence due to ε. The evidential relationship between the e_i's and *h* when *h* is not one of the e_i's is inferential, and inferential support between propositions isn't what's at issue here, so let's ignore *h* and instead focus on the relationship between ε and e_i.

View 1: e_i 'count[s] as evidence for one only because one is undergoing perceptual experience ε'

View 2: the non-conceptual content of ε is itself evidence for e_i

One possible objection to View 1 is that the richness of experience often seems to outstrip our visual and conceptual resources, so the needed propositions just aren't available to do the necessary work.

TW: obviously, it's not always possible to *convey* our perceptual evidence in a straightforward way; often we rely on demonstratives. Example: I have a visual experience of a mountain, and it provides evidence for a belief about the shape of the mountain. Which shape? Hard to say, other than '*that* shape' (pointing at mountain).

But just because it's hard to convey a content doesn't mean that there isn't one, or that one hasn't grasped it: maybe you have to have the visual experience yourself in order to grasp the propositional constituent denoted by 'that shape'. The resulting proposition will be contingent and a posteriori, just as one would expect from perceptual knowledge.

Second possible objection to View 1: in favorable circumstances, when I have an experience with content *there's snow*, I come to know *there's snow*, so that proposition is part of my evidence. But what's my evidence when conditions are unfavorable, e.g. when I'm a BIV? In that case there isn't any snow, so I can't know that there's snow, so by BIV *there's snow* can't be part of my evidence.

TW: In that case you could still know *it appears that there's snow*, which would provide some evidence for *there's snow*, even though *there's snow* is cannot be knowledge or evidence because it's false. NB: in this case you wouldn't know that *there's snow* isn't part of your evidence, but that's just an instance of KK failure.

Objection to the response to the second objection: what about animals and small children who can't grasp the distinction between appearance and reality? If you lack the concept *appearance* then you can't know that *things appear thusly*, so you can't have evidence in the form of propositions about how things appear. So what evidence does a squirrel-BIV obtain from their perceptual experience?

TW: the squirrel-BIV might have propositional evidence such as *the situation is like this* (mentally pointing). Or she might have no evidence at all (TW thinks this is the case for very simple creatures).

[BTM] Some lingering issues:

1. Variant of the cases: Jeffrey's case of viewing a cloth under poor lighting, distributing your confidence between *it's blue, it's green,* and *it's violet*. Here the experience doesn't outstrip our conceptual resources, but neither does it provide any of those propositions as evidence (in TW's sense): in that case $P(blue)$ (etc) would have to be 1, which it isn't. Can TW's responses handle this case?
2. What's the criticism of View 2, that non-conceptual content plays some evidential role? Is the falsity of View 2 supposed to follow from the central-functions-of-evidence argument at the beginning of the section?
3. What's the substantive difference between 'ε is evidence for e'

and 'ε makes e evidence'? On the former, ε is evidence, and since ε is an experience and not a proposition, TW must deny that this is possible. But what does it even mean to say the latter — that 'ε makes e evidence' — if not that ε is evidence for e? Is something important at stake, or are we just playing games with the word 'evidence' here?

9.6 – Propositional Evidence as Knowledge

Point of this section: argue that *e* is evidence for S iff S knows *e*

The argument here mirrors TW's positive account from §9.5: the central function of evidence is best served by propositions that are *known*, so all and only known propositions are evidence.

Claim: JTB in *e* is not enough to make *e* evidence; K is required

> Suppose that balls are drawn from a bag, with replacement... I have seen draws 1 to n; each was red (produced a red ball). I have not yet seen draw $n + 1$. I reason probabilistically, and form a justified belief that draw $n + 1$ was red too. My belief is in fact true. But I do not know that draw $n + 1$ was red. Consider two false hypotheses:
>
> *h:* Draws 1 to n were red; draw $n + 1$ was black.
>
> *h*:* Draw 1 was black; draws 2 to $n + 1$ were red.
>
> It is natural to say that *h* is consistent with my evidence and that *h** is not. In particular, it is consistent with my evidence that draw $n + 1$ was black; it is not consistent with my evidence that draw 1 was black. Thus my evidence does not include the proposition that draw $n + 1$ was red.

In this case I have a JTB that draw $n + 1$ is red, but that's not part of my evidence, so having a JTB in *p* is not sufficient for *p* being part of my evidence. What's missing? TW: knowledge.[144]

[144] NB that this is a version of the lottery paradox, and some will say that you do in fact know that $n + 1$ is red. To stoke that intuition, imagine that n is a really, really large number.

TW generalizes the objection:

> If evidence required only justified true belief, or some other good cognitive status short of knowledge, then a critical mass of evidence could set off a kind of chain reaction. Our known evidence justifies belief in various true hypotheses; they would count as evidence too, so this larger evidence set would justify belief in still more true hypotheses, which would in turn count as further evidence .... The result would be very different from our present conception of evidence. (201)

BTM: won't this also be the case if we take all known propositions to be evidence? After all, our known evidence doesn't just 'justif[y] belief in various true hypotheses' – it often yields *knowledge* of those various true hypotheses. For example, inductive inference of p from evidence sometimes results in knowledge that p, which becomes part of my evidence (by E=K). Since the inference is inductive, $\neg p$ consistent with my evidence, but that's impossible when p is itself part of my evidence (since p is certainly not consistent with $\neg p$). Assuming that knowledge entails JTB but not vice versa, this effect will be less extreme if E=K, but we are left with the same general problem.

[End BTM]

First consequence of E=K: all evidence propositions are *true*[145]

Knowledge is factive — K(p) implies p — so E=K implies that all evidence propositions are true.

TW: that's a good thing

1. 'if one's evidence included falsehoods, it would rule out some truths, by being inconsistent with them. One's evidence may make some truths improbable, but it should not exclude any outright.'
2. 'There is no suggestion, of course, that if e is evidence for h then h is true.'
3. 'The rival view, that a false proposition can become evidence through a sufficient appearance of truth, gains most of its appeal from the [rejected] assumption... that we must have an infallible way of identifying our evidence.'
4. Why should we care about having evidence, and about believing what our evidence supports? If all evidence consists in true propositions, the answer is clear: it facilitates true belief.[146]

Second consequence of E=K: all evidence propositions are *believed*[147]

Problem: intuitively, we can obtain perceptual evidence even when we don't believe what our senses are telling us:

> it is perceptually apparent to me that it is snowing; I am not hallucinating; but since I know that I have taken a drug which has a 50 per cent chance of causing me to hallucinate, I am not in a position to know that it is snowing. According to the radical critic[148], **my evidence nevertheless includes the proposition that it is snowing**, because it is perceptually apparent to me that it is snowing; thus my evidence

[145] Since TW thinks that he's already established that all evidence is propositions, it follows that all evidence consists in true propositions

[146] The contrast TW has in mind is one on which evidence consists in a coherent body of non-factive states or propositions, which might not track truth at all (as with the BIV).

[147] As before, since TW thinks that he's already established that all evidence is propositions, it follows that all evidence consists in believed propositions

[148] radical relative to TW, that is, not in some objective sense

> is inconsistent with the hypothesis that I am hallucinating and it is not snowing, even though, for all I am in a position to know, that hypothesis is true. According to E =K, my evidence includes at best the proposition that it appear to be snowing. Surely[149], if I proportion my belief to my evidence, I shall not dismiss the hypothesis that I am hallucinating and it is not snowing. E = K gives the better verdict. Perceptual cases do not show that we sometimes fail to believe our evidence.

[149] Ratcheting up the rhetoric!

BTM: There's a much more plausible objection in the neighborhood.

TW is assuming a picture on which all evidence is propositional, and having a proposition as part of your evidence is an all-or-nothing matter. He's supposing that the 'radical critic' shares this picture, and that the disagreement is simply over whether a perceptual experience as of *p* makes *p* part of my evidence whether or not one believes it.

Given TW's background assumptions, the 'radical critic' is forced to say not only that *it's snowing* is evidence, but that having that evidence requires one to reject the proposition that one is hallucinating, even though one has good grounds to believe that one *is* hallucinating and that it's not snowing.

It's unfortunate that TW picked such an implausible target, one that's made out of straw and shaped like a man.

Here's a more reasonable alternative position: under normal circumstances, when I have a (non-factive) experience as of snow, I should become more confident that there's snow, and I will often come to know that there's snow. When I antecedently believe that there's a 50 percent chance that I'm hallucinating, then my experience doesn't result in knowledge. But that doesn't mean that it's evidentially moot vis-à-vis *there's snow:* rather, the experience should lead me to become somewhat more confident that there's snow, but not too much. In that case I would not be in a position to infer that I'm not hallucinating from the experience as of snow.

This picture is inconsistent with E=K for a couple of reasons. First, it suggests that the *experience* as of snow provides some reason to become more confident that *there's snow*, which in turn suggests that an experience (rather than a proposition) is functioning as evidence. Second, even if one insists that the proposition *it's snowing* is evidence, it's not all-or-nothing evidence: I don't 'have it' in some binary way.

Yet this picture seems very plausible: it's at least possible for an experience to provide some evidence — some rational imperative to change my (partial?) beliefs — even though I don't entirely believe what my senses are telling me.

Moreover, note that if E=K, then in many cases we are in a position to bootstrap[150] our way out of skeptical worries. If I have an experience as of snow, and thereby come to know *there's snow*, then according to TW I'm in a position to know that I wasn't hallucinating.[151]

[150] *bootstrapping* is the process of using a perceptual faculty to come to know that *p* and also that *I've had an experience to know that p*, and on that basis ruling out skeptical hypotheses

[151] NB this is a really tricky issue: TW might simply claim that unless I was already certain that I'm not hallucinating, the experience doesn't result in knowledge, in which case the anti-skeptical inference is irrelevant – you already know the conclusion.

9.7 – Knowledge as Evidence

Recall TW's argument for E=K:

(P1) All evidence is propositional.

(P2) All propositional evidence is knowledge.

(P3) All knowledge is evidence.

(C) All and only knowledge is evidence.

The point of this section is to establish (P3)

First objection to (P3): there's a spectrum of views from extreme holism (= bodies of evidence support conclusions all together – it's impossible to isolate the contribution of a single piece of evidence[152]) to extreme atomism (= every piece of evidence makes an individual contribution which is insensitive to the contributions of other pieces of evidence). The objection is that E=K pushes us toward the extreme holism end of the spectrum which is implauisble: if my evidence includes *there are two cats in the room* and *I feel hungry*, then my conclusion *there's at least one cat in the room* is much more strongly supported by the former than the latter. 'The concern is... that if all one's knowledge is treated as a single body of evidence, its internal evidential interconnections will be obliterated, and therefore that such an account would falsify the nature of our knowledge.' (204)

[152] Essentially Quine's web of belief, but without sensory stimulations affecting the periphery

Response: EV (from §9.2) illustrates how there can be evidential interconnections within a body of evidence:

EV e is evidence for h for S if and only if S's evidence includes e and $P(h \mid e) > P(h)$.

P is a function from bodies of evidence (e) to credences in hypotheses (h). If we exclude some particular proposition e^* from body of evidence e we can compare:

*EV** $e*$ is evidence for h for S if and only if S's evidence e includes $e*$ and $P_e(h \mid e^*) > P_e(h)$.[153]

[153] TW actually means for partial bodies of evidence to be already built into P, so my revisions only make explicit what was already (very) implicit

$P(h \mid e) > P(h)$ compares the probability of h on no evidence to the probability of of h on evidence e. On this picture the evidential significance of e is considered independently of any possible holistic effects from the rest of one's body of evidence.

In contrast, $P_e(h \mid e^*) > P_e(h)$ compares the probability of h given background evidence e together with new evidence e^* to the probability of h given background evidence e without new evidence e^*. Here the holistic nature of e^*'s evidential significance for h are already built into the probability function.

Second objection to (P3): we are rarely certain of what we know, so if E=K then we can rarely be certain of our evidence, in the sense that our certain will never diminish in the future. This will be rejected by Cartesians and some Bayesians.

Response: it's implausible to think that any significant part of our evidence satisfies this demand for certainty. We might forget our evidence, or we might obtain evidence to the contrary.

BTM: NB that TW hasn't even attempted a positive case for (P3) – he's only defended it from a couple of objections.

Here's another objection: suppose I know that p and I know that q, and on that basis I infer and come to know that $p\&q$. By E=K, $p\&q$ is now part of my evidence. Is that plausible? Can a proposition inferred from evidence be evidence itself? Given the moderate foundationalist picture that TW likes, shouldn't there be more of an asymmetry between foundational knowledge and that which is supported by it? Is this just a version of this section's first objection?

9.8 – Non-pragmatic Justification

Are there non-truth-directed forms of justification?

Maybe: perhaps the athlete preforms better when she believes that she's the best, and this gives her a form of justification for so-believing. But this isn't the type of justification that TW cares about.

10 - Evidential Probability

10.1 – Vague Probability

Goal of the chapter: to formulate a theory of 'evidential probability' that can account for the fact that even the propositions we treat as evidence are uncertain.

Three standard interpretations of probability:[154]

[154] NB that each of these are identity claims – they seek to explain what probabilities *are*, not just to make some weaker observation about probabililties.

Frequency: probabilities are actual or infinite-sequence frequencies

Problem: non-repeatable events (Obama winning the '08 election, a coin that lands heads in a particular instance) have probabilities of 1 or 0, but often that's not what the evidence supports

NB: observed frequencies might provide *good evidence* about actual or counterfactual frequencies, but the frequentist's claim is one of *identity*, which is much stronger.

Propensity: probabilities are properties of objects, like dispositions [TW doesn't consider this interpretation]

Problem: Propensities of non-repeatable events needn't have probabilities of 1 or 0, but as with frequency view there's no clear connection between evidential support and propensities (subject to the same NB above)

Degree of belief: (also: subjective probability, credence) a subjective mental state of the agent measuring degree of confidence

Problem: the fact that people become more confident in h upon learning that p doesn't imply that e is evidence for h – maybe people are just being irrational

Proposed solution: probabilities are the credences of perfectly rational agents, not potentially irrational agents like us.

Note: much of formal epistemology is dedicated to spelling out the norms governing credences. Those norms tend to be unrealistic for normal people (e.g. they tend to imply that logical omniscience is a norm of rationality), so it's quite common for formal epistemologists to retreat to this proposed solution: the norms in question apply only to ideal agents, and our obligations are to approximate those norms to the degree possible (this last bit isn't well understood).

Problem with the proposal: q is a highly non-obivous logical truth, so non-obvious that we have strong evidence that no one will have a rational high credence in q.[155]. But since q is a logical truth, a perfectly rational agent would have a credence of 1 in q*, and so

$$P(q \text{ \& no one will have a high credence in } q)$$
$$= P(\text{no one will have a high credence in } q).$$

Since *no one will have a high credence in q* is part of the evidence, that credence will be very high.

[155] perhaps we know that no one has ever had a high rational credence in the past, and we know that the world is about to end in a nuclear holocaust

Two reasons this is problematic:

1. It's Moore-paradoxical, so no ideally rational agent would ever believe it, so a perfectly rational agent would never believe it.[156] Presumably she'd know that someone has great evidence for q because she herself has great credence in q, so she'd have a very *low* credence in *q&no one has great credence in q*.
2. But it's perfectly rational for us (rationally imperfect as we are) to be highly confidence on the basis our evidence that no one will ever have high credence in p, so the evidential probability of ϕon evidence e is *not* the credence that a rational agent would have in ϕgiven evidence e

[156] Note that agents ideally rational in a strictly *formal* sense might believe all the Moore-paradoxical propositions they please – whatever the norm ruling out belief in Moore-paradoxical propositions turns out to be, it's not a formal one.

This shows that the probability of *q & no one will have a high credence in q* on our evidence is not the same as an ideal agent's credence in that proposition would be. To generalize: the probability of a hypothesis on our evidence — evidential probabilities — are not the credences of rational agents.

*Background: *probabilism* is the thesis that one norm of coherence for belief is probabilistic coherence. Why accept that thesis?

A Dutch Book is a series of bets s.t. no matter how things turn out, you lose money. If you're habitually inclined to make such bets then you're a money pump: people can keep making these bets with you and you'll continually pay out more than you bring in. There's something pathological about being a money pump: it's a sign of practical irrationality. A Dutch Book *Argument* aims to show that probabilistically incoherent beliefs are irrational because[157] they turn you into a money pump.

[157] given further plausible assumptions from decision theory about how beliefs translate and values into betting behavior

Example:

I'm 70% confident that the coin will land heads and 70% confident that it will land tails. My beliefs are incoherent: I now think that there's a 140% chance that it will land either heads or tails.

If I'm 70% confident that the coin will land heads, then I should be willing to pay $7 for a bet that pays $10 if it really does come up heads and nothing if it comes up tails: in that case I think that there's a 70% chance that I'll win \$10 and a 30% chance I'll win \$0, and (\$10 x .7) + (\$0 x .3) = \$7.

For parallel reasons I'll be willing to pay \$7 for a bet that pays \$10 if it comes up tails and nothing if it comes up tails.

But that means that I'll be willing to pay \$14 to take both bets, even though that's guaranteed to lose me \$4 overall. So I'm a money pump.

So, decision theory requires that credences be probabilistically coherent.

So what *are* evidential probabilities?

> We should resist demands for an operational definition; such demands are as damaging in the philosophy of science as they are in science itself. To require mathematicians to give a precise definition of 'set' would be to abolish set theory. Sometimes the best policy is to go ahead and theorize with a vague but powerful notion. (211)

and

> Consider an analogy. The concept of possibility is vague and cannot be defined syntactically. But that does not show that it is spurious. In fact, it is indispensable. Moreover, we know some sharp structural constraints on it: for example, that a disjunction is possible if and only if at least one of its disjuncts is possible. The present suggestion is that probability is in the same boat as possibility, and not too much the worse for that. (211)

10.2 – Uncertain Evidence

In order to set up his formal system for expressing evidential probabilities, TW first criticizes rivals.

Subjective Bayesians think that new evidence is incorporated into one's credence function by a process of *conditionalization:*

BCOND $P_{new}(h) = P_{old}(h \mid e) = P_{old}(h\&e)/P_{old}(e)$

Remember P is a credence function: a function from propositions to numbers between 0 and 1 representing the agent's degree of confidence in those propositions.

Because functions aren't allowed to produce two different values for a single argument, if one's credence in h is going to be different at t_1 and t_2, the agent must have different credence functions at those different times. BCOND relates the credence functions that one has a different times: P_{old} is the credence function that one has before obtaining evidence e and P_{new} is the credence function that one has after obtaining e.

First criticism of BCOND: suppose that all credence revisions proceed via BCOND. In that case, once one conditionalizes on e, $P_{new}(e) = P_{new}(e \mid e) = 1$. The problem is that once a proposition has credence of 1, it's impossible to reduce that credence in light of any future evidence: if $P(e) = 1$ then for any future evidence f, $P(e \mid f) = P(e\&f)/P(f) = P(f/f) = 1$. Call this the *Invincibility of Evidence* objection.

But if conditionalizing on e requires having the highest possible confidence in it, and if it's impossible to reduce that credence on any future evidence, then my confidence in e is indefeasible, which seems wrong.

Second criticism of BCOND: in response to the first puzzle, some Bayesians retreat to a conception of evidence as phenomenal states on the idea that we can't be mistaken about those. But this move has trouble capturing the intersubjectivity of evidence as required in the sciences.

Bayesians can (sort-of) avoid this problem if the propositions updated upon are about the world, but not if they're about mental states. But propositions about the world are defeasible, while evidence propositions are not. Potential solution: generalize BCOND into JCOND:

JCOND $P_{new}(h) = \sum_i P_{old}(h \mid e_i) P_{new}(e_i)$

NB: BCOND is a special case of JCOND: the case in which there are only two propositions in the input partition, one of them weighted to 1 and the other to 0. Usually this will be a proposition and its negation, e.g. e and $\neg e$. It is common to simply fail to mention the partition elements weighted to 0, in which case one ends up updating on e alone; that's precisely what Classical Bayesians – those committed to BCOND – do.

Jeffrey conditionalization allows the credence in evidence propositions to change without going all the way to 1, so problem averted.[158]

[158] But if a credence *does* happen to go all the way to 1, it's stuck there, just like with BCOND.

TW's criticism of JCOND: it offers no account of where the input partitions come from. This is unlike BCOND, on which the inputs are propositions, which can in turn be identified with one's propositional evidence, which can in turn be supplemented with an auxiliary theory of evidence such as TW's E=K hypothesis.

The invincibility of evidence objection comes from two claims:

PROPOSITIONALITY The evidential probability of a proposition is its probability conditional on the evidence propositions.

MONOTONICITY Once a proposition has evidential probability 1, it keeps it thereafter.

PROPOSITIONALITY entails that evidence propositions have probability 1, and MONOTONICITY ensures that they keep that probability.

JCOND rejects PROPOSITIONALITY but keeps MONOTONICITY

TW's suggestion: reject MONOTONICITY and keep PROPOSITIONALITY

Why reject MONOTONICITY? It implies an evidential asymmetry: we can gain new certainties but never lose them. But that's false: it's possible to forget some of your evidence, in which case it's probability should be reduced

TW's suggestion:

ECOND $P_\alpha(h) = P(h \mid e_\alpha) = P(h\&e_\alpha)/P_{old}(e_\alpha)$

P (with no subscript) is the One True Probability Function, representing the probability that each proposition should have in the absence of all evidence, i.e. it's *initial plausibility.*[159]

P_α is the probability function representing the credence function one ought to have in a case α, in which one has total body of evidence e_α.

P_α is equivalent to what would happen if you start out with P and then update upon e_α using BCOND or JCOND[160].

TW's objection to BCOND and JCOND is that they impose implausible constraints upon how the probability of propositions change over time: they prohibit future reductions in propositions that now have a credence of 1. On these rules, certainty is cumulative: once gained it can never be lost. The problem is exacerbated for BCOND because it requires that all evidence propositions be certain, i.e. have a credence of 1.

NB that ECOND avoids the problem by imposing no constraints upon how the probabilities of propositions can change over time. If

[159] NB: Bayesians have something similar: the credence function that one accepts in the absence of any evidence. Generally Bayesians accept many different 'starting credence functions' as epistemically permissible – the details are contentious (we can talk more about this if you'd like), but they all agree that rationally permissible starting credence functions must be probabilistically coherent (i.e. they must be *probability functions.* But the Bayesian's initial credence function is making claims about beliefs, or credences. What are evidential probabilities in the absence of evidence? Are we back to frequencies or propensities?

[160] e_α determines a partition, so no problem updating via JCOND

at t_1 some proposition e is part of the evidence set e_α, then $P(e \& e_\alpha)/P_{old}(e_\alpha) = P(e_\alpha)/P_{old}(e_\alpha) = 1$, so $P(e \mid e_\alpha) = 1$, so $P_\alpha(e) = 1$. But if at t_2 my evidence β does not include e, then there's no constraint upon the probability of e at t_2. That's because the rule imposes no constraints at all upon what evidence one possesses at a given moment — that's determined outside of the model — and since we're told nothing about the nature of P, as long as e is not in β, the model imposes no constraints upon $P_\beta(e)$.

But, if no evidence is lost between t_1 and t_2[161] then $P_\beta(h) = P_\alpha(h \mid f)$, where f is just the conjunction of all of the propositions that are in β and not in α. When this happens the transition from P_α to P_β proceeds exactly as it would according to BCOND, which is itself a special case of ECOND.[162]

[161] and hence that the evidence in α is a proper subset of the evidence in β

[162] BCOND is a special case of ECOND in these situations because when the change in evidence is cumulative, they give the exact same prescription for updating probabilities; BCOND thinks all changes of evidence is cumulative, so it thinks that all updates are of this sort, but ECOND allows non-cumulative evidence changes, so it can handle a wider range of cases.

10.3 – Evidence and Knowledge

TW thinks that without an account of the nature of propositional evidence, formal epistemology is 'empty'.

TW thinks that your evidence consists in all the propositions that you know: E=K

He also thinks that subjective Bayesians think that your evidence consists in all the propositions that you *believe*: E=B

Problem with E=B: you could manipulate your evidence by manipulating your beliefs (assuming that's possible), thereby manipulating what it's rational for you to believe

BTM: TW presents this as a problem for subjective Bayesianism, but his view has a similar problem. Assume E=K. K entails B, so if some piece of evidence is inconvenient then by refusing to believe[163] it would ensure that I don't know it, so it's not part of my evidence. Of course this only works in one direction: by choosing to believe something I don't thereby come to know it, so although I could lose evidence in this way, I can't gain it, so the problem is worse for E=B than for E=K.

[163] this objection depends on belief voluntarism, but so does TW's. NB that both my version and TW's are evidence of a broader phenomenon of people manipulating their evidence in order to manipulate what their evidence supports. This is possible on a broad range of theories of evidence, and it's not entirely clear what to say about it, other than that it clearly involves some sort of irrationality.

But there's a deeper problem with this objection. Subjective Bayesians needn't accept E=B. Bayesianism imposes certain coherence norms on beliefs, but it doesn't imply that there aren't other norms of belief. It may be that one should conditionalize upon only rational beliefs, and in that case TW's objection fails (assuming that I can't generally be rational in believing p when I will myself to believe it).

End BTM.

The rest of the section rehashes previous material about how it's possible to follow a rule even it you're not always in a position to be able to do so, and hence it's not a problem that we're not always in a position to conform our beliefs in ECOND + E=K due to the fact that we're not always in a position to know what we know.

10.4 – Epistemic Accessibility

Point of this section is to utilize epistemic modal logic to develop a formal framework for combining ECOND and E=K. This logic has the following features:

- rather than cases, which are centered on specific agents, propositions are true at worlds, which are not
  - worlds are equivalent to mutually exclusive and jointly exhaustive sets of propositions
  - worlds needn't actually be possible: we allow worlds in which water is composed of XYZ (*water*, not twin-water)[164]
- a priori probability function P specifies the evidential probability that each world has conditional on no evidence.
  - The sum of the probabilities of all worlds is 1.
  - The probability of any specific proposition q is the sum of the probabilities of all of the worlds in which q is true.
  - If q is true in all possible worlds its probability is 1.
- Worlds are related by an accessibility relation R. A world w' is accessible to S from w iff every proposition that S knows in w is true at w'. In other words, R connects this world with all of the worlds that are consistent with the evidence that I have here.
  - Accessibility relations have directions: just because wRw', it doesn't follow that $w'Rw$. We can impose that (symmetry) condition if we want, but that's a further step (see below).
  - if some proposition p is consistent with what I know, then there exists a world accessible to me at which p is true. This provides the semantics for $\Diamond$: $\Diamond p$ is true at w iff there exists a world w' s.t. p is true at w' and wRw' (we allow that w might be identical to w')

[164] TW says that the worlds needn't be *metaphysically* possible. Though he says nothing about *logical* impossible worlds, he's going to need to include them if he want's to allow that a person would not be logically omniscient. Let's keep an eye on this issue as the section unfolds.

- p follows from what I know iff p is true in every world accessible from w. This provides the semantics for $\Box$: $\Box p$ is true at w iff for any world w', if wRw' then p is true at w'.
- Since knowledge is factive, each world sees (i.e. is accessible to) itself; R is *reflexive*.

- With this framework in place, we can combine ECOND and E=K
  - take e_w to be S's evidence in w
  - by E=K, e_w consists in all the propositions that S knows at w
  - by the factivity of knowledge, e_w is true at w
  - since e_w consists in all the propositions that S knows in w, e_w is true in all the worlds accessible to w (by our account of the accessibility relation)
  - P_w is the function specifying S's evidential probability for every proposition in w.
  - by ECOND, $P_w(\cdot) = P(\cdot \mid e_w)$. Intuitively, this says that the evidential probability of p for a person with evidence e_w is equal to the weighted sum the probabilities of p in each of the worlds consistent with evidence.

BTM: Interesting consequence of what's been said. TW has committed to the claim that: p follows from e_w iff p is true in every world consistent with e_w. In that case the evidential probability of p at w is 1. But, he also says that 'one need not know that which follows from what one knows.' (225) So on this account, although all known propositions have evidential probability 1, not all propositions with evidential probability 1 are known.

That's not an objection, just an observation. Remember that unlike the subjective Bayesian, TW's probabilities are not mental states like belief or (maybe) knowledge. But this leaves open the question of what rational demands the account imposes on agents: if p has an evidential probability of 1 on my evidence, what should my attitude be toward p? What are the rational norms governing evidential probabilities? The straightforward answer would be to say that if p has an evidential probability of 1 then S should believe it, but in that case it looks like rationality requires logical omniscience, a requirement that TW rejects: 'The account will not assume any general principle about knowledge, except that a proposition is true in any world in which it is known. In particular, it will not assume logical omniscience; if p and q are true in exactly the same worlds, one may know p and not know q.' (224)[165]

There's a second way of interpreting what's going on here. I've been assuming that the 'follows from' in the semantics of $\Box$ and $\Diamond$ is equivalent to 'is a logical consequence of'. If that's right then once e_w is determined, many facts about the worlds consistent with e_w are determined as well. In particular, if p really is a logical consequence of e_w then the worlds consistent with e_w must be p-worlds. But TW might instead be appealing to facts about the worlds in order to define the expression 'follows from'. In that case p could be a logical consequence of e_w, but some of the worlds consistent with e_w could be $\neg p$ worlds (remember that the worlds don't actually have to be *possible* worlds). In that case, what TW has told us about what 'follows from' what implies that p does not in fact follow from e_w, in spite of their logical relationship. But if some of the worlds consistent with e_w are $\neg p$ worlds, then $P_w(p) \leq 1$, so the puzzle of why probability 1 propositions needn't be known (or believed) does not arise.

[165] I'm not sure what to make of the bit after the semi-colon: do all permissible failures of logical omniscience have that form? Am I still required to know all of the logical truths that don't share that form? NB that if I know one logical truth then one might think that all questions of logical omniscience are of this form.

I find TW in KAIL 10 to be very unclear on this point, but in his response to Kaplan he's much more explicit: the connection between evidential probabilities and credences is pretty weak. In particular, p's having a probability of 1 conditional on my evidence *does not* require that I have a credence of 1 in p. Of course this leaves open the question of what evidential probabilities mean for credences...

[End BTM]

In this framework, constraints on R are constraints upon knowledge: Kp implies $\Box p$. What else can be said about R?

TW has told us that R is reflexive. This ensures that in our logic $\Box p \rightarrow p$ (this is the 'M' axiom). Suppose that's false, and that in w I know that p. That implies $\Box p$. A failure of reflexivity means that p would be false in w, i.e. that in w I know something that's false. But knowledge is factive, so that's impossible. So R must be reflexive.

He also tells us that R can't be transitive. Transitivity ensures that $\Box p \rightarrow \Box\Box p$ (this is the '4' axiom). This is essentially the KK principle (actually something slightly weaker, but presumably TW thinks that's false too).

He also tentatively rejects the symmetry[166] of R, which ensure that $p \to \Box\Diamond p$ (this is the 'B' axiom). This says that if p is true then in every world w' consistent with my evidence, there is at least one world w'' s.t $w'Rw''$ and p is true at w''. Problem case:

> Let x be a world in which one has ordinary perceptual knowledge that the ball taken from the bag is black. In some world w, the ball taken from the bag is red, but freak lighting conditions cause it to look black, and everything which one knows is consistent with the hypothesis that one is in x. Thus x is accessible from w, because every proposition which one knows in w is true in x; but w is not accessible from x, because the proposition that the ball taken from the bag is black, which one knows in x, is false in w. Let p be the proposition that the ball taken from the bag is red. In w, p is true, but that p is consistent with what one knows does not follow from what one knows, for what one knows is consistent with the hypothesis that one knows $\neg p$.

[166] There are lots of possible constraints upon R, and modal logics are individuated by which of them are imposed. Presumably TW singles out M, 4, and B because a logic that accepts these three is called S5, which the most familiar modal logic for most philosophers. The modal logic that TW is utilizing here is just called M (after the single axiom that it accepts).

11 - Assertion

11.1 – Rules of Assertion

The point of his chapter is to argue that knowledge is the norm of assertion.

There are lots of praiseworthy properties of assertions: they can be brave, truthful, honest, well timed. So why think there's just one norm?

TW: we're looking for 'the constitutive rule(s) of assertion, conceived by analogy with the rules of a game.'

Method of inquiry: 'suppos[e] that [assertion] has such rules, in order to see where the hypothesis leads and what it explains. That will be done here.'

So what's a constitutive rule?

- '...a rule will count as constitutive of an act only if it is essential to that act: necessarily, the rule governs every performance of the act.'
- but, we're not actually looking for 'non-circular necessary and sufficient conditions' for asserting something
- it must be possible to violate these rules while still performing the speech act of assertion, just as it must be possible to speak English ungrammatically
- norms of assertion are not *moral* norms. If such a violation (e.g. lying) is immoral that's just because morality is made possible by the rules of the assertion game, just as cheating at Monopoly is immoral (maybe), an immorality that's made possible only by the rules of Monopoly

The goal is to find a single rule. Some options:

Truth rule: One must: assert *p* only if *p* is true

Warrant rule: One must: assert *p* only if one has warrant to assert *p*

Knowledge rule: One must: assert *p* only if one knows *p*

TW defends the knowledge rule.

11.2 – The Truth Account

Point of this section: criticize truth as the norm of assertion

First problem: lots of speech acts are better when they involve true content, so this rule isn't special for assertion.

Example: conjecturing that p, asserting that p, and swearing to p are all better when p is true, so in some sense truth is *a* norm of conjecturing, asserting, and swearing.

But: note that the seriousness of conjecturing, asserting, and swearing to p seems to depend not just on whether p is true, but depends also on how much evidence one has for p, and that the amount of evidence required in order to be subject to condemnation varies: conjecture requires less evidence than assertion, and swearing requires more evidence than assertion. Why would that be the case if the norm of each is merely *truth*?

Second problem: assertion has *some kind* of evidential norm. Suppose that truth is the basic norm of assertion, an hence that the evidential norm is derived form the truth norm. In that case that the whole point of satisfying the evidential norm is that it helps ensure that you satisfy the truth norm.

Underlying this idea is principle (1):

1. If one must (ϕ only if p is true), then one should (ϕ only if one has evidence that p is true).

(1) is very general as a principle underlying the evidential standards for beliefs relevant to action (i.e. to ϕ), assertion being merely one type of action.

How much evidence is required in a particular case? That depends on the type of action and the badness of ϕ-ing when p is false. If ϕ is the action of driving down the street and p says that there's nobody walking on the street in front of you, then the evidential standard is pretty high. If ϕ is the action of making annoying noises and p is the same, then the evidential standard is lower.

Question: can (1) '...explain the weight of evidence which we require speakers to have for their assertions in terms of the degree of badness which we attribute to making an untrue assertion[?] Is the former proportionate to the latter?' (246)

TW thinks not. Consider a lottery case: I assert 'your ticket lost' solely on very strong probabilistic evidence. In fact your ticket lost, as we'll find out in two minutes when the result is announced. You really don't care whether you won or lost, so the stakes are really low, so presumably the evidential standards are low. So this is a case with low evidential standards and very strong evidence.

Nonetheless, says TW, your assertion is defective. But that's not what you'd expect if truth is the truth is the norm of assertion an hence that violations of (1) explain the defects in assertions. So truth isn't the norm of assertion.

Alternative explanation for the defect in the assertion:

The defect lies not in violation of the norm of assertion, but of some broader Gricean rules for conversation.[167] Example: might be a general rule of conversation to avoid stating the obvious, so when I state what should be obvious (that your lotto ticket will lose) I'm implying that I have some inside info about the outcome. I don't have inside info, so in that case my statement would be misleading.

[167] Everyone know what that means?

Problem with the alternative account: I'd also be stating the obvious if I said 'Your ticket is almost certain not to have won', so that assertion should also be subject to censure, but it isn't.

Second problem: if the problem with asserting 'your ticket lost' is that it implies that I have inside information, then canceling that implication should remove the problem: I should be able to say 'your ticket lost, but I don't mean to imply that I have inside information' without censure. But I can't.

Separate defense of truth as the norm of assertion:

For each ticket in the lottery, you have the same grounds for asserting of it that it lost. But one of those ticket's didn't lose, so if you made all those assertions you'd assert something false.

Problem: how does this translate into a prohibition on asserting of one of them that it lost? You're very unlikely to assert something false in that case.

Lesson: if truth is the central norm of assertion then we should be able to explain the evidential norms of assertion in terms of the truth norm. But we can't, so truth isn't the central norm of assertion.

NB that knowledge, unlike truth, is closely connected to evidence, so it's possible that a knowledge norm of assertion won't have that problem.

11.3 – The Knowledge Account

The preceding section relied heavily on thinking about lottery propositions. Why think that the lessons generalize to propositions about other things?

> Let p be a proposition whose truth value is known to an expert but about which you have no evidence. The expert holds a lottery. There are a million tickets, of which you have one. However, she does not announce the number of the winning ticket; she merely hands each participant a slip of paper. If your ticket won, the true member of the pair $\{p, \neg p\}$ is written on your slip; if your ticket lost, the false member of the pair is written there. There is no doubt that this is the arrangement. You are not in a position to confer with other participants. Suppose that $\neg p$ turns out to be written on your slip. On your evidence, there is a probability of one in a million that your ticket won and $\neg p$ is true, and a probability of 999,999 in a million that your ticket lost and $\neg p$ is false. Thus, if you assert p, the probability on your evidence that your assertion is true is 999,999 in a million. Intuitively, however, you are not entitled to assert p outright. (249-50)

We can make the same point more simply if we accept the principle: if p is less probable than q on one's evidence, and one has warrant to assert p, then one has warrant to assert q. Take p to be a proposition with probability n. Then take q to be the claim that my lotto ticket lost, where probability of my ticket losing is greater than n. We're not in a position to assert that q (says TW), so by our principle we're not in a position to assert p.

NB: p could be any proposition, and in particular it needn't be a proposition about a lottery. But the lesson is very broad:

> Thus the argument indicates that, for almost any kind of proposition at all, very high probability on one's evidence does not imply assertibility... The obvious moral is that one is *never* warranted in asserting a proposition by its probability (short of 1) alone. (250)

BTM:

The quote above amounts to the claim that one's assertion of p is open to criticism any time one's evidence fails to entail p. Given E=K and TW's theory of evidential probability, if you know that p then p is part of your evidence, so the evidential probability of p is 1. Hence knowing that p ensures that this necessary condition of assertibility is met, though for all he's said so far there might be other necessary conditions for the assertibility of p besides p having an evidential probability of 1, and knowing that p does not guarantee that those further conditions (if there are any) are met.

It's also consistent with this quote together with E=K that some propositions that aren't known are assertable. Example: logical truths have evidential probability 1, but they might not be assertable because one might lack evidence for them.

End BTM

NB: TW's argument for knowledge as the norm of assertion does not rely on his account of evidence. But, if you start out thinking that knowledge is the norm of assertion, and you also accept that one's evidence consists of just those propositions which one is licensed to assert outright.

Further support for knowledge as the norm of assertion (KNA):

- E=K helps avoid skeptical worries associated with limiting assertions to probability 1 propositions. Worries arise because no empirical belief is certain. But our probabilities are not credences. If E=K, then anything you know has evidential probability 1, and since seeing that p is a way to know that p, seeing that p produces knowledge that p and hence ensures the assertibility of p. Just because empirical knowledge is uncertain doesn't mean that it rests on probabilistic evidence. There's also a positive point: E=K plus KNA treats lotto propositions differently from uncertain propositions that aren't based on probabilistic evidence, which comports with intuitions (he claims).
- When someone makes an assertion, they're often challenged with 'how do you know?' We don't say 'where did you read that' because there's no presupposition that they read it somewhere; we do ask about knowledge because there's a presupposition that if they assert that p then they know that p
- Variation on Moore's paradox: it's weird to say 'p and I don't know that p'. But often one would speak the truth when they say so: take p to be some complex mathematical theorem, and then assert the above conjunction for both p and $\neg p$; at least one of your assertions is true, but both are defective in some way. This comports nicely with KNA.
  - relatedly: KNA makes sense of the belief version of Moore's paradox: 'p and I don't believe that p'. Assuming KNA, if you're in a position to assert p then you know it; that entails believing it; so the truth of the first conjunct guarantees the falsity of the second conjunct
  - sometimes presuppositions are cancellable: if I say 'Tom doesn't drink' I imply that he's an alcoholic, but it's perfectly sensible to cancel that implication and say 'Tom doesn't drink, but I don't

> mean to imply that he's an alcoholic; he just doesn't like it'. Asserting that p implies knowledge of p, but that presupposition isn't cancelable, as attempts to do so are Moorean paradoxical: 'p but I don't mean to imply that I know that p'

Objection: it's also weird to say 'p and I cannot be certain that p', so on the above line of reasoning, shouldn't we say that the norm of assertion is certainty rather than knowledge?

TW: the standards of knowledge are set contextually, and in some contexts those standards don't require certainty. Illustration: 'p and by Descartes's standards I cannot be absolutely certain that p'. The 'by Descartes's standards' bit serves to shift the standards to those of a long dead Frenchman sitting by a fire in a dressing gown. TW thinks that this sounds fine.

11.4 – Objections to the Knowledge Account, and Replies

First objection: TW's account allows for beliefs that are false but justified. In such cases, isn't assertion reasonable? Why not take justified belief as the norm of assertion?

TW: it's possible that: $JB(p)\&K\neg K(p)$. Example: I'm justified in believing that I won't be run over by a bus tomorrow, but I know that I don't know that. But assertion is unwarranted in this case so $JB(p)$ isn't a sufficient condition for warranted assertibility.

Sometimes it *is* reasonable to assert p even though you don't know that p, as when your evidence makes it extremely likely that you know that p. KNA 'makes knowledge the condition for permissible assertion, not for reasonable assertion. One may reasonably do something impermissible because one reasonably but falsely believes it to be permissible. In particular, one may reasonably assert p, even though one does not know p, because it is very probable on one's evidence that one knows p. In the same circumstances, one may reasonably but impermissibly believe p without knowing p.' (256)

Second reason one might reasonably assert what one does not know: overriding features of the case. One might assert 'there goes your train' in urgent circumstances even when one does not know. But that's not a counterexample to to KNA: urgent circumstances might make it reasonable to assert that sentence ungrammatically when speaking a foreign language, but that wouldn't be a counterexample to the rule of grammar for that language.

TW: A reasonable belief that the norm of assertion is satisfied typically makes it reasonable to make the assertion. But that doesn't mean that I've satisfied the norm of assertion

BTM: NB that TW here is implicitly appealing to an analogue of Littlejohn's distinction between *personal* justification and *doxastic* justification. TW thinks that one might be blameless, or at least excused, for violating the norm of assertion when these mitigating circumstances obtain. When he talks about 'reasonableness' he's talking about whether the person is behaving reasonably, not about whether the assertion itself is meeting some normative standard.

Second objection: KNA implies that 'speakers should always be at great pains to verify a proposition before asserting it.' (258) But in lots of cases that's false: gossip, causal conversation, philosophy seminar... So KNA is false.

Response: we often violate the rules of a game, and it's not a big deal. These cases really do involve breaches of the rules of the assertion-game, but we overlook it because in such cases it doesn't really matter. Compare: when a friendly chess game we might let someone take back a dumb move; that violates the rules, but who cares.

11.5 – The BK and RBK Accounts

Consider the believes-that-one-knows norm of assertion (BKNA):

> One must: asset *p* only if one believes that one knows *p*

Claim (that TW rejects – he's just entertaining it at this point): although one is not warranted in asserting that the lotto ticket lost or that one won't be hit by a bus tomorrow, one is at least sometimes warranted in asserting false propositions.

KNA is inconsistent with that claim. Can BKNA account for this data ('data')?

Strengths of BKNA:

- like KNA, it provides grounds for explaining why it's problematic to assert '*p* and I don't know that *p*'. If I believe that I know the conjunction then (presumably) I believe that I know each conjunct, but I believe K(*p*) and I believe K(*I don't know that p*); that second conjunct implies $\neg$ K(*p*); contradiction.
- also explains why we sometimes challenge assertions by saying 'how do you know?'

Problem for the BKNA:

- my belief that I know that *p* might be irrational: even if I believe that I know that I'm Napoleon, I still shouldn't assert 'I'm Napoleon'
- intuitively there's something wrong with the asserter – I have an irrational belief – and also something wrong with the assertion. BKNA allows that there's something wrong with the asserter, bus assuming that I really do believe that I'm Napoleon, on that view the assertion itself is beyond reproach.

So, revise BKNA for the *rationally*-believes-that-one-knows norm of assertion (*R*BKNA):

One must: assert *p* only if one rationally believes that one knows *p*.

This allows us to fault both the assertion and the asserter in the 'I'm Napoleon' case

Problem for RBKNA:

- problem of conjunctive assertions: we might rationally believe each conjunct in a large conjunctive sentence and at the same time recognize that the conjunction itself is logically false. According to RBKNA we're warranted in asserting each conjunct but not warranted in asserting the conjunction. So *warrant to assert* is not closed under conjunction.
- Rational beliefs can be false, including rational beliefs about what one knows. So according to RBKNA there's nothing wrong with false assertions (assuming that the falsely asserted propositions is something that the asserter rationally believes that they know). But there is something wrong with assertions of false propositions.
- the burden of proof is on the more complex theory, and both the BKNA and RBKNA theories are more complex than the KNA. KNA is more complex than TNA (the *truth* norm of assertion), but TW has provided reasons to prefer KNA to TNA, so the burden is met.
- Possible motivation for both BKNA and RBKNA: the thing that warrants assertion should be a mental state, and both belief and rational belief are mental states. TW: knowledge is a mental state too, so no advantage here.
  - moreover, if assertions really are defective any time they're false, and if knowledge really is the most general factive mental

> state, then taking knowledge as the norm of assertion is the only way to ensure the truth of norm-satisfying assertions while at the same time taking the thing that warrants assertion to be a mental state.[168]

[168] This might be a little too strong. The norm might be some less-general factive mental state like seeing or hearing. In that case it's possible that not all known propositions would be assertable.

11.6 – Mathematical Assertions

Consider the case of assertion in mathematical contexts.

First approximations:

In mathematics: one has warrant to assert *p* iff one has a proof of *p*. Why is that?

Explanation, according to KNA: that's because in mathematics, one knows *p* iff one has a proof of *p*

There are exceptions to the proof norm of assertion in math, but in those cases, whether or not one has warrant to assert *p* tracks whether one knows that *p*.

Suppose that I have warrant to assert mathematical proposition *p* but I lack a proof. Possible reasons:

- I know by testimony that a poof exists
  - in that case I don't have a proof, but I do know that *p*, so assertibility tracks knowledge rather than proof possession
- I have non-deductive evidence for *p* that doesn't amount to a proof
  - again, in that case I know, so assertibility tracks knowledge rather than proof possession

In both cases, warrant to assert tracks knowledge, not proof-possession. So these counterexamples to the proof-having norm of assertion in math are not counterexamples to the more general knowledge norm of assertion.

What's not permitted by KNA is the assertibility of *p* when I mistakenly believe that I have a proof of *p*: in that case I don't know that *p*, so I'm not warranted in asserting that *p*

BTM:

TW uses the example of of mistakenly believing that *p* because your mathematician friends have asserted that there's a proof of *p* as part of a practical joke. In defense, he makes a claim about the epistemology of testimony:

> Testimony is a special source of warrant because one speaker can pass on a warrant to another. Since the expert mathematicians have no warrant to assert *p* themselves, they have none to pass on to you.

It's contentious that testimony merely 'passes on' warrant from person to person. On other views, testimony can generate knowledge, even if the testifier doesn't know. If these generative accounts of testimony are correct then I *do* know that *p*, so by KNA I'm warranted in asserting that *p*. TW judges that I'm not so-warranted, and he accepts KNA, so he's forced to accept the transmissive picture of testimony. It sounds like TW is fine with that, but it's worth noting that accepting TW's account of assertion and his intuitions about what's assertible in particular cases forces us to accept particular views elsewhere in epistemology.

End BTM

Sometimes I have a proof but I lack warrant:

I have a proof of *p*, but it's really long and complicated. My math friends tell me (falsely) that my proof contains an error, or perhaps I just remember all the times I've mistakenly believed myself to have a proof. Intuitively, I'm not warranted in asserting *p*. So I have proof but not warrant.

Do I know that *p* in those cases? TW: no, your doubts/ the false testimony undermines your knowledge, so according to KNA you lack warrant to assert, which matches the intuition.

Mathematical assertions are not so untypical as they first appear. Like empirical assertions, they are defeasible.

NB: *proofs* are not defeasible: they're deductive, and so they're monotonic. But the *warrant* provided by having a proof is defeasible, as above in the case where someone tells you that your proof is fallacious, or where you have self-doubts.

In this sense at least the warrant to assert provided by a proof is defeasible just like the warrant to assert provided by experience is defeasible. As TW sees things, both proof-possession of *p* and an experience of *p* (of the right sort) are factive, so both entail that *p* is in fact true. Hence the two ways of having a warrant to assert that *p* — having a proof and experiencing — are both factive and defeasible.

11.7 – The Point of Assertion

What's the point of assertion? Why do we perform the speech-act of assertion?

> To make an assertion is to confer a responsibility (on oneself) for the truth of its content; to satisfy the rule of assertion, by having the requisite knowledge, is to discharge that responsibility, by epistemically ensuring the truth of the content. (268-69)

If the norm of assertion were truth rather than knowledge, then you could discharge your responsibility accidentally: you'd discharge it anytime you asserted something that happened to be true, even if that's just by coincidence. A knowledge norm rules that out, since you can't know things on accident.

1 - A State of Mind