Intuitions and Meanings

Brian Weatherson, Cornell University, September 26

Incomplete version - references etc missing

One common response to the experiments by Weinberg, Stich and Nichols on diversity of epistemic intuitions is that the subjects who do not agree with philosophical orthodoxy about Gettier cases have a different concept of knowledge. This line of response links up with an attractive defence of reliance on intuitions about possible cases in philosophy. Intuitions about possible cases, the defence goes, partially determine the intension of the concepts being considered. I will argue that the response fails, and that the experimental results are not good reason to think that the subjects have different concepts. This means the general defence of intuitions also is questionable. I also argue that a related defence of the reliability of our intuitive judgments discussed by Timothy Williamson fails. But Williamson's argument we don't actually use intuitions *as such* in philosophy as much as some philosophers (including my earlier temporal parts) have supposed is more successful. This line of argument does successfully rehabilitate some uses of appeals to possible cases in philosophy, though I'll argue at the end that it still doesn't amount to a conclusive defence of the use of Gettier cases because of peculiar facts about those cases and their relation to norms of belief, assertion and practical action.

1. Clashing Intuitions and Clashing Concepts

Even before the Weinberg, Stich and Nichols (hereafter WNS) results were published, the idea that agreement on the possible cases used to defend various epistemological theories, especially Gettier cases, was necessary for sharing the concept KNOWLEDGE had been endorsed in print. Here, for example, is Frank Jackson's view of the situation.

> I have occasionally run across people who resolutely resist the Gettier cases. Sometimes it has seemed right to accuse them of confusion… but sometimes it is clear that they are not confused; what we then learn from the stand-off is simply that they use the word 'knowledge' to cover different cases from most of us. In these cases it is, it seems to me, misguided to accuse them of error (unless they go on to say that their concept of knowledge is ours). (Jackson, 1998, 32)

Since those results came out, this view has been more prominently endorsed. For example, Ernest Sosa suggests that many of the apparent disputes, not only between cultures but between philosophers about these cases are 'spurious'. And William Lycan makes a similar conjecture.

> [S]uppose that the survey results are impeccably produced and robustly replicated: 60% of an Asian ethnic group and 25% of European-descended American undergraduates firmly reject Gettier and insist, clearheadedly and understanding the terms and the issue, that a Gettier 'victim' does know. In that eventuality, I submit, we have a conceptual difference. In the speech of the 60% and the 25%, 'know' really does mean justified true belief, period. We would have to regard that speech as a dialect that differs from our own. It would be interesting to go on to ask those subjects whether they see any important difference between the two kinds of 'knowers,' ordinary ones and Gettier victims. Perhaps they would stigmatize the Gettier victims in some way for which there is no simple convenient expression. Or, less likely, they would see no important difference, and simply have no stronger conception of successful cognition. (Lycan, forthcoming)

I don't quite know why Lycan thinks speakers of this dialect have "no stronger conception of successful cognition". Such speakers are fallibilists about knowledge, so they have a conception of *infallible* knowledge, which is a stronger conception. But I think it is worthwhile to continue the thought experiment he suggests and asks for how these people treat differently ordinary 'knowers' and Gettier 'victims'. For there is a lot of indirect evidence that *this* is what is relevant to whether they have the same concept, or they have different opinions about how the shared concept is applied.

For most philosophically interesting concepts, the test we use for whether two subjects share the concept is not whether they apply the concept in the same cases, but whether they are prepared to make the same relatively simple relatively general inferences from propositions containing the concept(s).[1] Here are three cases that support that general consideration.

First we'll consider CAUSE. Aaron and Barry are having a philosophical disagreement about causation. Aaron thinks that causation is transitive, but Barry does not, and currently Barry is having fun with trying to come up with more and more ridiculous cases where transitivity yields surprising results. He has just convinced Aaron that his parents' actions in conceiving him caused his death when Caleb arrives. Caleb willingly agrees with Aaron about all of Barry's cases. For a while Aaron thinks that Caleb not only shares his concept CAUSE, but shares his theory of causation. Then this conversation ensues.

---

[1] By sharing a concept here I don't mean to make any metaphysically deep claim about there being things *concepts* that literally reside in many different people's heads. What I do mean depends on just what kinds of things propositions are. If propositions are structured entities then people share a concept if they bear intentional relations (belief, desire, fear, etc) to propositions containing that concept. If propositions are unstructured things are a little more delicate to say *exactly* what sharing a concept comes to. But for our purposes we can say that they share a concept if the meaning of a word w_1 in the mouth of the first speaker is the same as the meaning of a word w_2 in the mouth of the second. It isn't necessary for present purposes to give a more general definition that covers, for instance, thinkers who do not speak.

Aaron: Could you turn out the light?

Caleb: That's a good idea. Let me go get a ladder.

Aaron: Why are you getting a ladder.

Caleb: To remove the light.

Aaron: Why not just flick the light switch?

Caleb: Why do that?

Aaron: Because flicking the light switch causes the light to turn off.

Caleb: I agree that flicking the light switch causes the light to turn off, and I am trying to turn the light off, but I don't see that's a reason to flick the light switch.

Aaron: Do you mean you have other reasons to not flick the light switch, a taboo or something?

Caleb: No, I just don't see why you think I have any reason to flick it.

Assuming Caleb is being sincere here, he *doesn't* have the same concept CAUSE as Aaron and Barry. For it is essential to possessing the concept CAUSE that Aaron, Barry and the rest of us share that you recognise that if the pair <X, Y> falls under that concept, and you want to accomplish Y, then you have a defeasible reason to do X. Just what concept Caleb possesses is unclear - perhaps the concept we would express by 'stands in the ancestral of the counterfactual dependence relation to' or something similar.

What matters here is that while Aaron has a reason to think he and Caleb do not share a concept here, he has no such reason to think he and Barry fail to share a concept. He and Barry are having a philosophical dispute in which, they both agree, one of them is right and the other wrong. They disagree about who it is, but they agree that one of them is wrong, and so should we.

The same pattern arises when we consider differences about the concept of belief. Consider three possible reactions to a small variant on the famous false belief experiments reported by Baron-Cohen, Leslie and Frith (1985). In their experiments they used the following question to test where the children thought the doll believed the marble was stored: "Where will Sally look for the marble?". This was even called the *Belief Question* in Baren-Cohen et al's paper. We can imagine breaking this question into two parts though, and seeing how the children respond. Three kinds of answers are interesting to us, and we'll imagine they are given by Aimee, Bianca and Cindy.

> Q: Where does Sally believe the marble is?
>
> Aimee: In the basket.
>
> Q: Where will Sally look for the marble?
>
> Aimee: In the basket.

> Q: Where does Sally believe the marble is?
>
> Bianca: In the box.
>
> Q: Where will Sally look for the marble?
>
> Bianca: In the box.

> Q: Where does Sally believe the marble is?
>
> Cindy: In the basket.
>
> Q: Where will Sally look for the marble?
>
> Cindy: In the box.

Here Aimee gives the answers we'd expect a normal child to give, Bianca gives the answers characteristic of autistic children, and Cindy's answers are just odd. If Baron-Cohen *et al* are right about labelling their question the *Belief Question*, then Bianca does share our concept of belief though. Crucially, she applies the concept BELIEVES in the right way when trying to infer where Sally will look for the marble. Although Cindy sounds like she agrees with us, and with Aimee, about what Sally believes, in fact she has a different concept of belief, so her first answer doesn't express agreement with us, or for that matter disagreement with Bianca. If we can infer anything about what Cindy thinks about Sally's beliefs at all, it's that she *agrees* with Bianca that Sally believes the marble is in the box. That's the conclusion we would get by using Baron-Cohen *et al*'s method for determining her belief, because it's how she answers the belief question. I think it's safer to say that she has a different concept, while Bianca shares our concept but misapplies it.

Our final example is from a famous passage by R. M. Hare.

> Let us suppose that a missionary, armed with a grammar book, lands on a cannibal island. The vocabulary of his grammar book gives him the equivalent, in the cannibals' language, of the English word 'good'. Let us suppose that, by a queer coincidence, the word is 'good'. And let us suppose, also, that it is really is the equivalent—that it is, as the *Oxford English Dictionary* puts it, 'the most general adjective of commendation' in their language. If the missionary has mastered his vocabulary, he can, *so long as he uses the word evaluatively and not descriptively*, communicate with them about morals quite

> happily. They know that when he uses the word he is commending the person or object that he applies it to. The only thing they will find odd is that he applies it to such unexpected people, people who are meek and gentle and do not collect large quantities of scalps; whereas they themselves are accustomed to commend people who are bold and burly and collect more scalps than the average. But they and the missionary are under no misapprehension about the meaning, in the evaluative sense, of the word 'good'; it is the word one uses for commending. If they were under such a misapprehension, moral communication between them would be impossible. (Hare 1951: 148)

Hare is mostly right here. The missionary and the cannibals can communicate, and that is evidence that they share a concept. And this is despite the fact that they would apply the concept very differently to particular cases. But Hare is wrong to immediately infer that the concept isn't descriptive, for we can note some things that aren't *entirely* evaluative that the missionaries and the cannibals agree about. They agree, for example, that if X is good, then anyone has a reason to do X. They agree that whether doing X is good is related in some way or other to duties to others. The cannibals seem to think it is duties to the rest of the tribe, as far as I can tell the missionary thinks it is duties to God, and others may think it is duties to all humans, or all sentient beings. Whether that means they agree about descriptive facts depends on just what the correct metaphysics of morality is, and on just what it means for a fact to be 'descriptive'. But that isn't relevant to our interests. What is relevant is that whether they share a concept depends on what inferences from claims involving that concepts to claims about *reasons*, and more generally whether they accept similar relatively general, relatively simple inferences whose premises involve the concept.

I suggest we take these three cases as indicative of a broader pattern. What it takes to share a concept is not to agree about particular cases, but to agree about what claims about reasons are made true by claims involving the concept, and more generally to agree about which general, simple inferences can be properly made from the concept. In the case of knowledge, the general inferences might involve the following kinds of principles. Knowledge is a norm governing belief in that beliefs that do not constitute knowledge can be properly criticised. Knowledge is a norm governing practical reasoning in that agents who use premises in practical reasoning that they do not know can be properly criticised. And knowledge is a norm governing assertion in that asserting what you don't know opens you to proper criticism. If someone agrees about those claims, or something akin to them for the precise statements of those claims is a matter of philosophical dispute, they share our concept of knowledge. By this I don't mean that if they agree about which believers, practical reasoners or asserters can be properly criticised they share our concept. I mean that if they agree about enough of the inferential connections, they share the concept.

This might look like a kind of inferentialism about concepts, and you might think that inferentialism about concepts suffers from some serious problems. Surprisingly enough, even though both conjuncts there are true, this view doesn't suffer from serious problems. In particular, what I've said is much too weak to be vulnerable to either of the major problems facing inferentialist accounts of concepts. I don't say that concepts are *entirely* individuated by the inferences they licence. All I say is that it is *necessary* to possess some concepts that the thinker accept certain inferences from those concepts. I don't say that there could be enough such inferences that we get all the way to necessary *and sufficient* conditions for applications of the concept. So we can still have conceptual disagreement, and we can explain why it is so hard to find necessary and sufficient conditions for the application of a concept. Further, the inferences in question are so weak that it is not possible for people to share the concepts and disagree about their application. Stronger inferential approaches to concepts have to say either (a) that genuine conceptual disagreement is very rare because when people accept different inferences that shows they have different concepts, or (b) that although disagreement is possible, there are still enough inferences to precisely individuate a concept and disagreement about them is *not* possible. Neither option looks attractive, so sticking to the weaker line I'm adopting seems sensible.

So far I've argued on fairly general grounds that the kind of disagreement we see in Gettier cases, and in thought experiments more generally, is not the kind of disagreement that should lead us to think that the parties are deploying different concepts. There is also a more direct argument to the same conclusion. It turns out that Lycan himself accepts that some cases sometimes described as Gettier cases are cases of knowledge. We can argue directly that Lycan deploys the same concept both as regular sceptical philosophers and their anti-sceptical brethren who accept the JTB analysis.

> I am more sympathetic to Hetherington's view than most will be. He very usefully distinguishes between 'helpful' Gettier cases and 'dangerous' ones: A helpful case is one in which the Gettierish 'strange occurrence' or fluke saves JTB itself, as when Havit owns a Ford even though Nogot does not. A dangerous case is one in which the 'strange occurrence' prevents knowledge despite existing normal JTB, as in Harman's unpossessed-defeater examples and the Ginet-Goldman barn case.
>
> As I declared in section iii, I reject the majority view that the victims in unpossessed-defeater cases and the barn case lack knowledge. And now Hetherington has shown that those examples have something distinctive in common, viz., being 'dangerous' as opposed to 'helpful.' Moreover, I think his interpretation of them is pretty much right: that although their protagonists' knowledge is failable and some luck is involved in a peripheral way, it is knowledge nonetheless. True, Jill and Henry nearly failed to know; it

> does not follow that they fail to know. With Hetherington, I maintain that they do know. (Lycan forthcoming)

Lycan takes himself, correctly, to be here disagreeing with (as opposed to talking past) philosophers who think the Harman and Ginet-Goldman cases are not cases of knowledge. So thinking there are some 'dangerous' cases is consistent with deploying the same concept of knowledge as the orthodox philosopher who thinks that none of these amount to knowledge.[2] Now it could hardly be evidence that we have a conceptual difference, as opposed to a normative disagreement, if two philosophers disagree on where the line between dangerous and helpful Gettier cases are.

To see how this line may be blurry, consider the case of someone who receives testimonial evidence from someone in a helpful Gettier case. More concretely, assume that Ezra is a usually reliable and cautious testifier, who has a justified true belief that Havit owns a Ford. Ezra tells Andrew that Havit owns a Ford. Now it turns out that Ezra's true belief was grounded in a false belief about which particular Ford Havit owns. So this is a helpful Gettier case. But if Andrew comes to believe Ezra, is he in a helpful or dangerous case? It seems implausible to me that disagreement on *this* point would be grounds for saying two philosophers deployed different concepts of knowledge. Now consider the philosopher who thinks that Andrew does know, because his case is more like Harman's unpossessed defeater case than like 'helpful' Gettier case. That philosopher could reasonably think that if Andrew now told Ezra that Havit owns a Ford, Ezra would have sufficient grounds (testimony from a knower) to know that Havit owns a Ford. And she could think that Andrew's testimony here, just repeating what Ezra told him, could hardly change Ezra's epistemic position, so Ezra knew all along that Havit owns a Ford.

The point of that paragraph is not primarily to argue that knowledge really is justified true belief. All I want to defend *here* is that a philosopher who made that series of inferences to go from the helpful/dangerous distinction to the JTB theory would not thereby have adopted a different concept of knowledge. She would just have changed her mind about whether Ezra knew that Havit owns a Ford. So whatever the merits of my general argument that disagreement about cases is usually insufficient grounds for differences about which concepts we possess, there are particular reasons for thinking that one can easily hold the JTB theory without thereby lacking the concept of knowledge most philosophers deploy.

---

[2] My use of 'orthodox' here is a little tendentious since I really don't know what the majority opinion is on either the Harman or the Ginet-Goldman examples. I suspect many philosophers these days think the fake barn example is a case of knowledge since it is a case of perception. But I haven't done any surveys to confirm this suspicion.

2. Justifying Intuitions via Meaning

The debate in the first section is important because it links to a possible defence of the use of intuitions in philosophy. Why should we trust the intuitive judgements that often seem to form the foundation of philosophical arguments? Ideally, an answer to that question will show why intuitions are guaranteed to be at least probably true. If intuitions are constitutively tied to meaning, we should have such a guarantee. That is, if it is true in virtue of how the content of, say, 'know' is defined that most of our intuitions concerning the truth of sentences containing 'know' are correct, then we have a pretty good reason to rely on intuitions in philosophy.

Williamson considers a similar proposal and pronounces it a version of 'silly idealism'.

> If we merely seek the best explanation of our having the intuitions, without any presumption in favour of their truth, we may find a psychological theory to explain them, but how are we to answer the questions about a mainly non-psychological universe that grip many metaphysicians and other philosophers? Perhaps intuitions about thought and language have a special epistemic status, because they help to constitute their own subject matter; but to generalize that claim to all intuitions in philosophy is to fall into a silly idealism. The nature of identity over time, for example, is not a matter of thought or language: the question is how things persist, not how we think or say that they persist. (Williamson forthcoming a: 15)

Let's grant, what may be controversial, that it would be a silly idealism to hold that when we said the Treaty of Versailles was a cause of World War II, we are somehow talking about our own intuitions because facts about causation are constituted in part by our own intuitions.[3] Still, it is hard to deny that whether the sentence labelled (1) is true depends on the meanings of the words in that sentence.

(1) The Treaty of Versailles was a cause of World War II.

And it is hard to deny that meanings are determined in part by shared dispositions to use terms. Since it is plausible that philosophical intuitions are grounded in the same mental states that ground those usage dispositions, the intuitions are a guide to the truth of the sentence via being a guide to its meaning. This is

[3] I've changed the example from persistence to causation because I think there are special features of the persistence debate that complicate the issue of just what the connection is between intuitions and the truth of the claims intuited. If, like me, you believe in perdurantism and mereological universalism, many of the interesting facts about persistence turn out to be facts about which fusions of temporal parts most interest us, so they are partially constituted by us. Since our interests play a smaller, and more easily identifiable, role in other debates, I'll use other examples.

entirely consistent with saying the proposition (1) expresses is entirely about facts external to our thought, and hence denying the idealist claim that the proposition is somehow about us.

This much should all be obvious, but there are two other points worth making.

First, this argument generalises to other kinds of facts that play a role in determining meaning. For instance, if David Lewis (1983) is right that facts about metaphysical naturalness play a role in determining the meanings of words, we can appeal to facts about metaphysical naturalness in a systematic investigation of whether the sentence (1) is true. And we can do that even though the proposition that sentence expresses is no more about metaphysical naturalness than it is about our intuitions. This was one of the main points I was making in Weatherson (2003).

Second, the argument is only as good as the claim that usage dispositions are constitutive of meaning. That's obviously true *to some extent*, but as argued in the previous section, dispositions to apply terms (or not) in philosophically interesting cases play at best a peripheral role in determining meanings. (I talked there mostly about the content of mental concepts, but the arguments carry across in an obvious way to the meanings of public expressions.) So to the extent that this is our justification for relying on intuitions, we should put less weight on intuitions about possible cases and more weight on intuitions concerning general principles.

Williamson of course acknowledges the point that intuitions could be connected to meaning without any kind of idealism being true. Indeed he offers a suggestion about meanings that offers the prospect of justifying a much wider appeal to intuitions than I have hitherto allowed. For reasons we will discuss at length in the next section, this isn't Williamson's considered defence of what we commonly call intuitions, but it is of some independent interest.

He proposes that we modify Davidson's principle of charity by replacing true belief with knowledge. The right assignment of content to a person's mental states is that which maximises the amount that they *know*. (Williamson only illustrates this with pronouns, but it seems that the proposal is meant to generalise.) If this is correct then *all* intuitions, and not just intuitions about general principles, have an important role in determining content, and we have our justification for using intuitions back. Unfortunately, there are three kinds of problem with this theory.

First, like many theories of content, it faces a problem with disjunctions. For most of us, whether our mental word 'horse' means *horse* or means *horse or horsey cow* makes no difference to how many of our beliefs constitute knowledge. Williamson could solve this problem by using Lewisian naturalness as a tie-

breaker in cases where two different assignments do equally well at maximising knowledge. So this is at worst a recoverable problem.

Second, Williamson's view assigns the wrong content to natural kind terms. At some stage (say the 14th century in poorly astronomically informed pockets of England) there were communities that used sentences like "Venus is a star" to express their false belief that Venus was a star. Now it would maximise their knowledge if we took their mental and public word 'star' to mean *heavenly body*, or perhaps *heavenly body visible in the night sky*, or perhaps *heavenly body visible in the night sky and not identical with the moon*. But those would be the wrong assignments, because they falsely imply that these people expressed a true belief when they said "Venus is a star".

Third, Williamson's view gets the content of normative terms badly wrong. Assume we get to sit down with a bunch of apparently English speaking Muslim terrorists, and try to do radical translation on their dialect. They say things like "Killing Westerners is good", "Sex outside marriage is bad", "Theocracy is good", "Drinking alcohol is bad" and so on. It's going to be rather unlikely that we'll maximise their knowledge by interpreting their word 'good' as meaning *good* and bad as meaning *bad*. We'll do much better by the standards of maximising knowledge if we interpret their word 'good' as meaning *approved of by Osama bin Laden*, or *approved of by them*, or *approved of by such-and-such moral code*. Yet none of these are the correct interpretations. The correct interpretation is that by 'good' they mean *good*, and they have an inordinate number of false moral beliefs.

This problem is particularly relevant to the debate at hand. Many times in philosophical analysis we are interested in working out the content of normative concepts. If intuitions about the application of normative concepts to particular cases are not, or at least not centrally, constitutive of the content of those concepts, this line of justification for the use of intuitions won't be broadly relevant. So we should look elsewhere.

3. What are the roles of Intuitions in Philosophy?

Intuitions have a wide role in philosophy, but they are usually judgments about possible cases. These possible cases are just short stories. The stories philosophers use usually have little literary merit, but that is inessential. We could do some philosophy taking most of our examples from, say, *Hamlet*. Let's look at some of the philosophical theories we can refute with cases from *Hamlet*. (These will generally be bad philosophical theories, hence we can easily refute them.)

Hypothesis: If X causes Y, then X and Y overlap temporally.
Refutation: Polonius's murder and Ophelia's suicide do not overlap temporally, but Polonius's murder causes Ophelia's suicide.

Hypothesis: Killing your father's killer is morally justified, and whenever X believes they are doing something that is in fact morally justified, what they do is morally justified.
Refutation: When Hamlet killed Polonius, he thought he was killing his father's killer, but his killing of Polonius was not a morally justified act.

It's not clear which conjunct this refutes, but it refutes their conjunction. Our last case is one we'll return to a little in what follows.

Hypothesis: If the causal origin of X's belief that *p* is being told that *p* by someone who has a vested interest in X coming to believe that *p*, then X does not know that *p*.
Refutation: The causal origin of Laertes's belief that Hamlet killed his father is Claudius's testimony, and Claudius has a vested interest in Laertes coming to believe that Hamlet killed his father, yet after this conversation (at the end of act IV) Laertes knows that Hamlet killed his father.

What exactly is the logical form of the refutation? It looks like it takes this form:

> *Demonstrative*
> In the possible situation S, this F is not-G
> So, not necessarily all Fs are Gs.

If that's the logical form of the underlying argument, it is hard to see just how this is connected to debates about intuitions. For there is no talk of intuitions in this argument, just facts about Laertes's epistemic state at the end of act IV. Why think that intuitions are relevant to the refutation of this little epistemic theory? One possible reason is the thought that the only way a philosopher could *know* the premise to be true is by using her intuitive capacities. This isn't that plausible phenomenologically. When watching *Hamlet* we don't have to drag out anything that feels like intuition to come to believe, and even to know, that Laertes knows who killed his father. But it's not generally true that we have excellent access to how our minds work, so let's bracket that worry for now. If the argument really rests on an intuition about knowledge, then the argument should look something like this.

Non-Demonstrative

In the possible situation S, intuitively this F is not-G

So, not necessarily all Fs are Gs.

The names for the forms should be obvious. The first argument is deductively valid, the second argument is not. If we want to defend an argument of the second form, what we need to defend is an inference pattern, if we want to defend an argument of the first form we just have to defend a proposition about, in this case, Laertes's knowledge. That suggests there is something importantly different about the two forms. We might have a general argument that arguments like *Non-Demonstrative* are not only always invalid, they always provide poor support for their conclusion. Perhaps the WNS experiments could ground such an argument, or perhaps general concerns about the reliability of our intuitions, as suggested by Cummins (1997). But these don't tell, at least in the first instance, against arguments like *Demonstrative*. To reject that argument you'd have to give a reason that Laertes's *didn't* know that Hamlet killed his father, or at least a reason we shouldn't believe that he did.[4]

Now one could stipulate that whatever we know about possible stories is to be called intuitive knowledge, and hence that in both *Demonstrative* and *Non-Demonstrative* we are using arguments from intuition. But that seems like an unfortunate choice of terminology. For one thing, it seems like an interesting theoretical question whether intuitions are unreliable, as Cummins suggests. But there's obviously no interesting question about whether our knowledge of stories is unreliable, let alone whether the facts of the story are unreliable. For another, intuitions should be things that are in our heads, and facts about what Laertes knows are in Laertes's head if they are anywhere at all, not in ours. So it is best to restrict the phrase "arguments from intuition" to arguments like *Non-Demonstrative*.

In his (forthcoming), Williamson chides some philosophers[5] for assuming in effect that all arguments from possible cases are of the form *Non-Demonstrative*. He thinks philosophers make this assumption for two reasons. First, they want their arguments to ultimately rest on luminous conditions, and facts about what is really true in the story are not luminous. But as he points out, facts about intuitions are not always luminous either. Second, they want their arguments to rest on premises that are agreed on by all sides in the debate. But this is a bad motivation if taken generally, because sometimes the only good arguments for a true conclusion (anti-scepticism about the external world, the falsity of all contradictions) rest on premises that are not common ground.

---

[4] One might think that the latter disjunct could be defended by just the same kind of general criticisms of intuition. But this move is risky for two reasons. First, it shows at best we don't know *Demonstrative* to be sound, not that it is unsound. Second, it relies on the unproven claim that intuition really is (part of) how we know that Laertes knows who killed his father.

[5] He writes of plural philosophers here, though the only one named is me.

We should take Williamson's point that not all arguments from possible cases are like *Non-Demonstrative*. And we should agree that requiring luminous premises or premises that are common ground is sometimes excessive. But it's not clear what follows from these. For it is *better*, ceteris paribus, to use premises that we can (in the circumstances) know to be true when they are true. (It might be true that nothing is always luminous, but there are significant differences in how nearby the counterexamples are to different luminosity theses, so this condition is not trivial.) And it is better, again ceteris paribus, to use premises that are common ground in a debate. If our opponents won't allow us enough premises to get to true conclusions we should give up this desiderata, but it's a useful starting point. So there is some reason to interpret a philosopher putting forward an argument from possible cases as using *Non-Demonstrative* rather than *Demonstrative*. And hence there is some reason to think that in these cases our evidence does bottom out in mental facts.

So Williamson's arguments do not provide conclusive reasons for thinking that philosophers are better off basing their arguments on *facts* about possible cases rather than *intuitions* about possible cases. I think we'll make better progress on this question if we get some more examples in front of us, and in particular look at whether the arguments being offered are best thought of as being of the form *Demonstrative* or of the form *Non-Demonstrative*. The next three examples are, in order, an example of a completely successful argument from a possible case, an argument from a possible case that fails despite being plausible, and one that fails and is implausible. (Sadly we have to leave *Hamlet* for a while, but we'll come back to it.)

Nozick and Kripke
Hypothesis: Nozick's (1981) sensitivity-based theory of knowledge.
Refutation (from Kripke's unpublished lectures on Nozick): Rosencrantz is in a county where there are many many fake green barns, and a few real red barns. He looks at a red barn and forms the belief that it is a red barn. By Nozick's lights, he knows it is a red barn but not that it is a barn. But this conjunction is false.

Bayes and Ellsberg
Hypothesis: The Bayesian theory of rational choice under uncertainty.
Refutation: According to the Bayesian theory, our preferences should satisfy the condition Patrick Maher (1993) calls *Independence*. But in the choice situation described by Ellsberg (1961) one of the rationally permissible choices violates Independence. So the Bayesian theory is false.

Bayes and Roulette

Hypothesis: The Bayesian theory of rational choice under uncertainty.

Refutation: According to the Bayesian theory, our preferences should be transitive. But when playing roulette at a table known to be fair, it is rationally permissible to prefer betting on red to betting on black, and betting on black to betting on 00, and betting on 00 to betting on red.

I think it is widely accepted that Kripke's criticism of Nozick is successful. And I think no one would think that the Roulette example generates difficulties for Bayesian theories of rational choice. (The standard complaint about the Bayesian approach is that it illegitimately analogises every decision problem to the problems facing a roulette player, not that it gets roulette wrong.) The middle case is controversial, but I'll assume for the sake of this paper that Ellsberg's case is a very plausible, but ultimately unsuccessful objection to the Bayesian theory of rational choice, and in particular to Independence. It is impossible to find an uncontroversial example of a case that is a powerful but ultimately unsuccessful challenge to a philosophical theory, so I don't expect everyone to agree with this claim about the Ellsberg case, but this is as good an example of this kind of case as we can expect to find.

Each of the three refutations contains a tacit argument, and in each case we could interpret it either as being like *Demonstrative* or being like *Non-Demonstrative*. I'll argue that neither general interpretative strategy is perfect.

Let's assume we interpret them all as being of the form *Demonstrative*. Then the argument in *Nozick and Kripke* is a valid argument with true, and indeed knowably true premises. As we should, we interpret it as a successful response to Nozick. The problem comes when we look at the anti-Bayesian cases. In each case we interpret them as valid but unsound arguments whose conclusions are that the Bayesian theory of rational choice is incorrect. In each case the argument is unsuccessful, and in each case the problem is the same: the argument contains a false premise. Now it is good that on this interpretation the two arguments are regarded as unsuccessful. But something seems to be missing in this analysis. Ellsberg's objection is historically, and philosophically, very important, and it calls for detailed response by those who want to defend the Bayesian theory. The roulette objection is a philosophical clanger, and calls for simple dismissal. If the two arguments fail for the same reason, we need to do some work to explain why this is so.

So let's assume instead that we interpret each of them as being of the form *Non-Demonstrative*. Now we can easily say what the difference is between *Bayes and Roulette* and *Bayes and Ellsberg*. Both arguments are deductively invalid, but they at least resemble sound inductive inferences. However the roulette argument we know could not be sound because it has a false premise, whereas the argument

Ellsberg presents has true premises. This seems like an important distinction between the cases, so we get the result we were looking for that the two arguments are not on a par. The downside is that we now seem to have too strong an analogy between Ellsberg's objection to Bayesian decision theory and Kripke's objection to Nozick. Both of them deploy the same argument form, and both of them deploy arguments with clearly true premises. The only issue then is whether they are successful or unsuccessful instances of that inductive schema. Now some people might think that kind of analogy is entirely appropriate. (That was what I assumed in my earlier paper.) The idea is that both parties are putting forward fallible arguments, and although the form they are doing is usually truth-preserving, there is some hard work to do in each case in determining whether we are in one of the cases where it fails to be truth-preserving. But as Williamson argues, it isn't altogether obvious why we even need to do that work to evaluate Kripke's argument, for it is clearly successful, and interpreting it as being of *Demonstrative* form removes questions about the reliability of the inference pattern in *Non-Demonstrative*. Moving to a non-demonstrative interpretation of the arguments has given us a genuine disanalogy between the two anti-Bayesian arguments, but has created a possible spurious analogy between Ellsberg's argument and Kripke's.

There is a relatively simple way out of this. We could interpret the philosophers who put forward these cases as presenting *both* kinds of argument, arguing 'in the alternative' as lawyers like to put it. Interpreted this way, we get the right disanalogies between the cases. Kripke is presenting a simple sound argument, and a more complicated non-demonstrative argument as a reserve. Ellsberg is presenting one argument that fails because it has a false premise, and one argument where the premises are true and support the conclusion, but which happens to be a counter-instance to a generally reliable schema. The existence of the second argument is why his case is more philosophically important than the person touting the roulette argument, who simply puts forward two arguments with false premises.

To a large extent, this way of looking at the cases agrees with Williamson. It is not the case that *all* our fundamental evidence consists of mental states such as intuitions. Some times we do just put forward arguments that rest on facts about possible cases, and in some cases, like Kripke's, those arguments are successful. But there is still an important place for intuition in our practices, because in some cases where it is controversial or unclear just what the facts are it might be less controversial or more clear what the intuitions are, so we can make more progress if we harness both facts and intuitions. So the question of what can justify our intuitions is still to some extent an important one.

Still, this tour through various cases has revealed some facts about just how urgent that question is, or isn't. We know from simple reflection on *Hamlet* that people know lots of things about what is true in stories, so facts about stories are available as evidence for philosophical argumentation. Hence we need a

special reason for thinking facts about the stories philosophers tell are not similarly available[6]. It might be thought that what is special about philosophers' stories, as opposed to the stories I took from *Hamlet*, is that they concern live philosophical questions. To some extent I sympathise with this objection. It is plausible that for many cases alleged to refute interesting philosophical theories, facts about the relevant to the evaluation of the theory will often be within the 'margin of error' of our capacity for spontaneous evaluation. But this need not always be the case. It certainly isn't the case that *all* facts about stories which involve philosophical concepts are too close to call, as the three examples from Hamlet above show. It need not always be the case that examples relevant to live theories are too close to call, as this example shows.

Hypothesis: If X's evidence that *p* is from a source that (a) X has not had access to before and (b) X has no reason to believe is reliable, then X does not know that *p*.

This hypothesis is the kind of thing that an internalist epistemologist could easily come to believe on reflection on Keith Lehrer's TrueTemp case. (I haven't credited it to any particular internalist because I have deliberately left off all qualifications and subtleties in order to make the methodological point. I suspect the example should be troubling even once we add those back in, but arguing for that would take us too far afield.) But *Hamlet* shows it is false.

Refutation: At the end of Act I, the only evidence Hamlet has that Claudius killed his father is the testimony of his father's ghost. Hamlet has never received evidence from ghosts before, and he has no reason to think that ghostly visions are reliable sources of evidence. But still he knows that Claudius killed his father.

As you may expect, the example is a little complicated. (It would be surprising if we had *clear* refutations of prominent philosophical theories from the most famous English language play of all time.) Hamlet's toying with Cartesian doubts at the end of Act II ("The spirit that I have seen May be the devil: and the devil hath power To assume a pleasing shape") may be evidence that he lacks the requisite subjective certainty for knowledge. I doubt this, I think Hamlet is just trying to justify to himself his delays in revenging his father. But even if you accept that Hamlet doesn't *know* Claudius killed his father, he is in a position to know it, and he has good evidence that it is true, and those facts refute slightly stronger (but really equally plausible) internalist hypotheses as the ones we mentioned. And the issue of whether at the end of Act I Hamlet knows (or is in a position to know, or has good evidence) that Claudius killed his

---

[6] Williamson (forthcoming b) makes a similar point. He notes that all we need to get, say, an argument from Gettier cases going is the truth of the counterfactual *If someone were in a Gettier case, they would have a justified true belief that is not knowledge*, and scepticism about counterfactual reasoning in general is very implausible.

father is no harder for the average reader to determine than the issue of whether at the end of Act IV Laertes knows (or is in a position to know, or has good evidence) that Hamlet killed his father. If we have sufficient skills to work out simple facts about the story, we have in principle sufficient skills to work out enough facts about the story to refute interesting philosophical theses.

4. Back to Knowledge

It's time for a quick review of what we've concluded so far. I've argued that it is best to conceive of arguments from possible cases as consisting *both* of arguments from (allegedly) known facts about those cases as well as arguments from intuitions about the cases. And I argued that when it comes to facts about particular cases, it is usually not good to try and defend the reliability of argument from intuition by arguing that intuitions are constitutive of concepts or meanings, for these intuitions rarely are constitutive. So in cases where we don't know what is going on in the case, we should to some extend focus our attention to intuitions about more general conceptual connections, particularly connections between the concept at hand and *reasons* of various kinds. This might all sound platitudinous, and its methodology so to some extent it is, but there is a lesson to be drawn here.

In many cases, for example trolley cases in ethics or the many variants of fake barn cases in epistemology[7], we face the following trilemma. In case one the concept intuitively applies, in case two the concept intuitively does not apply, and intuitively the differences between the cases do not matter to the applicability of the concept. Since general principles like the last intuition *are* constitutively tied to content, the methodology sketched here gives us a preference for hanging on to the principle that the two cases do not differ in any significant respects, and hence giving up one of the intuitions about the cases. That is exactly what we usually do in decision theory. Maher's (1993) argument against the Ellsberg intuitions just consists in showing the pair of intuitions Ellsberg wants to defend concern cases that differ only along lines that we have strong intuitive reason to think are irrelevant to the rationality of choices in the cases. And it's very hard to see any other approach as being remotely sensible when faced with the mess of intuitions about fake barn cases. The main take-home lesson from this discussion is that we have a general reason to take this approach on methodological grounds whenever faced with such a trilemma.

In an earlier paper I suggested we could use these kinds of reasons to cast doubt on Gettier intuitions. The argument I made there seems overstated. First, the principle I was appealing to, that important concepts should have simple analyses, isn't really that plausible. Second, as Williamson has stressed the Gettier argument can easily rest on facts about Gettier cases, not intuitions about them. And if they do, we really

[7] A genre well catalogued and advanced by Gendler and Hawthorne (forthcoming).

need to say something to undermine those facts, or at least our knowledge of them. I think some normative principles concerning knowledge we know to be true can do just this.

The first is that knowledge is a norm governing belief. That is, if a person does not know that *p*, they are violating a doxastic norm in believing that *p*. But the victim in a Gettier case is not violating any doxastic norms, so their belief counts as knowledge. It's notable that the general principle connecting knowledge and belief here is accepted even by some who disagree about Gettier cases. Jonathan Sutton, for example, agrees with the normative claim and uses it to infer that the victim in a Gettier case does violate a norm, i.e. he is *not* justified.

It's actually a little tricky to avoid concluding that knowledge is justified true belief once we accept that anything like this normative connection holds. What we have to do is accept that there are two different levels of doxastic normativity, and reaching the first level does for justification, while one has to reach the second level for knowledge. Of course one can find different kinds of virtues beliefs can have so that this is true, e.g. reliability to degree *x* and reliability to degree *y* greater than *x*, but this doesn't seem to undermine the *kind* of analysis the justified true belief theorist wants. Remember that the Gettier cases aren't usually thought to tell just against the analyses from Plato and Ayer than Gettier mentions, but against any analysis of this kind. But if knowledge is a norm of belief, then it follows trivially that there is some doxastic norm such that true beliefs which satisfy this norm, i.e. which satisfy a certain criteria of justification, are knowledge. It *doesn't* follow that this is a *reductive* analysis since the only access we have to this norm might be via our concept of knowledge, but this relatively trivial point does put a limit on how much we could possibly show via Gettier cases. We cannot show that no tripartite analysis of knowledge into a factual component, a doxastic component and a normative component succeeds, because knowledge itself could be the norm. In other words, we could just take the Gettier cases to show us that there are normative standards that are stronger than traditional standards of justification. For all we've seen, some of these may even allow for a reductive analysis of knowledge.[8] That might all seem a little trivial, so let's look at some connections that more strongly suggest the original Gettier cases *are* cases of knowledge.

Second, knowledge is a norm governing practical reasoning. This is a point frequently stressed by Hawthorne (2004). A person who does not know that *p*, but uses *p* as a premise in a practical syllogism does something wrong. But a Gettier victim who uses his new belief in practical reasoning does not do

[8] For example, the analysis offered by Lycan (forthcoming) has this tripartite structure, and it is designed to be immune to traditional Gettier cases. Lycan puts it forward as a *four-part* analysis, but two of the parts (the traditional justification part and the no false lemmas part) are normative, and we can take their conjunction to be an interesting doxastic norm.

anything wrong. Indeed, it seems hard to me to see how one could think that the victim is hereby doing something wrong unless one reasoned from Hawthorne's principle and the independent claim that the victim does not know.

Third, knowledge is a norm governing assertion. This is a point frequently stressed by Williamson (2000). A person who does not know that *p*, but asserts that *p* does something wrong. But the Gettier victim does not violate any norms in asserting his justified true belief, so it follows that this is something he knows. Again, I can imagine inferring that he violates a norm from Williamson's principle and the claim he does not know, but I can't imagine thinking that directly.[9]

All three arguments here are of a common form: there is a normative connection between knowledge and something else, and facts (or perhaps intuitions) about the something else plus the normative connection tell us that the Gettier victim knows. Now I don't have any *methodological* replies to the person who wants to deny the facts (and intuitions) about the something else. These are things which we are often wrong about, and about which there is no constitutive connection between the intuitions and truth, so these might all be wrong. So perhaps all three arguments here fail. But note that as long as these three arguments provide reasons that defeat our (alleged) knowledge that the Gettier cases are cases of justified true belief without knowledge, we'll have undermined that *argument* for the falsity of the justified true belief thesis. So the argument has to rest on intuitions, and has to be set against the three arguments from intuitions I've sketched here. These aren't conclusive reasons for thinking knowledge *is* justified true belief, but they are reasons for taking that doctrine more seriously than we have in the past.

[9] I'm here agreeing with Price (1998) that the norms governing assertion are justification, truth and belief. It's not obvious this clashes with Williamson's view that the norm governing assertion is knowledge, unless one takes intuitions about possible cases of (non-)knowledge too seriously.