Williamson, *Knowledge and its Limits*, Chapters 5 and 6

1. When you consider Ii (Mr. Magoo knows that if the tree is i +1 inches tall, then he does not know that the tree is not i inches tall) in its contrapositive form, as Branden did, it seems to me the intuition is overwhelmingly against it. For, it seems, Magoo's knowing that the tree is not i inches tall says nothing at all about whether what it is instead is i + 1 , i-1 or i + n for any other n. It's not *being* i doesn't say it's not i + 1, and how could Magoo's *knowledge* have any implications about how the world is?

Yet, in the positive form TW stokes the intuition for Ii well when he says (115, top):

> … even if he so judges [that it is i inches tall] and in fact it is i inches tall, he is merely guessing: for all he knows it is really i–1 or i+1 …

This is because Magoo can't discriminate visually between i inches and i+1 inches, etc. So he merely happened to get it right so far as the value is distinct from i+1 (so he *doesn't* know it is i inches tall, says TW). Thus, if it is i inches tall, then he does not know it isn't i+1 inches tall. Similarly, if it is i+1 inches tall, then he does not know it isn't i inches tall. And there we have Ii.

Why the two conflicting intuitions, about the positive and contrapositive forms of the claim? For me it was easier to start with the contrapositive and try to figure out why it's supposed to be true. Then that tells you exactly what underpins Ii. So:

Ii = K (i+ 1 ⊃ -K-i)
Consider just the thing claimed to be known: i + 1 ⊃ -K-i
Contrapose: K-i ⊃ -(i + 1)

What could Magoo's knowledge that it's not i have to say about it's not being a particular other value? Well, we know it's not coming from its being true that it's not i, so it must be coming from his knowing it. In other words, there's a requirement for knowledge which Magoo must have fulfilled if we count him as knowing, and this requirement implies that in the actual world the value is not i+1. The requirement is *safety*.

A belief is safe if in worlds 1) nearby to the one he actually believes it in and 2) such that he would still form the belief that he actually does, it's also true. Supposing Magoo knows it's not i, that belief must be safe on the assumption TW must be making here. That means it must be true in the worlds that are nearby to the actual world and in which he believes it. Given that Magoo can't discriminate between i and i+1, the actual-world tree must be, like, i+100 inches tall, so as to insure that the worlds nearby the actual world are *also* worlds in which his way of tending to form the belief that it's not i (i.e. using a faculty with imperfect discrimination) won't lead him to a false belief. If the actual world were i+1 then he'd form the false belief that it wasn't i in the close world in which it was i. That's not allowed because the antecedent says Magoo *knows* it's not i, and safety is required for knowledge. Safety is the precise counterpart of the intuition TW is pumping with his comment about "merely guessing".

So, as we've seen him do in the luminosity argument, he is trying to pass off Ii as an innocuous thing that's intuitively obvious, when in fact he's making a very particular assumption that knowledge requires reliability, and that the reliability in question is safety, not sensitivity or process, etc. For notice that Ii isn't true on a sensitivity view. This is easiest to see with the contrapositive:

K-i ⊃ -(i + 1)

If instead of safety sensitivity--if p weren't true then you wouldn't believe p--is required for knowledge, then K-i implies you *can* discriminate i from close possibilities (e.g. i+1 !), so the assumption that makes things go for TW, viz., that Magoo can't discriminate, isn't an assumption that's compatible with his knowing it's not i or it is i. More explicitly, K-i implies that if i were the height then you wouldn't believe it wasn't. To figure out which worlds this has to be true in we have to ask what the height is in the actual world—is it near to i or far? If it's far from i, like i+100, then the closest way for i to be the height is a thing very far away from the actual value. The actual Magoo could tell the difference, and so would fulfill sensitivity and and know it was not i. What if the actual world has a value very near to i, like i+1? Well, then the i-worlds Magoo has to be able to not believe –i in are very hard to discriminate from the actual world (in which it's i+1). But that's a problem for Magoo if he wants to have a sensitive belief. If we *assume* that Magoo has knowledge that it's not i (as in the antecedent above) he must *have* a sensitive belief. Nothing follows that says the world has to be such that having this sensitivity was easy for him. So nothing follows that says the actual world must not have had the value i+1. Maybe it did! If so, that just means his sensitivity organ must have been better than we were imagining.

Side note: Doesn't it seem *very strange* to assume as obvious that MAGOO *knows* i+ 1 ⊃ -K-i (as Ii assumes) when the reason it's true is so complicated and contentious?

2. Several complaints about this margins and iterations argument: I) it's circular: if you are going to try to argue against KK by assuming that knowledge requires reliability, then it seems to be begging the question. This is because if knowledge requires reliability it's perfectly obvious that the KK principle is false. This is due to the fact that "is that belief reliable?" and "does the subject know that belief is reliable?" are obviously distinct questions 1) because reliability is external so a person's having it says nothing about their access to the fact that they have it, and 2) because the second question is second order. Thus, answering the first in the affirmative says nothing about the second. II) while error certainly increases with increased iterations of K at least if you're a reliabilist, that's not an interesting reason why K doesn't imply KK. It's irrelevant to epistemology, it seems to me. (Analogous to what I said earlier in the semester about how lack of resolution isn't an interesting reason why the internal is not easier to know than the external, from the luminosity chapter.)

3. (124-125) Related to the second-order bit about KK: I think TW is wrong to take "reliably reliable" to be what KK means on a reliabilist view of K. The scope and meaning are all wrong. On, say, the process reliabilist view, *knowing p* means having a true belief that p and that belief having been formed by a reliable process. Hence, knowing that you know p is having a true, reliably formed, belief *that* your belief that p was (true and) reliably formed. "Reliably reliable" means something like that in most similar circumstances the process you use to form such a belief (as your belief in p) is a reliable one. This says nothing about whether you even *have* a belief that your belief in p is reliably formed. Suppose you have, since that's not the issue. It can be true that your belief that p is reliably formed and that you use that reliable process most of the time for forming such beliefs, so you're reliably reliable. We allowed that you even have a belief about whether the process you used to form the belief that p was reliable. Nothing about being reliably reliable says that belief is true, and nothing says it's reliable. (Your views *about* your actually reliable behavior could be thoroughly goofy for all "reliably reliable" tells us.) "Reliably reliable" is to one side of the KK principle. (Similar argument works for safety-type reliability.)

4. (136, ##1) I'm suspicious of the claim that the need for margins of error and the resulting failure of KK is the key to this paradox for lots of reasons, beginning with the fact that we can see use of another instance of Ii in this ##1 scenario, and yet it's not true for the same reason as with Magoo. It's not because of a requirement of safety on knowledge combined with our limited ability to discriminate. They can discriminate between the days quite easily. Rather, when it's far from the endpoint they don't have enough information to tell whether it will be today rather than tomorrow. It doesn't see right to say they need a margin of error.

5. (140, top paragraph) Voila, the key to the paradox is that there's a surreptitious use of the KK principle in understanding the reasoning of the students. However, that principle is general, and even if it fails systematically that doesn't in any way imply that it fails in all cases. KK may well happen to come along with K sometimes. All you need are instances to fill the gaps he's talking about here that are there in every step. But why is it remotely implausible that the students know that they know that the exam isn't on the last day? (We're told that they *know* that the teacher puts rings around all and only exam dates. So they *know* they're using a reliable indicator to form their belief that the exam is not on the last day.) And then, reasoning further, why not the same for the next step?

6. What bugs me about the section on conditionally unexpected examinations is that he's trying to show that the existential assumption that there is going to be an exam is inessential to the paradox, thus making it like the backward induction "paradox" in the iterated prisoner's dilemma, and that means he should take that assumption out and have the sense of paradox remain. Then explain how it gets solved. Problem is, when you take that assumption out you get a different setup that doesn't seem paradoxical, for all the reasons TW gives. So, it looks to me like a different problem.

Williamson on the Surprise Examination Paradox

10/18/06 (B.F.)

The surprise examination paradox has been around since the early 1940's. I think the best overall discussion of the paradox is given in Roy Sorensen's (terrific) book *Blindspots* (see website for reference). There are various renditions. Here's Ned Hall's (see website for a link to his paper):

> At the end of class one Friday afternoon, the professor announces to her students that she will give them an exam during one of next week's classes. (Class meets every day during the week.) She adds that the exam will be a surprise, in that the students won't expect, on the morning of exam day, that the exam will be that day. One of her cleverer students pipes up, saying that she cannot possibly fulfill her intention to give such an exam. "For it cannot be held on Friday: if it were, we would expect it on Friday morning (having noted that no exam had yet been given). So Friday is ruled out; the exam must take place on one of Monday through Thursday. But then, for exactly the same reason, it cannot be held on Thursday, else we would know that fact ahead of time (having noted that no exam had yet been given, and having ruled out Friday). And so on: It's really just a simple use of mathematical induction to show that your statement is inconsistent." The professor beams at her bright young student, and says nothing. Arriving in class next Tuesday, the students discover that they are to take an exam that day. None of them, of course, expect it. The exam consists of one question: "What was wrong with the clever student's reasoning?"

Wright and Sudbury (see website for link) give the following six *desiderata* for a solution:

(A) The account given of the content of the announcement should make it clear that it is satisfiable, since a surprise examination is, palpably, a logical possibility.

(B) The account should make it clear that the headmaster can carry out the announcement even after he has announced it since, palpably, he can.

These two conditions (A) and (B) require that the paradox not be construed as straightforwardly one of impredicativity or 'pragmatic self-refutation'.

- Some, like Shaw and Kaplan & Montague (see website for links), have claimed that there is some sort of vicious self-referntial character to the teacher's announcement. But, this 'liar-like' twist on the teacher's announcement seems inessential.

(C) The account must do justice to the intuitive meaning of the announcement. An extraordinary proportion of commentators have chosen to discuss quite unnatural interpretations of it.

- I think they have people like Shaw and Kaplan & Montague in mind here again.

(D) The account must do justice to the intuitive plausibility of the pupils' reasoning.

(E) The account should make it possible for the pupils to be *informed* by the announcement: we want the reaction of someone who notices no peculiarity but just gets on with his revision to be logically unobjectionable.

- Here, they are talking about Quine (see website) who seems to think that the students cannot know that the teacher's announcement is true (and this is the root of their fallacy).

(F) The account must explain the role, in the generation of the puzzle, of the announcement's being made to the *pupils*; there is, intuitively, no difficulty if, e.g. the headmaster tells only the second master or keeps his intentions to himself. Most of the interpretations in the literature which identify the problem as one of impredicativity fail to meet this condition.

Quine was the first to point out that the students are not really doing a *reductio* of the teacher's announcement, since they are importing extra assumptions into their reasoning that are not part of the announcement itself. For instance, when they rule-out Friday, they are assuming that they would (continue to) know (or expect) *on Thursday* that an examination was still going to occur. That's not part of the announcement itself. Quine's solution goes much fartrher than this observation, however. He concludes that the students are *never* in a position to know that the teacher's announcement is true (not even on the day it is made). This runs afoul of W&S's (E), and seems to be not fully diagnostic of what is going on. In effect, Quine's maneuver makes the n-day paradox tantamount to a *one*-day paradox: 'There will be an examination today and you do not know it'. Wright & Sudbury refine Quine's diagnosis, by positing a gap between what the students know at the beginning of the week and what they know at the end of the week. It's not that the students forget during the week, rather it's, as Williamson explains, that...

> ... their memory of examinationless days would undermine their earlier knowledge of the truth of the announcement, like misleading evidence. For them, to know on the last day that there will be a surprise examination, when there has been none so far, is in effect to know 'There will be an examination tomorrow and we do not know that there will be an examination tomorrow'. Such knowledge is impossible, for their knowledge of the first conjunct is inconsistent with the truth of the second.

This paradox (which the students would find themselves in if they did retain all their knowledge until Thursday evening) is intimately related to Fitch's knowability paradox (chapter 12). But, none of this shows that the students couldn't have known *at the beginning of the week* that the teacher's announcement was true. Intuitively, they could have.

Williamson introduces another paradox called The Glimpse, which has some similarities with the surprise examination paradox, but also a few differences. Williamson does argue convincingly, though, that his arguments in chapter 5 will diffuse both paradoxes in a unified way. The Glimpse:

> A teacher's pupils know that she rings all and only examination dates on the calendar in her office. At the beginning of term, the only knowledge they have of examination dates this term comes from a distant glimpse of the calendar, enough to see that one and only one date is ringed and that it is not very near the end of term, but not enough to narrow it down much more than that. The pupils recognize their situation. They know now that for all numbers i, if the examination is $i + 1$ days from the end of term then they do not know now that it will not be i days from the end ($0 \leq i < n$). In particular, they know now that if it is on the penultimate day then they do not know now that it will not be on the last day. But they also know now from their glimpse of the calendar that it will not be on the last day. They deduce that it will not be on the penultimate day. They also know now that if it is on the antepenultimate day then they do not know now that it will not be on the penultimate day. They deduce that it will not be on the antepenultimate day. And so on. They rule out every day of term as a possible date for the examination.

The Glimpse is analogous to the surprise examination in several respects:

- "The teacher's announcement corresponds to the claim in the Glimpse that if the examination is $i+1$ days from the end, then the pupils do not know that it is not i days from the end ... To say that in the Surprise Examination the pupils cannot know in advance that the announcement is true corresponds to saying that in the Glimpse the pupils cannot know the limitation on their knowledge by reflecting on the poverty of their perceptual knowledge in that case."

- "The Glimpse is not a Liar paradox, nor is the Surprise Examination on its natural reading."
- "That vagueness is not to blame is even clearer in the Surprise Examination than the Glimpse."

The analogy is less perfect when it comes to iterations of knowledge:

- "The paradox of the Glimpse depends on a concealed use of the KK principle."
  - This is for the same reason that Ramachandran's argument from last week did not work as an argument against the KK principle. We must carefully distinguish the students' reasoning from the theoretician's reasoning in the Glimpse. Let p be 'The examination will be on the penultimate day' and l be 'The examination will be on the last day'. And, let Kq express that the students know that q. Then, the theoretician, reasoning about the students' knowledge might be tempted to think they can reason as follows:
    1. $K(p \supset {\sim}K{\sim}l)$ Premise.
    2. $K{\sim}l$ Premise.
    3. $K{\sim}p$. 1, 2, closure of the students' knowledge under deductions *they* perform.

    However, step (3) is not kosher here (a similar problem plagued Ramachandran's argument against KK last week). If the students' premises are $p \supset {\sim}K{\sim}l$ and ${\sim}l$, then ${\sim}p$ does not follow. To deduce ${\sim}p$, the students need the extra premise $K{\sim}l$. And, the theoretician can only apply closure to this reasoning if the reasoning in question is from premises *known to the students*. This means the theoretician needs to assume, also, that $KK{\sim}l$ in order to get a kosher deduction of $K{\sim}p$. This is a hidden instance of KK, and, in fact, one instance of KK will be needed for *each step* of the backward induction.
  - It is interesting that a hidden application of KK is required for The Glimpse. Some have claimed that KK is required for the surprise examination paradox (*e.g.*, Harrison and McLelland & Chihara — see website for links). But, it is not. In fact, all that is needed for the surprise examination is that "the pupils know on the first morning of term that they will know on the second morning that ... they will know on the penultimate morning that they will know on the last morning the truth of the teacher's announcement." So, here we have a disanalogy between The Glimpse and the surprise examination. However, Williamson's arguments from §5.3 show that neither of these sorts of iterative properties of knowledge hold in general. He sees this as evidence of generality and unification.

- "In the Surprise Examination each iteration of knowledge involves a change in cognitive standpoint, whereas the Glimpse (like the KK principle) involves a fixed cognitive standpoint."

  > The iteration of knowledge operators leads sooner or later to falsity through a process of erosion resulting from the need for margins for error. This applies just as much when the knowledge operators refer to different cognitive standpoints.

  Williamson's general, unified diagnosis of both The Glimpse and the surprise examination:

  > ... a small difference between actual pupils and idealized ones is magnified by each stage of the reasoning until it is clearly visible. That difference is the margin for error.