Fall 2006 Sherri Roush

Williamson, *Knowledge and its Limits*, Chapter 7, "Sensitivity"

I. Stuff about counterfactuals
II. Sensitivity and Skepticism (and Closure)
III. Sensitivity and Methods (and Closure)
III. Ship example (and Closure)
IV. Sensitivity and Broad Content

I. Stuff about counterfactuals

a. Note that in the hardback version in the first paragraph of this chapter, there's a mistake where TW argues by example that sensitivity does not imply safety. This is true, but his example goes in the wrong direction, to show that safety doesn't imply sensitivity (which is also true). An example that serves the purpose is a necessary falsehood. (The problem is corrected in the paperback.)

More interestingly, though this fact, and the converse that safety does not imply sensitivity, are often cited that is a bit of missing the forest for the trees, since in the domain of *true* belief sensitivity provides uniform probabilistic assurances of safety (but not the converse), and beliefs that aren't true aren't knowledge anyway, before we get to the safety or sensitivity requirement on knowledge. (See *TT*, chapter 4: "Sensitivity and Safety: No Trade-off", pp. 123-124.)

b. TW claims on page 148 that "Any counterfactual conditional entails the corresponding material conditional, so in particular –p $\square\!\!\rightarrow$-Bp entails –p ⊃ -Bp." Since the material conditional is equivalent to – (-p . Bp), this implies that if a belief satisfies sensitivity then (in the actual world) it's not false (but true), in other words: "S sensitively believes p only if p is true". But this implication should seem bizarre, because it makes truth a variable that's not independent of the X condition on knowledge which is the classic way to define infallibilism. However, the tracking view was always understood as a fallibilist view, the obvious symptom of which is that the counterfactual only requires its consequent to be true in a subset of –p worlds. In those far off worlds where –p holds the counterfactual requires nothing about your beliefs so you might have been wrong if one of *those* worlds had been the case. You can satisfy the sensitivity criterion while you still *might have been* wrong—one notion of fallibilism—and yet because of properties of counterfactuals satisfying the sensitivity criterion implies your belief is true—a rejection of the other standard notion of fallibilism, that of truth as an independent variable. So there are two notions of fallibility at work that come apart here. However, notice what's going on: the only reason that it follows from satisfying sensitivity for p that p is true in the actual world is that you had to specify that p was true in the actual world in order to evaluate the counterfactual, in order to orient that evaluation. So it falls out essentially because it was put in, not because what your dispositions to believe in –p worlds give you is protection against the possibility of believing p falsely.

II. Sensitivity and Skepticism (and Closure)

Let's rehearse the derivation of the "**abominable disjunction**": either we have no knowledge or knowledge isn't closed under known implication.

Closure: If you know p, and you know that p implies q, and you believe q on the basis of p, then you know q.

(1) Necessarily, if S knows p then if p were false, S would not believe p.

Consider the following (the Moorean direction):

I have a hand.
Therefore, I am not a brain in a vat.

Skeptic: but you ***don't*** know you're not a brain in a vat (appeal to (1)), so you can't know you have a hand because if you knew you had a hand you *would* know you weren't a brain in a vat (appeal to closure). Note that the skeptic can't get this going merely in virtue of the fact that having a hand *implies* one isn't a brain in a vat. (Everyone—except modal logicians and David Lewis—agrees that knowledge isn't closed under implication.) Closure under known implication is required.

Thus, agreeing to (1), which is to say taking sensitivity (or tracking) as a *necessary* condition for knowledge "commits one to the disjunction of non-closure and rampant skepticism." (TW, 151)

DeRose's contextualism avoids this implication by taking knowledge to be closed only within contexts. (Though coming to know an implication sometimes changes the context (as with inferring the negation of the skeptical hypothesis from an ordinary truth), so technically there won't be as many failures of closure as you might expect.) Thus, within the skeptical context, there is closure and you don't know anything, but in a non-skeptical context – say, where no one has mentioned the skeptical possibility – you know ordinary things, such as that you have hands. Except for avoiding the abominable disjunction DeRose's view inherits all the problems of Nozick's view. But he says he's not doing theory of knowledge. Rather, he's trying to explain the appeal of skepticism (via a version of (1)) without giving in to it. My feeling is that contextualism has all the virtues of theft over honest toil. It is implausible to me that whether you have knowledge should change *merely* on the basis of whether people around you have talked about the skeptical possibility.

The **probabilistic (closed) tracking view** takes a version of (1) (plus an adherence condition we'll ignore) to be *sufficient* but not necessary for knowledge, and imposes closure on knowledge by means of a recursion clause. So, roughly,

S knows p iff S believes p, p is true, and

either P(-Bp/-p) > .95 and P(B(p)/p) > .95 and P(-B(-p)/p) > .95 [i.e., S tracks p]

or there is a q such that q is true, S believes q, S tracks q, and S knows q implies p. (*TT*, p. 47)

This means that you can know things either by tracking them or by knowing they're implied by some proposition(s) you track. It's a disjunctive account of knowledge where you have a "core" of more robust knowledge, the things you track.

Implications for skepticism (see *TT*, 54-57):

I have hands.
Therefore, I'm not a brain in a vat.

If things are good, then you *do* track that you have hands. You don't track that you're a brain in a vat (standardly—see last section below), but you don't have to since you know that your lack of envatment is implied by your having hands. The Moorean option is restored. (Of course the theory of what knowledge *is* doesn't tell you that you have this knowledge. Things in addition have to actually be good.) To go in the skeptic's direction we're going to have to find a reason to doubt that we're not brains in vats. For the skeptic to appeal to a tracking condition there would look circular from my point of view. (That doesn't mean it *is* circular.)

Implications: Once we have done the Moore argument (even implicitly), we have done all we can to make a difference to whether we know. The rest is up to the way the world is. This is because it's not *possible* to track "I'm not a brain in a vat" (standardly), so no further effort we make could make any difference to whether we know. (This is why scientists don't care about philosophical skepticism. Science is about improving knowledge through effort.) Also, the fact that it's not possible to track the negation of the skeptical hypothesis (standardly) explains why the skeptical question is so frustrating.

III. Sensitivity and Methods (and Closure)

Nozick had a problem, easily illustrated with the grandmother case. p = My grandson is well. Grandma sees the grandson and knows how to tell. But if p were false the relatives would spare her pain and lie to her, so she would believe that he was well just the same. Nozick makes what I take to be a bad turn by addressing this problem with a relativization to method. So:

(2) Necessarily, if S knows p via method M, then if p were false and S were to use M to arrive at a belief whether p, S would not believe p.

So, when you vary the worlds to see if the grandma believes or not, you keep the method she actually used fixed. Relativizing to method leads to baroque problems because more than one method may actually be employed, so that Nozick has to introduce a new notion of "outweighing" between methods. But in those cases the right answer would have popped out if he had just stuck to allowing the methods to vary in evaluation of the counterfactual. In my view, the criteria that solve the generality problem also spit out the result that the method *happens* to be fixed in a lot of normal cases, in particular the cases where whether *that* method was used is probabilistically independent of the truth value of p. (See *TT*, pages 79-97fn.) This renders the outweighing nonsense unnecessary. The grandma case is solved in my view by closure: grandma tracks a lot of things that imply that the grandson is probably well: he is walking around, his cheeks have a healthy glow, etc. and she knows that these things imply he is probably well.

Nozick's approach to method would have to individuate methods coarsely in order to make it work for grandma. However, there are other cases where you'd have to individuate finely in order to get to the right answer. (E.g. Jesse James, *TT*, 68-71.) Closure removes those problems too. The dog problem (that Goldman originally presented against Nozick's view, and that he himself avoids by distinguishing methods by what's on the inside, which TW objects to) also gets resolved by closure on my account. (*TT*, 57-59)

Digression: Individuating finely vs. coarsely makes a big difference to the answers you get. In fact every choice about what to hold fixed and what to vary in these conditions has the potential to make a big difference to the answer you get as to whether it's a case of knowledge. Thus, in order to have a theory at all you need to give a general rule about this varying. With counterfactuals this means giving a similarity relation, but one seems to need more as well, as in individuating methods coarsely or finely (those two *words* aren't enough to do the job of pinning things down either). This is the (counterfactual) tracking view's version of the *generality problem* that all externalist views have. The problem hasn't been solved for the counterfactual version of the view, but I have a solution for the probabilistic version (*TT*, ch. 3).

Despite his criticisms of Nozick, Goldman, and the individuation of methods internally, TW still wants to relativize to method, and seems to think it is necessary. I don't want to relativize since I think it is not only unnecessary but *bad*. (See *TT*, 69-70.) This is because when method is fixed in the variation of worlds, the tracking conditions won't be checking whether the method actually used was used for a good reason or the person could easily have used *any* method. (There's an example on pp. 69-70 of *TT*.)

III. Ship Example (and Closure)

This example is supposed to show that knowledge is compatible with widespread slight insensitivity. It's a good point and good example. (Goldman's dog was the earliest probably.) However, with closure this kind of thing can be handled so we needn't say that sensitivity isn't part of knowledge in order to acknowledge it.

The example: Although I can distinguish 19 and 23 meters, I can't distinguish 19 and 21. (This gives rise to my insensitivity.) In actual fact the water line on the ship is 1 meter high. I judge it, correctly, to be < 20 meters high. Now, I can distinguish 1 and 20 meters from each other, so I know that the line is less than 20 meters high. However, the closest way for it not to be less than 20 meters high is for it to be 20 or 21 meters, which I cannot distinguish from 19, so I might well take it to be < 20 meters high. My belief that it is < 20 meters high is insensitive yet surely I know it's < 20 meters high since it's only 1 meter!

This problem can be handled with closure. (*TT*, 71-72) TW says the man can know it's < 20 meters without having any belief at all about what it actually is. This is implausible unless we're insisting on a belief about a specific number. The man must be willing to bet that the distance is roughly between .5 and, say, 3. (Or raise it if you like as long as you stay well below 20. If he isn't disposed to do something like this, it's unclear to me why we should count him as knowing it's less than 20 from it's being 1 and his looking at it.) If he is willing to so bet then (on my view) he has a belief that it's roughly between .5 and 3, and this belief is sensitive: if it weren't roughly so, then he wouldn't believe it was. The lack of sensitivity is simply built into which proposition we know in this case: only "roughly between .5 and 3", although this belief implies that it's < 20, so from it the man does know that. The key to the solution is the claim that someone can be sensitive to the "roughly" claim. TW doesn't canvas this option. I'm not sure whether there's a principled reason why. Perhaps he doesn't think propositions like "It is roughly 1 meter" are really things that have truth values, since vagueness is epistemic for him. (Of course, he wouldn't want to appeal to that view here.)

There now seems to me a simpler way to solve this problem. Look, we're imagining the alternative scenario in which –p as one where the distance is 20 or 21 (it must be so to lead to the lack of discrimination), and because the man can't discriminate 20 from 21 we're counting him as insensitive. But this doesn't follow. An epistemologically sound person who can't discriminate between 19 and 20 or 20 and 21 would simply not have a belief in that situation. (Thus –p □→ - Bp would not be violated.) If our guy *would* believe then we've got to wonder about his general tendency to believe when he can't discriminate. That restores the intuition that if all this is the case we shouldn't count the person as knowing it's < 20 meters, even though it's actually 1 meter and he's looking at it.

It is also arguable that this problem goes away just through switching to conditional probability. If the distance weren't less than 20 then the *closest* way for this to occur is for it to be 20 or 21, which the guy wouldn't be able to distinguish from being less than 20, but are those among the *probable* ways? I suppose we could imagine a case where they are, but then the previous objection still applies.

IV. Sensitivity and Broad Content – missing the forest for the trees

TW points out that externalism about content means that after all we do satisfy (1) for p = I am not a brain in a vat. (If I were a brain in a vat I wouldn't believe it because I *couldn't* believe it due to the fact that I failed to fulfill the requirements for referring to vats.) The Putnam version of this argument is limited, as has been known for a long time, since one could still have been envatted just yesterday and retain one's ability to refer for a while. But there's a way to maintain the point, by using q = I see this pen, from which I deduce p. You can't get the indexical to refer to this pen if you're a brain in a vat, even if you were just envatted yesterday. So you don't fail to be sensitive to q. Since being a brain in a vat (-p) doesn't explain why I would believe q falsely – if I were a brain in a vat I wouldn't be able to believe q at all, I satisfy condition (4). This works wrt (4), though not wrt (1) since the latter has no clause about deducing.)

TW's reason for bringing this up is to say that DeRose's project of using (4), which is a version of (1), to explain why skepticism is appealing must fail, since with content externalism we *do* satisfy (4). (And content externalism is surely appealing.) However, this misses a possible outlook for DeRose. It's clear that traditionally we were content internalists. It requires effort and thought to see the appeal of content externalism. So DeRose may just say that what explains the appeal of skepticism is the assumption of content internalism and criterion (4). I.e., it's insofar as we believe these two things that we find skepticism appealing. That this outlook is available is confirmed by the fact that if you think about it skepticism *does* cease to be appealing when you bring content externalism into your assumptions.

Notes on Counterfactuals, Likelihoods, and Tracking
10/25/06 (B.F.)

1 Setting the Stage

Williamson's arguments concerning sensitivity rely on two basic assumptions:

1. Sensitivity is to be explicated *counterfactually*: S sensitively believes p iff S believes p and had p been false, S wouldn't have believed p.
   Formally, SBp iff Bp and $\sim p \;\Box\!\!\rightarrow \sim Bp$.

2. The semantics for $\Box\!\!\rightarrow$ is a possible-worlds semantics, which presupposes:

   (a) $p \;\Box\!\!\rightarrow q$ is (actually) true iff the closest $(p \,\&\, q)$-world is closer (to the actual world) than the closest $(p \,\&\, \sim q)$-world.

   (b) The actual world is the closest world to itself.

Sherri, on the other hand, does not explicate sensitivity counterfactually. Rather, she understands sensitivity in *probabilistic* terms. For Sherri:

3. S sensitively believes p iff S believes p and the probability that S doesn't believe p conditional on p being false is high.
   Formally, SBp iff Bp and $\Pr(\sim Bp \mid \sim p) > t$.

Recently, Gunderson (see link on website) proposes a probabilistic semantics for $\Box\!\!\rightarrow$, which leads to a counterfactual account of sensitivity that is very similar to Sherri's. Specifically, Gunderson combines (2a) with the following.

4. The closest $(p \,\&\, q)$-world is closer than the closest $(p \,\&\, \sim q)$-world iff $\Pr(q \mid p) > 1/2$ and $\Pr(q \mid p) \gg \Pr(q \mid \sim p)$.

Gunderson is not totally clear about how to understand "$\gg$". Two possibilities:

(LD) $\Pr(q \mid p) \gg \Pr(q \mid \sim p)$ iff $\Pr(q \mid p) - \Pr(q \mid \sim p) > d$.

(LR) $\Pr(q \mid p) \gg \Pr(q \mid \sim p)$ iff $\Pr(q \mid p) / \Pr(q \mid \sim p) > r$.

If we assume (LD), then (with suitable choices of d and t), Gunderson ends-up with a counterfactual account of sensitivity that is equivalent to Sherri's probabilistic account. And, if we choose (LR), then Gunderson ends-up with a counterfactual relation of sensitivity that is *stronger than* Sherri's relation. In any event, Gunderson must reject (2b). His (4) entails that (2b) is sometimes *false*. This will become clear, below, when we discuss properties of these three approaches to sensitivity. The three approaches to sensitivity are:

(S_1) A standard counterfactual approach which presupposes (1), (2a), and (2b).

(S_2) Sherri's probabilistic approach, which presupposes only (3).
[Note: this is equivalent to Gunderson's (2a), (4), and (LD).]

(S_3) Gunderson's (stronger) counterfactual approach: (2a), (4), and (LR).

Next, I will list some important (logical) properties of standard counterfactuals (some of which are explicitly used by Williamson). Then, I will compare and contrast our three theories above with respect to these properties.

2 Logical Properties of Standard Counterfactuals

2.1 Materiality

The following property is satisfied by the standard counterfactual $\Box\!\!\rightarrow$.

Materiality. If $p \Box\!\!\rightarrow q$ is actually true, then $p \supset q$ is actually true.

- Proof that Materiality holds for the standard counterfactual. Assume $p \Box\!\!\rightarrow q$, *i.e.*, that the closest $(p \,\&\, q)$-world (to the actual world α) is closer to α than the closest $(p \,\&\, {\sim}q)$-world is. And, assume that α is the closest world to itself. These are the background assumptions of Materiality for the standard counterfactual. Now, assume (for the purposes of doing a conditional proof of $p \supset q$) that p is actually true. If q were actually false, then (contrary to our assumptions) there would be a $(p \,\&\, {\sim}q)$-world that is as close to α as the closest $(p \,\&\, q)$-world (namely, α itself!). Thus, α must be a $(p \,\&\, q)$-world. Hence, q must be actually true, which suffices to establish the actual truth of $p \supset q$ from the actual truth of $p \Box\!\!\rightarrow q$. QED.
- Neither of Gunderson's counterfactuals satisfy Materiality (in general). This is because there exist probability functions $\Pr(\cdot \mid \cdot)$ such that both conditions (i) $\Pr(q \mid p) > 1/2$ and (ii) $\Pr(q \mid p) \gg \Pr(q \mid {\sim}p)$ obtain, and this possibility is consistent with $p \supset q$ being actually false. If it weren't consistent with this, then whenever (i) and (ii) were satisfied in an actual case, one could infer q from p *with certainty*. This is absurd, since (i) and (ii) are consistent with $\Pr(q \mid p)$ being less than 1. Here, you can think of $\Pr(\cdot \mid \cdot)$ in either subjective (credence) or objective (chance) terms.

This property has serious consequences for our three theories of sensitivity.

- According to theory S_1, which is based on the standard counterfactual, Materiality implies that if S actually sensitively believes p, then p is actually true. Williamson makes good destructive use of this consequence. Tracking theorists (like Nozick) do not want the actual truth of p to be *redundant* in their account of knowledge. Thus, any account of the form:

  S actually knows p iff

  (a) S actually believes p (Bp)

  (b) p is actually true (p)

  (c) $p \Box\!\!\rightarrow Bp$ is actually true (adherence[1])

[1]Typically, it is also assumed that $p \Box\!\!\rightarrow {\sim}B{\sim}p$. See footnote 2 for further discussion.

(d) $\sim p \;\Box\!\!\rightarrow \sim Bp$ is actually true (sensitivity, *a.k.a.*, variation)

must be presupposing a notion of sensitivity that is *not* counterfactual in the standard sense. Else, the actual truth of p *would* be redundant, by the Materiality of the standard counterfactual conditional. Avoiding Materiality while maintaining (2a) requires a *radical* revision of (2b).

- On the other hand, S_2 and S_3, which are probabilistic (and in similar ways), are consistent with S actually sensitively believing actual falsehoods. This avoids some of TW's criticisms of sensitivity. Moreover, it avoids the redundancy of the actual truth of p in tracking accounts of knowledge. Gunderson's theory of $\Box\!\!\rightarrow$ maintains (2a), but *radically* rejects (2b). Indeed, on Gunderson's account, there can be worlds that are "closer" to the actual world than the actual world is to itself! In fact, this *must* happen whenever Materiality is violated by a Gunderson conditional. This, I think, gives one reason to doubt if Gunderson is even talking about *closeness* anymore. [Is he even talking about a *conditional* anymore? See below.]

2.2 Centering

The following property is satisfied by the standard counterfactual $\Box\!\!\rightarrow$.

Centering. If $p \,\&\, q$ is actually true, then $p \;\Box\!\!\rightarrow q$ is actually true.

- Centering holds for the standard counterfactual, since if the actual world (α) is a $(p \,\&\, q)$-world, and (2b) α is the closest world to itself, then the closest $(p \,\&\, q)$-world (α) is closer (to α) than the closest $(p \,\&\, \sim q)$-world.

- Neither of Gunderson's counterfactuals satisfy Centering (in general). This is because it is possible to have (i) $p \,\&\, q$ is actually true, and at the same time either (ii) $\Pr(q \mid p) \leq 1/2$, or (iii) $\Pr(q \mid p) \not\gg \Pr(q \mid \sim p)$, or both.

This property does not have serious consequences for our three theories of sensitivity. The consequences of Centering have more to do with *adherence*, which is traditionally explicated counterfactually as (c) above: $p \;\Box\!\!\rightarrow Bp$.

- Centering implies that an agent can trivially "adhere" to true p's, just by believing them. In other words, adherence is *redundant* in standard counterfactual theories of tracking. As such, anyone who wants a non-redundant theory of tracking (of the form given above) must presuppose a notion of adherence that is *not* counterfactual in the standard sense. In this case, however, we needn't reject (2b) in such a radical way as Gunderson does. Just by allowing that there is a *sphere* of worlds that are equally "close" to the actual world (where this sphere contains the actual world itself), we can avoid Centering (this is what Nozick does).

2.3 Lattice Properties

The following lattice properties are satisfied by most *conditionals* $\Box\!\!\rightarrow$ I know:

(A) $p \Box\!\!\rightarrow p$.

(B) If $p \Box\!\!\rightarrow q$ and q entails r, then $p \Box\!\!\rightarrow r$.

(C) If $p \Box\!\!\rightarrow r$ and $q \Box\!\!\rightarrow r$, then $(p \vee q) \Box\!\!\rightarrow r$.

(D) If $p \Box\!\!\rightarrow q$ and $p \Box\!\!\rightarrow r$, then $p \Box\!\!\rightarrow (q \,\&\, r)$.

It is easy to show that the standard counterfactual conditional has all of these lattice properties. But, neither of Gunderson's counterfactuals satisfy *any* of them. Property (A) will be violated for any p that has extreme (unconditional) probability. Property (B) will be violated whenever r is tautological. Counterexamples to (C)–(D) require a bit more ingenuity, but they all fail because the underlying relation "p is strongly correlated with q" violates them. In light of the fact that Gunderson's counterfactual violates *Modus Ponens* as well as (A)–(D), I think it's no longer clear why we should be calling it a *conditional*.

One motivation for Gunderson's unorthodox semantics for $\Box\!\!\rightarrow$ is that it leads to a better counterfactual theory of *tracking*. It is interesting that the theory he ends-up with is so similar to Sherri's theory. I think it's better to do what Sherri does (cut out the middle-man), since there is no reason to think that Gunderson is even talking about the semantics of *conditionals* anyway. [Note: It is no problem for Sherri that her *tracking* relation doesn't satisfy *Modus Ponens* or (A)–(C).[2] Why should the tracking relation be anything like a *conditional*?]

3 Probabilistic Sensitivity and TW's Ship Example

The example (Sherri's rendition): The water line on a ship is 1 meter high. I judge it, correctly, to be less than 20 meters high. I can distinguish 1 and 20 meters, so I sensitively believe that it's less than 20 meters high. However, the *closest* (in possible worlds talk) way for it not to be less than 20 meters high is for it to be 20 or 21 meters, which I cannot distinguish from 19. So I might take it to be less than 20 meters high. Then, my belief that it's less than 20 meters high is (counterfactually) insensitive yet I know it's less than 20 meters high since it's only 1 meter high.

This example seems tougher to run *probabilistically*. We would need a case in which p, Bp, and $\Pr(Bp \mid {\sim}p) > t$. That is, the probability that I believe the line is less than 20m, given that it is not less than 20m needs to be sufficiently high. It will take more work to make this a counterexample to sensitivity in *this* sense. We'd need to assume that the one-sided ${\sim}p$-distribution is heavily weighted near 20m. Undoubtedly, there are such examples. But, no matter, since Sherri's disjunctive combination of tracking and closure takes care of this problem anyway (as well as TW's dilemma concerning skepticism and closure).

[2] Since (D) is violated, Sherri needs to add $\Pr(Bp \,\&\, {\sim}B{\sim}p \mid p) > t$ to her tracking conditions. For the same reason, a Gundersonian would need to add $p \Box\!\!\rightarrow (Bp \,\&\, {\sim}B{\sim}p)$ to his conditions. For the standard $\Box\!\!\rightarrow$, (D), adherence, and $p \Box\!\!\rightarrow {\sim}B{\sim}p$ imply $p \Box\!\!\rightarrow (Bp \,\&\, {\sim}B{\sim}p)$. See footnote 1.