

On Skepticism and the Self-Knowledge Rule

Tomorrow I'll be presenting the paper "The Self-Knowledge Gambit." To make sure that what we talk about today is sufficiently different from what I will discuss tomorrow in the practice jobtalk, I would like to first talk about the philosophical motivations for the Self-Knowledge Rule. Then I will tell you what my main claims are. Then I will give you my main arguments—though you'll hear more about them tomorrow. Finally, I would like to discuss some worries about my account. I'd be very grateful for suggestions about what to say, and for an assessment of how pressing these worries are.

1. Philosophical Motivations

An opinionated view of the philosophical background (something I can't defend here):

- a) Externalism in the philosophy of mind is correct.
- b) Externalism in epistemology is correct.

Nonetheless skepticism is defensible! Usually skepticism is associated with internalism, in at least one of these senses. Yet I think that a skeptic can do without such an assumption.

I understand a) and b) in terms of the following theses: (I won't defend this either.)

A) Our evidence is different in the good case and the bad case. That's why, from a failure of knowledge in the bad case, we cannot infer that we lack knowledge in the good case. Because I fail to know that I am dreaming, while dreaming, I don't fail to know that I am not dreaming, while not dreaming.

I think that this could be made out to follow from a view about externalism in the philosophy of mind. Since the experiential states are different in the good case and the bad case, the evidential states are different in the good case and the bad case. That's why, when one is in the bad case, one cannot distinguish them from the good case.

B) The KK-principle is false.

KK-Principle: Necessarily, if S knows p , S knows that she knows p .

I think that this is the basic epistemological externalist commitment. There are forms of epistemological internalism that do without it, but then they either give up the spirit of internalism, even if not the letter, or they don't disavow commitment to this principle after all. —This is not something I can argue for, though.

So—how are we to get a doubly externalist skepticism? It's useful to consider Branden's analogy from the discussion of Williamson on skepticism:

“Let’s say you’re someone who values money for its *instrumental* power. You like having money only to the extent that it will actually do things for you. If it’s just sitting in the bank, never to be used, then (we shall say) it has no *cash value* for you. As it turns out, *unbeknownst to you*, you “possess” lots of money. Some benefactor has placed a large sum of money in an account that (legally) you own. But, the catch is that he didn’t tell you anything about this, and you will never find out about it. Are you rich? Well, there is a sense in which you are rich, but it’s not a very interesting sense, especially by your own lights. *Legally*, you “have” lots of money. But since – as far as you know – you’re the same poor schmuck you were yesterday, this is a sense of ‘rich’ that has no *cash value* for you. I suspect that something analogous may be happening in this discussion about what evidence you “have” in the good vs bad cases.”

I think that this is a useful analogy in many ways. So I’ll explain my view through this analogy.

Suppose that traditional skeptical arguments establish that we don’t know whether we have any evidence. (If $E=K$, then that amounts to the conclusion that we don’t know whether we know.) Put in terms of the analogy, suppose we help ourselves to the conclusion that we don’t *know* whether we have any money, whether we in fact have it or not.

[We must postpone the question why you don’t know whether you have evidence, rather than not knowing whether you have evidence for the thought that you have evidence. If you assume $E=K$, we must thus postpone the question why you lack second- rather than higher-order knowledge.]

1) Branden’s story illustrates how one should be a skeptic while accepting assumption A. In the bad case you have no money. In the good case you have money. Still, the good case is defective, because the money has no “cash value” for you since you don’t know about it.

Analogously: In the bad case you have no evidence. In the good case you have evidence. But that evidence is defective, because you don’t know about it.

2) Branden’s situation also illustrates how one’s skepticism would not rest on the KK-principle: From the fact that you don’t know that you have money, it doesn’t follow that you don’t have money. From the fact that the money doesn’t have “cash value” for you, it doesn’t follow that you are poor.

Analogously: From the fact that we conclude that we doesn’t know whether we have evidence, it doesn’t follow that we don’t have it.

I want to make two more points in terms of this analogy:

3) There is a way in which “possessing money,” even if it has no “cash value” for you, is different from knowing something. You can possess money without doing anything whatsoever. However, in order to know something, you have to believe what you know. (At least that’s quite generally, though not universally, accepted.) So if you give up your

belief, you'll lose your knowledge. And I argue that not knowing whether you have evidence should move you to give up your belief even if in fact, unbeknownst to you, you do have evidence.

4) Finally, what requires explaining is how exactly your epistemic situation is bad if you don't know whether you have evidence or not. What's the analogue of lacking "cash value," even if you do possess knowledge? I want to say: You are estranged.

(There is a problem here. It's worry number two, below. I'll come back to it.)

2. My Main Claims: The Self-Knowledge Rule and Transparency

a) The Self-Knowledge Rule.

"One must: assert or believe p only if one knows that one knows p ."

The Self-Knowledge Rule is modeled on TW's Knowledge Rule, except that it's more demanding. (I follow TW in treating belief and assertion as interchangeable.)

The Self-Knowledge Rule underwrites *descent*: It moves us from a failure of higher-order knowledge to a failure of lower-order knowledge. – Suppose that skeptical arguments show that we don't know whether we know the things we take ourselves to know (or that we don't know whether we have any evidence about the world). By the Self-Knowledge Rule, we should suspend our beliefs about the world. If we do so, we lose our first-order knowledge, even if we had some beforehand.

[To take up the question bracketed earlier: If we can descend, then it doesn't matter whether skeptical arguments establish that we lack second-order knowledge or higher-order knowledge.]

Yet what is 'descent'? It's not logically valid reasoning. It doesn't follow from "S doesn't know whether she knows that p ," that "S doesn't know that p ." Indeed, the rejection of the 'KK-principle' makes it the case that this doesn't follow. Rather, on my proposal, skepticism is reached through *suspension of belief*. So the skeptical conclusion will have to be stated in the first person, singular or perhaps plural, present tense.

I think that that's not at all implausible: Skepticism is a first-personal, deliberative affair. Indeed, you can see this in the First Meditation. I think that for skeptical possibilities to be threatening, they have to be considered from the first-person perspective. For example, the evil demon possibility will only appear epistemically threatening if you consider it from the first-person point of view. – And I want to say: that's when externalist maneuvers won't be of use to you. Even if in fact you have better evidence when you are not deceived by an evil demon than when you are not so deceived, and even if you in fact know things when you are not deceived (for instance, because you then *perceive* them), that won't help you in settling the question whether here and now you know that you are not deceived.

The attraction of the Self-Knowledge Rule is this: You don't need any kind of internalism! The norm for belief replaces the internalist claims about what constitutes knowledge or evidence or justification. That's why the Self-Knowledge Rule serves my doubly externalist purposes.

Now there is a big problem with the Self-Knowledge Rule: I don't know what kind of norm it is. I am attracted by the thought that it's a constitutive norm for something, but I can't say for what. I'm open to the suggestion that it's derivative from some other norms (for example, Brandom's, though as Branden correctly pointed out, Brandom doesn't have truth as a requirement). I also still think that these two options needn't be exhaustive.)

[Compare the 'passive offside' rule: It's derived from the 'offside rule' but it's constitutive of the state of being in passive offside.]

Even though I can't say what kind of norm the Self-Knowledge Rule is, I can say why there would be a norm that requires descent: When we descend, we preserve our first-personal authority. – That brings me to the second main claim.

b) Transparency extends to Knowledge and Self-Knowledge.

Transparency explains why there would be a norm that requires us to *descend*. It doesn't say why we should endorse the Self-Knowledge Rule rather than a number of other possible rules, but it does offer a story about why there should be a norm requiring descent. And that's all we need for the doubly externalist skepticism.

My proposal is this: Transparency, in Richard Moran's sense, governs deliberation on knowledge and on self-knowledge. [Accepting Moran's account of transparency is compatible with rejecting transparency in TW's sense.] Here is how Moran characterizes transparency:

“The relation of transparency...concerns a claim about *how* a set of questions is to be answered, what sorts of reasons are to be taken as relevant. The claim, then, is that a first-person present-tense question about one's belief is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world.” (2001, 62)

I propose that Moran's claim should extend to questions about knowledge and questions about self-knowledge.

To see the appeal of this, think about a situation in which you're deliberating about a certain question. You genuinely want to get an answer. In that situation, the questions, “Is *p* the case?” “Do I believe *p*?” “Do I know *p*?” and “Do I know whether I know *p*?” won't come apart. If you answer the first one affirmatively, you should answer the other ones affirmatively, too, provided you answer them at all.

What if you don't? What if you violate the transparency norm? (It is, after all, possible to violate the norm.) —Then you are estranged. You lack first-personal authority over your mental states. For example, you might say, "I know p but I don't know whether I know p ." (I know that I have a hand, but I don't know whether I know I have a hand.) (Assuming $E=K$, this would amount to saying that one is in the situation described initially and illustrated through Branden's analogy.) – What's wrong with you in this case? Following Moran, I would like to say this: You lack authority over your first-order knowledge, because you answer the question about knowledge purely as a psychological question about yourself. Here is how Moran puts the point:

"The vehicle of transparency...lies in the requirement that I address myself to the question of my state of mind in a *deliberative* spirit, deciding and declaring myself on the matter, and not confront the question as a purely psychological one about the beliefs of someone who happens also to be me." (2001, 62-63)

Worry: I am not totally clear on what I want to say here. Is one estranged when one lacks higher-order knowledge but possesses first-order knowledge, or is one estranged when one knows or thinks that one lacks higher-order knowledge? Or is one estranged when one thinks that one lacks higher-order knowledge but possesses first-order knowledge? (The latter is not something one could know.)

Since the issues about Moran don't concern the topics of the seminar, I would like to leave them at this. I turn now to the main arguments for the Self-Knowledge Rule.

3. My Main Arguments for the Self-Knowledge Rule

There are three parts to this. First, I seek to piggyback on TW's arguments for the Knowledge Rule. Second, I seek to argue that the Self-Knowledge Rule is preferable to the Knowledge rule. Finally, I address some of the big objections to the Knowledge Rule.

3.1. The Piggyback Arguments

First, Williamson's argument by elimination also works in favor of the Self-Knowledge Rule. If you rule out the Truth Rule, the Warrant Rule, the Belief-Knowledge Rule, and the Reasonable Belief Rule, you will have ruled out competitors of the Self-Knowledge Rule as well. Moreover, Williamson's arguments against these rules are consistent with the Self-Knowledge Rule.

Second, Williamson considers evidence from ordinary epistemic practice that supports the Knowledge Rule. For instance, the Knowledge Rule explains why we challenge an assertion by asking, "How do you know?" or saying, "You don't know that!" Furthermore, it explains why it is impermissible to assert or believe that one will lose the lottery. Yet both of these things are also explained by the Self-Knowledge Rule: Any violation of the Knowledge Rule will necessarily constitute a violation of the Self-

Knowledge Rule, because, necessarily, if one fails to know, one fails to know that one knows. Hence the challenge, “How do you know?” is explained by both rules, and both rules prescribe that it is impermissible to assert or believe that one will lose the lottery.

[By the way, if you treat belief and assertion as interchangeable, you can rule out Matt Weiner’s Gricean story as follows: Since you shouldn’t believe that you will lose the lottery (because that belief exceeds your evidence), and since you shouldn’t take it as a premise in reasoning that your ticket will lose (e.g. you shouldn’t throw away your ticket), you shouldn’t assert that your ticket will lose either. All of these cases are explained by a lack of knowledge. The Gricean story, however, could only explain the impermissibility of assertion, not of belief and reasoning.]

Third, Williamson points out that the Knowledge Rule explains why it is odd to assert or believe the following versions of Moore’s Paradox (2000, 253):

- (1) I don’t believe p yet p .
- (2) I don’t know whether p yet p .
- (3) I don’t know whether I know p , yet I know p .

Asserting or believing any of (1)-(3) necessarily violates the Knowledge Rule, because in each case it is permissible to believe or assert the second conjunct only if the first conjunct is false. This holds straightforwardly in (2) and (3), and it extends to (1) with the further assumption that belief is necessary for knowledge. Yet exactly the same explanation is available in terms of the Self-Knowledge Rule.

3.2 Independent Arguments for the Self-Knowledge Rule

In the paper I offer three arguments. I’ll only discuss the first two here.

First, challenges to one’s assertion that p often demand self-knowledge. Such challenges are not always appropriate, but when they are, one’s failure to meet them requires one to qualify or retract one’s assertion that p , and not merely one’s assertion that one knows p . This is explained by the Self-Knowledge Rule.

Consider as an example the following conversation: I come back from the movies and claim, “There’s a new De Niro movie coming soon.” You ask, “How do you know?” I reply, “I saw the preview!” You then counter, “How do you know it’s De Niro? You’re terrible at recognizing movie stars. You always confuse Robert De Niro with Al Pacino, and the other day you even confused him with Dustin Hoffman!” I reply, “I *read* the credits.” —In this conversation the second, “How do you know...?” is a challenge to my second-order knowledge: The question challenges me to show how I know that by seeing the preview I came to know that a new De Niro movie is coming soon. It would be possible for me to come to know that a new De Niro movie is coming soon by seeing the preview even if I didn’t know that I came to know this: For example, if in fact I read the credits, I would know that a new De Niro movie is coming soon even if I didn’t know that I know it because I didn’t remember whether I read the credits or whether I relied on

my notoriously poor skills of facial recognition. Indeed, such cases are the paradigm counterexamples to the ‘KK-principle’.

In the imagined conversation I meet the challenge to my second-order knowledge by establishing that I came to know in a reliable way: by reading the credits rather than by facial recognition. Yet if I couldn’t establish this, *I would have to retract or qualify my original assertion that a De Niro movie is coming*. —This observation constitutes evidence for the Self-Knowledge Rule: If I fail to know that I know, the Self-Knowledge Rule prescribes that I must retract or qualify my assertion that a De Niro movie is coming, whereas the Knowledge Rule merely prescribes that I must retract or qualify my assertion that I know that a De Niro movie is coming.

My second argument for the Self-Knowledge Rule is that we sometimes deploy this rule in ordinary deliberation: A failure of self-knowledge moves us to suspend our first-order belief. Consider the following example: Jurors are reexamining their beliefs about a crime. They realize that all their beliefs stand and fall with the testimony of one witness, and they conclude on independent grounds that they do not know whether the testimony is reliable. They will suspend their belief in the testimony and conclude that they don’t know whether the defendant committed the crime. —An example of such a situation arises at a late stage in Reginald Rose’s *Twelve Angry Men* (1957). At one point Juror No. 3 grants that the whole case stands and falls with the testimony of the woman who claims to have seen the defendant stab his father. In the course of the deliberation another juror recalls that the woman had marks from glasses on her nose. The jurors establish that it is improbable that she would have been wearing glasses at the time of the stabbing, which happened in the middle of the night, and they conclude that they don’t know whether her testimony is reliable. Ultimately they render the verdict ‘not guilty’.

In this example the marks on the witness’s nose are the independent grounds on which the jurors conclude that they don’t know whether her testimony is reliable. This constitutes a failure of second-order knowledge: If the testimony were reliable and true, because the witness did wear her glasses and did see the murder, the jurors would know that the defendant committed the crime even though they would fail to know that they know it. After all, in order to come to know something by testimony, one does not have to know that the testimony is reliable. Yet if all that the jurors have to go on is the testimony, and if they don’t know whether the testimony is reliable, they are required to suspend their belief in its truth. This is exactly what the Self-Knowledge Rule prescribes: The jurors’ failure of self-knowledge must lead them to suspend belief in the defendant’s guilt. In the course of suspending their beliefs they might *lose* their first-order knowledge. Hence their descent from a failure of higher-order knowledge to a failure of lower-order knowledge might actually unravel their epistemic achievements.

By the way, Sherri invokes something like the Self-Knowledge Rule in her discussion of ‘dumb luck’ and ‘blind luck’ cases of knowledge (2005, 102-107, esp. 107). She holds that when one knows by dumb luck, one doesn’t know that one knows and because of that one will, upon reflection, give up one’s first-order belief, thus losing one’s first-order knowledge (105). When one knows by blind luck, one *ought*, upon reflection, to give up

one's first-order beliefs for the same reasons (107). – However, Sherri doesn't think that there are any such general norms for belief or assertion.

Worry: Couldn't such unraveling be explained in terms of a norm that is derivative from the Knowledge Rule? In that case, the Self-Knowledge Rule would be superfluous, and 'descent' would be underwritten by the Knowledge Rule. (Yet if that is so, then I can use TW's own account to argue for skepticism. I'd revel in the perverseness of that.)

How would this go? TW formulates the Evidential Rule deriving from the Truth Rule as follows (thanks to Fabrizio here):

One should not: [(assert or believe p) and lack evidence that p is true.]

By parity, a derivative Evidential Rule deriving from the Knowledge Rule would be:

One should not: [(assert or believe p) and lack evidence that p is known.]

(Compare DeRose's (2002) and Weiner's discussion of violations of secondary rules.)

I am not sure that this is sufficient to underwrite descent. For, when the skeptic concludes that we don't know whether we know the things we ordinarily take ourselves to know, he does not say that we lack evidence that we know things. We just don't know what our evidence amounts to. – But maybe this will do after all?

In recent work, Williamson considers the possibility of such unraveling but deems it illegitimate. He writes,

For creatures who can know without being in a position to know that they know, or who can know that they know without knowing that they know that they know, the capacity in principle to withstand all...interrogations is simply not a good test for their knowledge: as often, the rules of dialectic are a poor guide to epistemology. High-level failure does not legitimately unravel low-level success (2005, 234).

I agree with Williamson that interrogation is not a good test for whether one in fact knows: After all, I acknowledge that interrogation sometimes unravels lower-level epistemic success. However, I do not agree that such unraveling is *illegitimate*.

Does Williamson have the resources to explain the illegitimacy of such unraveling? The Knowledge Rule does not explain why unraveling would be illegitimate: The Knowledge Rule does not state that knowledge is sufficient for permissible belief; it only states that knowledge is necessary. If Williamson can explain the illegitimacy of unraveling, then it is only in virtue of his claim that knowledge warrants belief. Thus unraveling is illegitimate because if one knows p , one has warrant to believe p . — However, I think that this element of Williamson's account is problematic. For, Williamson acknowledges that unraveling is part of conversational practice: After all, interrogation is part of conversational practice. Thus interrogation constitutes evidence *against* the claim that knowledge warrants belief. The practice of interrogation provides better support for the Self-Knowledge Rule than the Knowledge Rule.

Worry: Is there more to the Knowledge *Account* than the Knowledge Rule? It's not clear whether Williamson's slogan, "Only knowledge warrants assertion" is supposed to imply that knowledge warrants assertion (2000, 243). (Compare: "Knowledge and only knowledge warrants assertion.") Yet in responding to objections to the Knowledge Account, Williamson does explicitly claim that knowledge warrants assertion (2000, 257). Also, he writes, "If an assertion satisfies the [constitutive] rule, whatever derivative norm it violates, it is correct in a salient sense" (2000, 241). Presumably the point carries over to belief. – Does here the claim to be giving a constitutive rule play a bigger role than it initially seemed? – In any case, the Knowledge Account seems less plausible to me than the Knowledge Rule.

3.3 Replies to Objections

Here I'll try my replies to the two objections that I take to be the most threatening.

First, one might object that my arguments for the Self-Knowledge Rule could be used to support the even more demanding KKK Rule: "One must: believe p only if one knows that one knows that one knows p ." I concede that I do not have a fully satisfactory answer to this objection: I do not know why we should prefer the Self-Knowledge Rule to the KKK Rule. However, the KKK Rule is as useful for the skeptic's purposes as the Self-Knowledge Rule. For, a double application of the Self-Knowledge Rule effectively yields the results of the KKK Rule: If I don't know whether I know that I know, then, by the Self-Knowledge Rule, I must not believe that I know. If I adjust my belief accordingly, I will no longer know that I know. At that point the Self-Knowledge Rule requires me to qualify or suspend my first-order belief. Thus both rules require one to descend from a failure of higher-order knowledge to a failure of lower-order knowledge. Hence even if there is no decisive argument for the Self-Knowledge Rule over the KKK Rule, the main point of my paper stands: Skeptical descent is required even if the KK-principle is false.

(By the way, his objection parallels objections to Frankfurt's and Lewis's discussion of higher-order desires. (Frankfurt understands the concept of a person in terms of higher-order desires, and Lewis analyzes the notion of valuing in terms of higher-order desiring.) The objection (given, e.g. by Smith) is that it's arbitrary that their analyses are given in terms of second-order desires rather than higher orders of desires. – I never found this objection compelling, but I don't know what the best way to respond would be. Any ideas?)

Worry: There is a quick and dirty way to skepticism, which I don't like: Obviously we lack knowledge of some higher-order. At the very least we don't have beliefs of all higher orders. And so we can descend to skepticism via the Self-Knowledge Rule. – What's wrong with this argument?

Second, one might object that we never ask the question, "How do you know that you know?" in ordinary conversation. Again I can offer two points in defense: First, the

question “How do you know...?” does double duty in ordinary conversation. Sometimes it is used to ask how one knows p , and sometimes it is used to ask how one knows that one knows p . Similarly the question, “Do you want to ϕ ?” is sometimes used to ask about one’s first-order desires and sometimes to ask about one’s second-order desires. Which question is meant depends on the context. In suitable circumstances the following would be questions about higher-order knowledge: “How do you know that the newspaper report was accurate?” “How do you know that he wasn’t blinded by her beauty?” and “How do you know that you’re not dreaming?” —The Self-Knowledge Rule furthermore explains why we almost never use the awkward phrase, “How do you know that you know...?” We don’t need to distinguish between knowledge and self-knowledge in ordinary conversation, because failures of knowledge and failures of self-knowledge alike require us to give up or qualify our assertions. The import of a successful challenge to one’s knowledge and of a successful challenge to one’s self-knowledge will be the same. That is why we don’t need to formulate challenges to self-knowledge explicitly.

4. Things I worry about.

(1) Could descent be explained in terms of a norm that is derivative from the Knowledge Rule? In that case, the Self-Knowledge Rule would be superfluous.

(2) Is one estranged when one lacks higher-order knowledge but possesses first-order knowledge, or is one estranged when one knows or thinks that one lacks higher-order knowledge? Or is one estranged one thinks that one lacks higher-order knowledge but possesses first-order knowledge? (The latter is not something one could know.)

(3) Is there more to the Knowledge *Account* than the Knowledge Rule? It’s not clear whether Williamson’s slogan, “Only knowledge warrants assertion” is supposed to imply that knowledge warrants assertion (2000, 243). (Compare: “Knowledge and only knowledge warrants assertion.”) Yet in responding to objections to the Knowledge Account, Williamson does explicitly claim that knowledge warrants assertion (2000, 257). Also, he writes, “If an assertion satisfies the [constitutive] rule, whatever derivative norm it violates, it is correct in a salient sense” (2000, 241). Presumably the point carries over to belief. – Does here the claim to be giving a constitutive rule play a bigger role than it initially seemed?

(4) Why not the KKK-Rule?

(5) Quick and dirty skepticism: Obviously we lack knowledge of some higher-order. At the very least we don’t have beliefs of all higher orders. And so we can descend to skepticism via the Self-Knowledge Rule. – What’s wrong with this argument?