

Lecture 24: Defending Externalism

I. Srinivasan's Challenge

Recall the following problem cases for reliabilism (and for externalist theories of justification more generally):

Clairvoyant: Norman's belief *that the President is in New York City* arises as a result of his clairvoyant powers, but he has no evidence for or against the existence of such a power in general and no evidence for or against his having this power.

Brain-in-a-Vat: Jane_{BIV} is a handless brain-in-a-vat being subjected to a compelling electrochemical illusion in which she is a normally embodied person who believes *that she has hands*.

Srinivasan proposes two real-world cases that, she claims, are exactly parallel to these two cases (here I quote Johnson King's summaries of the cases):

Racist Dinner Table: "Nour, a British woman of Arab descent, goes to dinner with a friend's family. She leaves the table [believing *that her friend's father is racist*], though she can point to nothing specific that happened at dinner to support [her belief]. In fact, the friend's father is indeed racist, and Nour was subconsciously picking up on subtle behavioral cues to which she is reliably sensitive."

Domestic Violence: "Rahda lives in a community suffused by patriarchal ideology, which leads her and everyone around her to believe *that she deserves the . . . beatings that she receives from her husband* whenever he judges that she has been insufficiently obedient or caring. She has never doubted this, nor has anybody given her any reason to."

Srinivasan claims that because *Clairvoyant* and *Racist Dinner Table* are analogous, the beliefs in both cases must have the same justificatory status, and that because *Brain-in-a-Vat* and *Domestic Violence* are analogous, the beliefs in those two cases must have the same justificatory status.

Moreover, she insists, our intuitions and epistemic theories yield the following verdicts about these cases:

	<i>intuitive verdict</i>	<i>internalist verdict</i>	<i>externalist verdict</i>
<i>Clairvoyant</i>	unjustified	unjustified	justified
<i>Racist Dinner Table</i>	justified	unjustified	justified
<i>Brain-in-a-Vat</i>	justified	justified	unjustified
<i>Domestic Violence</i>	unjustified	justified	unjustified

Srinivasan draws the following conclusions:

- If her new cases are analogous to the old ones, then everyone must concede that intuition goes wrong about some of these cases.
- So either internalists must explain away our externalist-friendly intuitions about her new cases, or externalists must explain away our internalist-friendly intuitions about the old cases.
- Moreover, her cases "are far less *recherché* than their traditional analogues," so "if any intuitions should be thrown out as less reliable, it should be those we in response to the traditional cases" (p. 425).

This leaves internalists with three options (Srinivasan neglects the third):

- i. Find a difference between the new and the old cases that allows internalists to give different verdicts.
- ii. Explain away our externalist-friendly intuitions about Srinivasan's new cases.
- iii. Dispute her claim that we have externalist-friendly intuitions about her new cases.

II. Explaining Away Our Internalist-Friendly Intuitions (about the Old Cases)?

Srinivasan offers a diagnosis of why our intuitions about her new cases differ from those about the old cases:

Srinivasan's diagnosis: In the new cases, the subjects are operating under conditions of *bad ideology*: “conditions in which pervasively false beliefs have the function of sustaining (and are in turn sustained by) systems of social oppression” (p. 408). In such cases, we focus on “structural” explanations that explain why a belief is justified by averting to a larger system of which that belief is a part. Externalism is well-poised to offer such a structural explanation, but internalism, with its focus on “individualistic” explanations, is not.

But this diagnosis falls short of an explanation of why our internalist-friendly intuitions about the old cases are not to be trusted, which is what we were after.

Srinivasan gestures at such an explanation in the case of our intuitions about *Brain-in-a-Vat*: we are mixing up the (true) claim that Jane_{BIV}'s belief is *excused* with the (false) claim that her belief is *justified*.

(But excuses involve *the removal of a negative status* whereas justifications involve *the bestowal of a positive status*, so is it really so plausible that we are mixing these two up?)

III. Explaining Away Our Externalist-Friendly Intuitions (about the New Cases)?

It is tempting for the internalist to offer the following error theory for our intuitions about the new cases:

the internalist's error theory: Our intuitions about the new cases are unreliable precisely because these cases involve bad ideology. It feels politically unsavory for us to go with the internalist's verdicts about these cases (by judging, for example, that people subjected to racism are unjustified in believing themselves to be so subjected), so we convince ourselves that we in fact side with the externalist on these cases.

Srinivasan objects that we can construct variants of her new cases in which it is stipulated that our subjects are reliably connected to the truth of some moral view that goes against her own leftist political inclinations. Supposedly we are just as inclined to have externalist-friendly intuitions about these cases. (Is that true?)

IV. In Search of a Disanalogy between the Old and the New Cases?

Why Srinivasan thinks the Norman and Nour cases are parallel: “like Norman, Nour has nothing that is introspectively available to her—no experience or phenomenology—that could potentially serve as the grounds for her belief” (p. 403).

(Srinivasan thinks it is important that we interpret Norman as not having “anything like a quasiperceptual vision of the president's being in New York” [p. 416], but I wonder whether this is so. Would we really be comfortable saying that Norman's belief is justified if he had such a vision?)

Two potential disanalogies between the Norman and Nour cases:

- It is explicitly stipulated that Norman has no idea about how his clairvoyance works, no idea that he has such a power, and no idea that such a thing is even possible, but no such stipulations are in place for Nour with regard to her sensitivity to racism.

(But if we add in those stipulations to the Nour case, do your intuitions change?)

- Nour subconsciously detects certain racist-making behaviors on her host's part, but (on Srinivasan's understanding of the case) Norman does not detect (subconsciously or otherwise) various features that make it the case that the President is in NYC.

Srinivasan concedes this last difference but disputes its relevance on the grounds that Nour's detection of these behaviors is *subconscious* and hence not accessible via introspection.

But it is telling that the intuitive case for Nour's belief being justified completely disappears if we stipulate that she does not even subconsciously register these behaviors but instead has mechanisms that directly spit out a belief *that the host is racist* (even when those mechanisms remain reliable/safe).

But what about mentalist internalists, who can accept subconscious mental states as justifiers?

Srinivasan extends her argument against them by considering “Sarah, who shares all of Nour’s subconscious mental states: she subconsciously ‘sees’ and ‘hears’ subtle facial and verbal behaviors. Except that Sarah’s subconscious is not picking up on actual facial and verbal behaviors, but fabricating them: she is subject to a kind of subconscious hallucination . . .” (p. 426).

Supposedly it is intuitive that Sarah’s belief *that the host is racist* is unjustified, but mentalist internalists who hold that Nour’s belief is justified are committed to Sarah’s belief being equally justified.

But Sarah is very close to a brain-in-a-vat variant of Nour, albeit one where the illusion is internally generated. So people with internalist-friendly intuitions about *Brain-in-a-Vat* are unlikely to share Srinivasan’s intuitions about the Sarah case.

Why Srinivasan thinks the Radha and Jane_{BIV} cases are parallel: “Radha’s belief is . . . the product of a convincing, and systematic, patriarchal illusion. . . . Jane_{BIV} is literally envatted, while Radha is, as it were, envatted in patriarchal ideology” (pp. 399, 407).

Johnson King on the strength of this analogy:

“[Srinivasan’s] vague imagery masks a crucial disanalogy. In the brain-in-a-vat case, what drives the intuition that the brain is justified in believing that it has hands is the fact that its experiences are literally indiscriminable from those of a veridical perceiver. . . . But no analogous argument can be made in *Domestic Violence*, because there is no veridical believer whose experiences are indiscriminable from Radha’s. . . . There are no distant possible worlds that are just like Radha’s except that women deserve to be beaten if they are insufficiently obedient or caring” (“Radical Internalism,” p. 9).

Srinivasan in effect replies: there are no *metaphysically possible* worlds of that sort, but there are *conceivable* worlds like that.