Wedgwood 2002, Internalism Explained

Kevin Dorst
kmdorst@mit.edu

24.805, Fall 2024
Knowledge and its Limits and its Limits

I. The Goal

Internalist intuition: Consider two possible worlds, w_1, and w_2. In both worlds, you have exactly the same experiences, apparent memories, and intuitions, and in both worlds you go through exactly the same processes of reasoning, and form exactly the same beliefs. In this case, it seems, exactly the same beliefs are rational in both worlds, and exactly the same beliefs are irrational in both worlds. Now suppose that in w_1, you are bedevilled by an evil demon who ensures that many of your experiences are misleading, with the result that many of the beliefs that you hold in w_1, are false. In w_2, on the other hand, almost all your experiences are veridical, with the result that almost all the beliefs that you hold in w_2 are true. Intuitively, this makes no difference at all. Exactly the same beliefs are rational and irrational in both.

When we are assessing a state for its rationality, we're assessing it "on the basis of its relation to the agent's beliefs, desires, and other such mental states—not on the basis of its relation to facts about the external world that could vary while those mental states remain unchanged"

In contrast with other evaluations, like 'correct' or 'advisable' or 'jolly good'

Challenges: (1) **What** are 'internal' states? And
(2) **Why** should rationality supervene on them?

Aside on broad content.

II. Bog-Standard internalism

***Access* Internalism:** Rationality is determined by facts that you can know by reflection alone.

Reflection = a priori [?] reasoning and introspective awareness of your mental states

Being rational is being cognitively blameless—and you can't be blamed for not responding to a fact that you weren't in a position to know.

Wedgwood's objections:

1) Justifications vs excuses. Rationality is former.
2) Why restrict to 'by reflection alone'?
3) Anti-luminosity: there's *no* nontrivial condition C such that we're always in a position to know whether C obtains. (Hang tight.)

Q: Example (self-defense vs insanity) seems to elide reasons-based and causal explanations.
Q: Does j/e difference remain under uniqueness? If there's only one thing I'm excused to do, surely I'm justified.

III. Following basic rules

Rationality consists in following a certain set of rules.

Some rules—like the truth rule—we (try to) follow by means of following other rules.

You follow R1 'by means of' following R2 if we can analyze your following R1 into sub-processes *at the folk-psychological level*.

Must be (1) mental and (2) person-level explanations. (No brain modules.) And must make intuitive sense of the behavior.

eg follow rule 'when the water boils, add salt' by following the rule 'when you believe the water boils, add salt'.

NOT: 'when retinotopic images $i_1, i_2, ...i_n$ are processed in V1, add salt.'

If there are no intervening steps we can capture are the folk-psychological level, then the explanation is **fully articulated**.

There must be some rules we follow directly—basic rules.

Analogy: basic actions

Proposal: the rules of rationality are the basic rules that it 'makes sense' to follow.

IV. Basic rules use internal facts ['What?']

Internal facts = those that supervene on your *non-factive mental states*[1], as well as explanatory relations between such facts.

[1] Unlike knowing, seeing, remembering; like believing, it appearing that, and quasi-remembering

Brain states are not internal! Multiple realizability

Claim: a fully-articulated folk-psych explanation will always identify an internal fact as the proximate cause of why you hold/revise a belief.

Arguments:

1) *Open Questions:* If I say 'I'm from Missouri. So Jack believes that', that invites follow-up questions.

Vs(?): Jack's experiences suggest that I'm from Missouri. So he believes that.
Q: What about: 'I told Jack that I'm from Missouri. So he believes that.

2) *Screening Off:* Suppose I say Jack believes I'm from the midwest because he knows I'm from Missouri.
   This fails a **counterfactual test:** If Jack had merely believed (and not known) I was from Missouri, he'd still believe I'm from the midwest.

Non-factive mental states fail *proportionality* criterion. Elephant on a scale.
Q: Is that right?

Empirical question? Science can teach us all sorts of weird facts about what *causes* our beliefs (nudges, fallacies, and the like)—but those won't be folk-psychological explanations.

Q: Is that right? Why couldn't we get both (1) mental and (2) person-level explanations that didn't rationalize; eg people commit the conjunction fallacy because they conjure an image of Linda in their head.

V. Making sense is internal ['Why?']

Suppose you were just programmed by a neuroscientist to follow a given set of rules R. Even if the neuroscientist chose R because they wanted you to be rational, and those rules do indeed 'make sense' to conform to, *still* you are not 'genuinely following' the rule—at least not *directly*. (You're merely *conforming* to it, non-accidentally.)

'The link between this fact and your conforming to the rule... is mediated by the intervention of the neuroscientist'

In order for your belief revision to *directly* follow a rule that it makes sense to conform to, 'the fact that it makes sense for one to conform to the rule must be part of the proximate explanation of the belief revision in question' (368).
→ But only *internal* facts can play that role.

Because(?) *Open Questions* and *Screening Off*.

Short story: internalism is true because there must be some rules you follow directly, not as means to following other rules—and the only ones that make sense for this are those that have non-factive mental states as their conditions.