Responses to the Kripkenstein Paradox

Millikan's Solution

Millikan agrees that the (attempted) solutions that Kripke criticizes won't work. However, she thinks there's another solution that will succeed.

Rather than following Kripke in offering a skeptical solution, she offers a new "straight" solution to the paradox. The main idea behind her approach is that we can ground facts about both rule-following and meaning in an organism's evolutionary history. On this approach, natural selection is what confers determinate content on thought and – by extension – language.

Millikan's main proposal:

> **Rule Following and Purposes:** For an individual *x* to mean to follow a rule *R* is for one of *x*'s purposes to be following *R*.

Millikan says that while some purposes are explicitly represented by the subject, others are more basic – "root" purposes are not explicitly represented by the subject.

Three ways of conforming to a rule:

(i) One's behavior "merely coincides" with a rule (as when our arithmetical behavior coincides with quaddition)
(ii) Purposefully following an explicit or expressed rule
(iii) Purposefully conforming to an implicit or unexpressed rule

Millikan argues that the unexpressed purposes that lie behind acts of explicit purposing are biological purposes, and one's competence to purposefully conform to an unexpressed rule is a biological competence.

A function or purpose of X is to Z iff:

(i) X has a long line of ancestors that Z'd
(ii) It was thanks to their Z-ing that they survived to pass on the Z-ing trait to X

For example, the function of the heart is to pump blood because (i) earlier hearts pumped blood, (ii) present hearts wouldn't exist if earlier hearts had not succeeded in pumping blood.

One of Millikan's primary examples of purposive rule-following comes from the behavior of hoverflies. Male hoverflies will fly at a specific angle to intercept potential mates, though this response is often trigged by lots of other things (pebbles, birds, other bugs).

> **Proximal Hoverfly Rule:** Fly 180 degrees away from the target minus 1/10 the angular velocity of its image across the retina.
> **Distal Hoverfly Rule:** If you see a female hoverfly, catch it.

Conforming to the **Proximal Hoverfly Rule** enables the hoverfly to conform to the **Distal Hoverfly Rule**, which enables the continuation of the species.[1]

[1] Proximal rules depend on incoming sensory information; distal rules depend on the actual state of the environment.

But can't we come up with a "bent" or "gruesome" rule that's consistent with the hoverfly's behavior?

> **Quoverfly Rule:** Fly 180 degrees away from the target minus 1/10 the angular velocity of its image across the retina, unless the target's angular velocity is in range *R*, in which case fly to the moon.
>
> (where *R* is some very small and never previously instantiated range)

Millikan says that while the Quoverfly Rule is consistent with hoverfly's behavior, it is not part of the hoverfly's biological purpose to follow the Quoverfly Rule, for it is not *because* the hoverfly's ancestors conformed to the Quoverfly Rule that they caught mates and proliferated.

Why not? Millikan writes:

> "In saying that, I don't have any particular theory of the nature of explanation up my sleeve. But surely on any reasonable account, a complexity that can simply be dropped from the explanans without affecting the tightness of the relation of the explanans to explanandum is not a functioning part of the explanation."

Examples:

> "My coat does not keep me warm because it is fur-lined and red, nor because it is fur-lined in the winter, but just because it is fur-lined."

Millikan notes that she is taking on board a metaphysical distinction between natural properties and artificial/gruesome properties.

Applying the account to addition/quaddition:

> "[U]nless doing arithmetic results from a total breakdown of the cognitive systems (in which case there may be nothing your purpose when you encounter 'plus'…) then whatever you mean to do when you encounter 'plus', that content has been determined by your experiences coupled with evolutionary design. But reasonably, whatever you mean by 'plus' is the same as what other people mean who are endowed with the same general sort of cognitive equipment and have been exposed to the same sort of training in arithmetic. This meaning has been determined by the application of *Homo sapiens* rules of some kind to experience. It is likely that these are extremely abstract general purpose *Homo sapiens* rules, in accordance with which human concept formation takes place…"

Swampman

For Millikan, the fact that my thoughts – and by extension, my words – have the meanings that they do is attributable to facts about my evolutionary history. Some have argued that this yield counterintuitive results in cases like the following:

> Lightning strikes a swamp. This causes the swamp to bubble, sending its molecules in random motion. By some extraordinarily improbable fluke, these molecules coalesce into a person – or thing? – that is physically indistinguishable from you. This person [/thing] talks like you, behaves like you, etc.

Question: Does your swamp-doppelganger succeed in meaning anything by their terms? Do they manage to talk (or think) about things?

- Some have thought it is obvious the answer is "Yes" - after all, they would answer questions in the exact same ways as you, and carry on (seemingly) intelligible conversations.
- But Millikan's theory seems to predict the answer is "No" – after all, your swamp doppelganger lacks any evolutionary history, and hence lacks any biological purposes (in Millikan's sense).

Whether this is a genuinely counterintuitive result or not has been the subject of considerable debate.

Lewis' Solution

We saw above that Millikan relies on a distinction between natural properties and artificial/gruesome properties. Lewis relies on a similar (perhaps the very same?) distinction to respond to the Kripkenstein Paradox. However, Lewis does not combine this with any account of biological functioning, or any story on which evolutionary history plays a prominent role. Rather, Lewis tries to solve the problem directly in terms of the distinction between natural and unnatural properties.

For Lewis, the natural properties are those that are well-suited to do various metaphysical jobs, in particular:

1) To "ground objective resemblance."
2) To determine the "causal powers of things."
3) To provide a "minimal basis for characterizing the world completely."

(*Exercise:* Go through these jobs one by one and try to get a sense of what Lewis means. Is the property of *being green* better-suited to performing these jobs than the property of *being grue*?)

For Lewis, naturalness comes in degrees – some properties are better-suited to performing these jobs than others; the former will be comparatively more natural than the latter.

Equipped with this distinction, Lewis suggests that natural properties impose an important constraint on the *meanings of our terms*. Here's the idea. Suppose I use a particular term *t*. Suppose that all my actual uses of the term *t* (and maybe all of my dispositions to use *t*??) are consistent with both of the following hypotheses:

H1) I mean some property *X* by *t*.
H2) I mean some other property *Y* by *t*.

Which of these two hypotheses is correct will depend on which of the two properties – *X* or *Y* – is more natural than the other. If X is more natural than Y, then this is a reason for preferring H1 to H2. If, by contrast, Y is more natural than X, then this is a reason for preferring H2 to H1.

One way of putting this point: the more natural a property is, the more eligible that property is to be the referent of our terms. Another way of putting it: natural properties are "reference magnets" – they are things are terms tend to refer to, absent some reason to think the terms refer to something else.

Applied to addition/quaddition:

> "The property of adding is not perfectly natural, of course, not on a par with unit charge or sphericality. And the property of quadding is not perfectly unnatural. But quadding is

> worse by a disjunction. So quaddition is to that extent less of a way to go on doing the same, and therefore it is to that extent less of an eligible thing to intend to do.
>
> "It's not that you couldn't possible intend to quad. You could… But you have to go out of your way. Adding and quadding aren't on a par. To intend to add, you need only have states that would fit either interpretation and leave it to charity to decree that you have the more eligible intention. To intend to quad, you must say or think something that creates difficulties of fit for the more eligible intention and therefore defeats the presumption in its favor."

Question to consider: Is this plausible? If Lewis is right, why is it that more natural terms tend to be reference magnets?

Merino's Criticism of Lewis

Merino criticizes Lewis' account on the ground that any solution to the Kripkenstein paradox should be able to capture the following guidance constraint:

> **Guidance Constraint:** Meaning facts – i.e., facts determining what a subject means by a term – must be constituted by something capable of guiding the subject in applying the term in accordance with what she means by this term.

Merino argues that facts about naturalness are not guiding facts in this sense.

Questions:

i) Is this a compelling objection to Lewis' account?

ii) If Merino's criticism of Lewis is successful, does it generalize to Millikan's account as well?