

Ridgeview Publishing Company

Self-Knowledge & Semantic Luck

Author(s): Stephen Yablo

Source: *Philosophical Issues*, Vol. 9, Concepts (1998), pp. 219-229

Published by: Ridgeview Publishing Company

Stable URL: <http://www.jstor.org/stable/1522973>

Accessed: 10/08/2009 09:27

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=rpc>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.



Ridgeview Publishing Company is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Issues*.

<http://www.jstor.org>

Self-Knowledge & Semantic Luck

Stephen Yablo

1 Paul's Paradox

There is no prize in philosophy for the shortest *a priori* proof of an external world; longest would be more like it. If a prize were to be given, though, it would have to go to the argument just set out:

- (1) If I have water-thoughts, then water exists.
- (2) I have water-thoughts.
- (3) So, water exists.

I mean, of course, that this argument would win the prize if it worked: if, in particular, its premises were *a priori* knowable. Paul maintains that it doesn't work. What we have is rather a *paradox* in which two enormously plausible hypotheses —one backed by privileged access, the other by externalism about content— are found to entail an incredible result, viz. the *a priori* knowability of there being such a thing as water.

2 Saul's Paradox

Before Paul's paradox, there was Saul's. Saul's paradox is in a way more the urgent of the two, because it casts doubt on the compatibility, not of calling (1) and (2) *a priori*, but of calling them (or statements very like them) *true*. Here is a passage from *Naming and Necessity*:

[...] it is said that though we have all found out that there are no unicorns, ... [u]nder certain circumstances there *would* have been unicorns. And this is an example of something I think is not the case... Perhaps according to me the truth should not be put in terms of saying that it is necessary there should be no unicorns, but just that we can't say under what circumstances there would have been unicorns.¹

Notice that *two* things are "said": first, that there aren't any unicorns, and second, that under certain circumstances there would have been unicorns. It is only the second claim that Kripke disputes. He takes it for granted that unicorns don't exist; this indeed is why "unicorn"-sentences lack truth conditions, and why unicorn-thoughts—which presumably *need* truth conditions to exist—are ruled out entirely.

Now, suppose it is really true that for there to be *X*-thoughts, there have to be *X*s.² Then what is to prevent a philosophically reflective 18th century thinker—what for that matter is to prevent you or I—from arguing like so:

(1') If I have caloric-thoughts, then there is caloric.

(2') I have caloric-thoughts.

(3') So, there is caloric.

Since the conclusion is false—there is no caloric—one of the premises must be false as well. Which one? I call this a paradox because one has, as Kripke likes to say, a considerable feeling that *both* premises properly understood are correct.

Having just asserted that there is no caloric, the second premise will be hard for me to deny; the thought that there is no caloric is a caloric-thought if anything is. And I have other caloric-thoughts

¹ *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980), 24.

² A restriction is obviously needed on "*X*"; I won't attempt to formulate it here.

as well: people used to believe in caloric; if there is caloric, then so much the worse for the atomic theory of heat.

But the first premise is hard to argue with too. A thought is something with truth conditions; it can be compared to a possible world and found to be either true of that world or false of it. And in the absence of any real caloric to set the standard, there seems to be just no saying which of the caloric-like substances in other possible worlds are relevant to the counterfactual truth-values of my caloric-thoughts.

3 Kripkean *vs.* Fregean Truth Conditions

The Kripke passage paints a picture of unicorn-thoughts as true without possessing truth conditions. How is this possible? A thought is true, it would seem, only if it puts conditions on reality, which conditions are in fact met. These conditions though could hardly be other than the thought's truth conditions. And wasn't it truth conditions that were supposed to be going missing on the Kripkean picture?

An analogy may help. Suppose I introduce *H* as a predicate that, no matter how the world may turn out, is to be true of all and only the hedgehogs. Now consider the statement, or thought, that there are *H*s. Are you in a position to compute its truth conditions?

Not in the sense at issue in the Kripke passage, for I haven't told you anything about what it takes for *H* to be true of a *counterfactual* object. (And I'm not going to; the whole meaning of *H* has now been explained.) And yet there would seem to be little doubt that the thought is true. Because whatever *there are Hs* says or doesn't say about other worlds, *this* world it describes as containing hedgehogs. Here then is a case of a truth that is lacking in truth conditions.

But wait a minute, you say. Of course the thought that there are *H*s has truth conditions; it is true under the condition that hedgehogs exist.

There is clearly something to this. The thought has truth conditions *of a sort*: its (actual) truth-value depends on whether such and such conditions are (actually) met. It remains, though, apparently, that the thought lacks truth conditions of the kind intended by Kripke: conditions that determine whether a (possibly counterfactual) world is correctly *described* by the thought. To have some language for this unusual state of affairs, let's say that *there are Hs* has "Fregean" truth conditions, but little or nothing in the way of "Kripkean" ones.

4 Consequences for the Paradoxes

What does this tell us about Saul's paradox? If "having a thought to the effect that there is caloric" means "having a thought with the *Kripkean*, or modalized, truth conditions that there is caloric", then (1') looks true; without actual caloric to set the standard, how can there be a fact of the matter about caloric's counterfactual career?³ But (2') is false. I am *not* thinking that there is caloric because my "thought" leaves it (more or less) wide open which worlds are caloric-worlds.

If on the other hand we are talking about "having a thought with the *Fregean*, or actualized, truth conditions that there is caloric", then matters are reversed. I am indeed thinking that there is caloric—so (2') is true—but, (1') to the contrary, my ability to do this is not hostage to the real existence of the stuff.

Will the same approach work with Paul's paradox? Take first Kripkean water-thoughts, or water-thoughts individuated by their Kripkean truth conditions. *A priori* reflection suggests that these require actual water, as stipulated in (1). But that I am thinking a Kripkean water-thought is, it may be argued, not something I can tell *a priori*. So (1) is *a priori* but (2) is not.

If we switch now to the Fregean truth conditions of my thought—the way its truth-value in the actual world depends on the actual facts—this does seem to be *a priori* detectable. Even the full Fregean *meaning* of my thought—the way its truth-value across *all* worlds⁴ depends on the actual facts—is *a priori* knowable. But so what? Thoughts individuated by their Fregean truth conditions, or meanings, don't call *a priori* for actual water. This time then it is only (2) that's *a priori*.

5 Let's Be Kripkean

Terrific; except that all this time we have been walking into a neatly laid incompatibilist trap. Here is what the incompatibilist will say:

I don't care if Fregean thoughts are *a priori* knowable, because Fregean thoughts are not *externalist* in my sense. Your view appears to be that externalism and privileged access can both be true, *but not of the same thoughts*. Why should I disagree?

³I'll take this back in a little while.

⁴Its Kripkean truth conditions, in other words.

This reply forces us to put the Fregean notions aside for a bit, and look again at Paul's claim that Kripkean truth conditions go missing on Dry Earth. (Truth conditions are henceforth Kripkean; the one Fregean notion that will recur is "meaning", and that not until the last section.)

One thing is clear: if this be Dry Earth, then the great majority of worlds are not classifiable either as containing water or lacking it. It's the next step that bothers me. Does it really follow that *there is water* is lacking in truth conditions?

That depends. It probably does follow, if truth conditions are seen as singular propositions made up *inter alia* of full-blooded properties; in the absence of water, there can be no full-blooded property of *being* water, which is curtains for the proposition.

But the singularist conception is a surprising one in a context where truth-conditionality is being treated as a condition of thought. After all, the capacity for water-thought is intuitively quite *independent* of ontological disputes about what sorts of properties there may be, including disputes about whether properties exist at all. I suppose one could say: let's have a pleonastic conception of properties on which the needed properties come for free. But they come for free only when the predicate is suitably meaningful, and the meaningfulness of "water" on Dry Earth is just what we are arguing about.

What sort of object should play the role of truth conditions, if not a full-blooded singular proposition? The answer is that any object will do that encodes the possibility of interrogating a world on such matters as whether it contains water; that's what it takes for water-thought, hence that's all that can be asked of truth conditions considered as a requirement of such thought.⁵ For the sake of definiteness, let me suggest that the encoding role is most directly and efficiently played by (truth conditions conceived as) rules or recipes for classifying worlds. Singular propositions can serve in some cases as handy repositories of classificatory information. But if we're talking about truth conditions in the sense essential to thought, it's the rule that matters.

6 Degrees of Taxonomic Power

Now rules, as we know, are apt to have blind spots: sometimes big ones. It is not much remarked on, but the rules defining "true" leave

⁵Up to and including singular-proposition-like objects adapted to avoid Paul's worry about the lack of a property.

us hanging as often as they deliver a verdict; there are fully as many paradoxes as definite truths and falsehoods. (One can imagine alternative rules that would make the ratio still worse.) That thought about truth and falsity is somehow nevertheless possible suggests a conjecture: “there is water” *does* have truth conditions on Dry Earth,⁶ but truth conditions with not much resolving power as between cases. A few utterly dry worlds are ruled out, but on most worlds they just fail to pronounce.⁷

The truth analogy has its limits, because the sniffing out of paradoxes has a significant *a priori* component. A better analogy is with Carnap’s semantics for theoretical terms in *Testability & Meaning*.⁸ Carnap lays it down that an object immersed in water is soluble iff it dissolves; and that chemical analogues of *solubilia* (*insolubilia*) are themselves soluble (insoluble); and that’s it. As he is quick to point out, this leaves a great many cases completely undecided—for all we can tell *a priori*, it leaves *every* case undecided.

These empirical uncertainties notwithstanding, Carnap has, it seems, infused the word “soluble” with substance enough to allow for solubility-thoughts—and hence with substance enough for truth conditions, to the extent that solubility-thoughts require them. One does not feel the truth conditions of *this is soluble* to be dwindling away into nothing as the number of actual immersals declines. One feels rather that they are pronouncing on fewer and fewer cases, and to that extent falling short of their destiny as truth conditions. But thoughts with underperforming truth conditions are still thoughts.

7 Paradox Redux

You can guess where this is heading. If we are unlucky enough to be living on Dry Earth, then as Paul says, there is no (full-blooded) property of being water. That doesn’t in itself make nonsense of the question “which actual and counterfactual stuffs deserve to be described as water?” It doesn’t rule out that “there is water” puts a condition on worlds, albeit a condition with less taxonomical power

⁶E.g., a world contains water if there is a unique watery stuff on Dry Earth which it sufficiently resembles, and lacks water if it is thoroughly dry. (This is crude.)

⁷The similarly pathetic truth conditions of “there is caloric” rule out different worlds; this is enough to mark the two as distinct. (I assume the physics of heat on Dry Earth is exactly as here.) They are distinct too in being *poised* to rule different worlds *in*, given more favorable empirical conditions.

⁸*Philosophy of Science* (1936/7), Vol. 3, 419–471, and Vol. 4, 1–40.

than might ideally be wanted. (And it does have *some* taxonomical power, for it rules negatively on Dry Earth.)

The claim then is that the compatibilist can simply *deny* that water-thoughts *a priori* require water. But while this may be enough to counter the paradox as presented (specifically the first premise), what about the following variant, where “taxonomical” means “has a good deal of taxonomical or classificatory power”:

(1*) If I have taxonomical water-thoughts, there is water.

(2*) I have taxonomical water-thoughts.

(3) There is water.

I want to say that the revamped paradox just reverses the problem with the original one. It may be true, and true according to externalism, that (1*) is *a priori*.⁹ But privileged access ought *not* to be understood as making (2*) *a priori*. The hypothesis that I’m thinking *taxonomical* water-thoughts is too close to the hypothesis that there is water for one to be *a priori* and the other not.

8 Knowing What

At this point the original question —how am I to tell without empirical research *whether* I’m thinking?— begins to transmogrify itself into a more familiar one: how am I to tell without empirical research *what* I’m thinking?¹⁰ To claim *a priori* knowledge of my thoughts, in particular of their truth conditions, I should at a minimum be able to tell whether these conditions possess any genuine bite.

Such an argument may seem only common sense. But it trades on a very particular conception of “knowing what” —a conception that may itself seem only common sense, but that requires scrutiny. I call it *absolutism* about knowing what:

⁹I have doubts about the *a priority* of (1*) too, because I suspect that “water” could stand on Dry Earth for a superficial phenomenological kind. “Natural kind” terms stand for the most natural kind available; they rise to their own level, in Kripke’s phrase. That some of “air”, “earth”, “fire”, and “water” strike us now as more natural-kindly than the others reflects no preexisting semantic commitments, but just that they are the ones that got lucky. (I grant that there *could* be terms that denote natural kinds or nothing, and so I am not pressing the point.)

¹⁰This is of course the “other” externalist threat to privileged access.

knowing what X is is knowing inherently important facts about what X is, that is, X's identity or essence or nature.

If the truth conditions of my thought are not terribly taxonomical, then that would seem to be an important fact about their nature—the kind of fact that, according to absolutism, someone who knows what the truth conditions are ought to be cognizant of. Since I can't attest *a priori* to their resolving power, it seems that I don't know *a priori* what the truth conditions of *there is water* really are.

Is absolutism correct? A look at a few of its consequences will help us decide. Absolutism entails that (i) facts about X's nature are crucial to knowing what it is; (ii) facts not about X's nature are irrelevant to knowing what it is; and (iii) the same facts are relevant to knowing what X is regardless of how it is described.

I hope I'm not the only one who finds all of this highly debatable. As against (ii), to know what magenta is, you have to know how it makes things look, *even if* magenta is an intrinsic property of external objects with no essential relation to human experience. Similar remarks apply to the north pole (it's on top of the world), LSD (it gets you high), dirt (it's cheap), and the Earth (it's the planet we live on). To go by (i), only a philosopher can tell you what the least prime number is; the mathematicians talk a good game, but they aren't even decided whether two is a set. As for (iii), squareness and diamond-shapedness are the same property, but different recognitional abilities are required to know what they are. Whoever doesn't know what salt is must have been living under a rock, but sodium chloride is a different story. The anhedonic physiologist knows what p-fiber firings are but not pleasure, even if pleasure and p-fiber firings are one and the same.

Of course, the application that interests us is to knowledge of truth conditions. But absolutism seems wrong here too. I know what the truth conditions of Goldbach's conjecture are and yet I am ignorant of as basic a fact as this about them: whether they are necessary or impossible.¹¹ (Think too of theoretical identities in science, genealogical conjectures, etc.) If I can get by without *a priori* knowledge of one "basic metaphysical fact" about truth conditions, their satisfiability, why not another, the fact of how taxonomical they are?

¹¹Suppose some mathematician tells me they are impossible. It is far from clear that this in itself helps me to understand the conjecture any better.

9 A Non-Absolutist Alternative?

After all this shirking of epistemic obligations, someone might ask: what *do* I have to know to know the truth conditions of my thought? As an alternative to the “metaphysical” conception just scouted, suppose we try the following:

- (*) I know *a priori* what the truth conditions of my thought are iff I know *a priori* that it has the truth conditions that *P* —for *P* an appropriate (!?) sentence of my language.¹²

This leaves the incompatibilist one final opening. To know *a priori* that my thought has the truth conditions that *P*, I need to know *a priori* that it *doesn't* have the (alternative) truth conditions that *Q*. And if externalism is correct, then for some values of *Q*, I don't. E.g., for all I can tell *a priori*, the thought I express with the words “there is water” might be true under the condition that there is XYZ.¹³

Here is how I would *like* to respond; the details are still under construction. *A priori* knowledge resembles ordinary knowledge in an important respect: it requires us to rule out some counterpossibilities but not all the counterpossibilities there are. This is clear from consideration of simple examples. I know *a priori* that my location is here, but I *don't* know *a priori* that my location is not Kinshasa, despite the fact that being here (in Ann Arbor) is strictly incompatible with being in Kinshasa. I know *a priori* that Kabila is Kabila, but I don't know *a priori* that Kabila isn't Mugabe, despite the fact that being Kabila is incompatible with being Mugabe.

And now a speculation, offered in the spirit of something that would be neat if true: the alternatives I have to rule out *a priori*, to know *a priori* that I am thinking the Kripkean thought that there is water, are ones that I *can* rule out *a priori* just by virtue of my *a priori* grasp of Fregean meanings (as described at the end of section IV).

To see why this is not completely insane, suppose we ask why I *don't* have to know *a priori* about not being in Kinshasa to know *a priori* about being here.

Answer: for all I can tell *a priori*, “I am here” and “I am in Kinshasa” have identical Kripkean truth conditions; whence for all

¹²This is ignoring expressibility worries, which are legitimate but not to the present point.

¹³This could be questioned, since “XYZ” is explicitly introduced as standing for a substance distinct from the actual watery stuff, viz. water. But let that pass.

I can tell *a priori*, “my thought has the truth conditions that I am here” and “my thought has the truth conditions that I am in Kinshasa” have *compatible* Kripkean truth conditions.

Obviously, though, I cannot be required to rule out *a priori* scenarios that are, for all I can tell *a priori*, *compatible* with what I think, as a condition of that thought’s constituting *a priori* knowledge. This would empty the category of *a priori* knowledge altogether; even logical truths like *Kabila = Kabila* would lose their *a priori* status, since I cannot rule out *a priori* that *Kabila = Mugabe*. I conclude that

- (**) To know *a priori* that *A*, I have to know *a priori* that not *B* —but only in cases where *B* is *a priori* incompatible with *A*.

Applied to knowledge of truth conditions, this means that

- (***) To know *a priori* that my thought has the truth conditions that *P*, I have to know *a priori* that it does not have the truth conditions that *Q*— except in cases where it is *a priori* possible (i.e., not *a priori* false) that *P* and *Q* have the *same* truth conditions.

The connection between (**) and (***) is simply this. The cases where it is *a priori* possible that *P* and *Q* agree in their truth conditions are all and only the ones where *B = “my thought has the truth conditions that Q”* fails to be *a priori* incompatible with *A = “it has the truth conditions that P”*.

Now, do I have the *a priori* knowledge that (***) requires of me? I see no reason to doubt it. If *P* and *Q* are *a priori* different in their truth conditions, it will be *a priori* too that no *P*-thought has the truth conditions that *Q*. To infer *a priori* that *my* thought doesn’t have the truth conditions that *Q*, I will need to know *a priori* that *my* thought is a *P*-thought. But that is the commonsense view; it is the incompatibilist’s job to undermine it, not mine to shore it up.

As a matter of fact, though, it can be shored up, if we allow ourselves an assumption to which incompatibilists are not *per se* opposed. The assumption is that I have *a priori* knowledge of *Fregean meaning*: of how the truth conditions of my thoughts and sentences depend on the actual-world facts.¹⁴ Why should incompatibilists deny this? Fregean meaning is intrinsic and their problem is about *extrinsic* content.

¹⁴See the end of section IV.

Suppose then that I am granted a complete *a priori* grasp of the Fregean meanings of my thoughts, and of relevant sentences of my language. This tells me, for each *Q* meeting the proviso of (***), that however the actual world comes out, my thought never acquires the truth conditions that *Q*. From this I conclude *a priori* that I am not thinking that *Q*. Any block raised by (***) to my knowing *a priori* that I am thinking that *P* is thus removed.

So much is to defend against an objection. But it seems possible to make a *positive a priori* case that I am thinking that *P* by the same method: “noticing” that my thought comes out with the truth conditions that *P* on every hypothesis about actuality. By (*), nothing more is required for *a priori* knowledge of what it is that I am thinking. The apparent result is that a “complete” *a priori* grasp of my Fregean thoughts¹⁵ provides me with “ordinary” *a priori* knowledge of my Kripkean ones.

¹⁵ And the Fregean meanings of relevant sentences.