Beyond Rigidification: The Importance of Being Actually Actual

Stephen Yablo

May 2003

$^{1}/_{3}$ draft

Rereading Naming & Necessity in the light of later developments, it can seem that Kripke was not playing fair in his critique of Frege's sense theory. The sense theory for our purposes says that with each name is associated a bunch of properties. The name is linked to the properties in three ways.

modally	being n goes **necessarily** with having the properties
epistemically	being n goes **apriori** with having the properties
conceptually	being n goes **in understanding** with having the properties

Each of the links leads us to expect a phenomenon that turns out not to obtain:

modal	Water is possible without hydrogen
epistemic	Cats are as an apriori matter furry purring animals.
conceptual	Nothing counts as Peano unless it discovered PA.

Because of these false predictions, the sense theory is rejected. The thing is, Kripke then immediately turns around and points out other phenomena, also predicted by the sense theory, that are in his view there.

modal	A related possibility makes water *seem* possible w/out H.
epistemic	A meter is as an a priori matter the length of stick S.
conceptual	Nothing counts as heat unless it feels a certain way.

Not only are are these predictions correct by Kripke's lights, his own account of them, in terms of reference-fixing descriptions, look a lot like the rejected explanation in terms of sense. This could make a person suspicious. Perhaps Kripke's radical-seeming conclusions are a function less of his evidence than the order of presentation. A more logical approach would be to first use the sense theory's true predictions to motivate the theory, then bring in its false ones as a guide to the theory's proper development. Senses should be chosen with an eye to the importance of not falling into these particular traps.

This way lies the 2-dimensionalist reinterpretation of Kripke, first convincingly elaborated in Davies and Humberstone's "Two Notions of Necessity." Names are equivalent in meaning to definite descriptions, only not the ones we'd supposed. Sometimes the false predictions reflect just a bad choice of associated properties, 'Cats' means something more like 'whatever shares deep explanatory features with these things'. Peano' means something more like 'whoever the people I learned the word from were talking about'. This is what's going on with the epistemic and conceptual problems.

Other times however (this is the interesting case) the false prediction shows we have misjudged the character of the association. Being Hesperus goes with evening-visibility not across all counterfactual worlds but all worlds "considered as candidates for actuality" – all counteractual worlds for short. This is addressed by switching to 'the actual so and so'. This description is rigid in one dimension, since the actual so and so would have been one and the same whatever world had obtained. (That takes care of the modal mis-prediction.) But along another dimension it refers to whatever turns out to have the properties, supposing for argument's sake that the given world is actual.

I have called the 2D semantics a reinterpretation of Kripke. Not everyone sees it this way. The 2D line has taken on such an air of inevitability over the years that it can seem, at times, that what separates Kripke's position from later developments is just that Kripke is more confusing. I find it all the more interesting, then, this is not D&H's attitude at all. They see the 2D view as distinct from Kripke's, and point to an issue about language use that resolved one way supports the 2D view and resolved another way supports Kripke. It is because they are not sure how to resolve this issue they describe themselves as "not confident that the suggested [2D] view is correct" (20).[1] The reason this matters is that if 2-dimensionalism is correct no matter how we talk, then the view is lacking in substantive content. Because D&H are, to my knowledge, the last 2-dimensionalists to associate the view with a potentially falsifiable claim about English, they are giving us here a rare opportunity that we would be wise not to pass up.

What is the issue that D&H are not sure how to resolve? They start by looking at the cases where the 2D account works best: descriptive names a la Evans. Part of what makes 'Julius' a descriptive name, on their reading of Evans, is that

> One can understand sentences containing 'Julius' without knowing of any object that it is being said to be thus and so (7).

It will be hard for the name to retain this feature, they think, if it comes into everyday use.

> Imagine that every speaker of the language…had a visual confrontation with Tom and was told 'This man is Julius.' ….Given the knowledge which each speaker would now have

[1] Later: "it is no part of our position that the suggested view is the ultimately correct view of the way 'red' functions in English" (22).

> (knowledge by acquaintance) of Julius it would be natural for the semantical function of 'Julius' to change" (20).

This is (I suppose) because knowing Julius just as the zip-inventor no longer suffices for understanding. Imagine someone who walks in fresh from an Evans seminar and hears someone shout, "run, Julius has a gun!" She grabs your arm and says, "how are we supposed to know which way to run when we don't know who invented the zip?" Her failure to realize that she is to run from that guy, not just from whoever might have invented the zip, would seem to mark her as having a defective understanding of the word.

> But now consider the fact that practically every speaker of our language has had a visual confrontation with (a sample of) the chemical kind H_20 accompanied by the words 'This stuff (this chemical kind) is water (is called 'water')'. Is it not unlikely that 'water' remains, in our language, a merely descriptive name of H_20? (20).

If water is known to us as this familiar stuff of our acquaintance, then when someone says "water is refreshing" we know that this familiar stuff of our acqaintance is being called refreshing. Suppose a Martian chemist walks in who knows water only by description. She drinks a glass of water and says, "Mmm, that's refreshing. I wonder if water is refreshing?" That might again appear to mark her as not fully understanding the word.

The claim so far is that if water is this familiar stuff of our acquaintance, then it will be hard for 'water' to be a descriptive name. Contraposing, if water is to be a descriptive name, water had better not be known as this familiar stuff of our acquaintance. This is the conclusion that D&H draw. It had better be that

> physical ostension of a sample of H20 accompanied by the words 'this stuff'…is similar not to physical ostension of a

> man accompanied by the words 'this man' but rather to physical ostension of a screen accompanied by the words 'the man behind this screen' (20),

Of course, the analogy here – pointing to water is like pointing to whatever is behind the screen is strained at best – not to mention that it's unclear why pointing to Julius qua that man should be any easier than pointing to water qua that stuff. I take it this is part of the reason D&H are "not sure the suggested view is correct." That having been said, the strained analogy's connection to 2-dimensionalism has still not been quite made out.

Suppose we agree with D&H that 'water' is not a descriptive name if the referent is known as that familiar stuff of our acquaintance, as opposed to whatever lies behind that sheet. This still doesn't tell us why D&H are worried that 'water' doesn't have a 2D semantics. It is not enough to know why 'water' wouldn't satisfy somebody's (Evans's) definition of a descriptive name; we want to know why 'water' stops behaving in the characteristic descriptive-name way when we encounter its referent.

Here is a story that seems of the right general type, drawing on work of Jim Pryor and John Campbell. Judgments can be made by way of other judgments. E.g., I judge that the President is holding a dog by judging that Bush is holding a dog in the belief that Bush is the President. What is special about 'Julius' is that the one and only way to judge that Julius is F is by judging that the zip-inventor is F in the belief that Julius is the zip-inventor. The minute we learn how to go in the other direction, judging that the zip-inventor is F by judging that Julius here is Fin the belief the zip-inventor is Julius, 'Julius' has ceased to be a descriptive name. That of course is what happens when we meet the guy; we see that Julius is drunk and conclude that the zip-inventor is drunk.

Now, why should descriptive-name-hood as just explained have the result that '<u>n</u>' refers on all counteractual hypotheses to whatever is actually so and so, and why loss of descriptive-name-hood interefere with that result? When it comes to deciding whether Julius is so and so, I am doing something that is of its nature done when and because one decides the zip-inventor is so and so. So if you ask me how I decide whether Julius is on the given hypothesis drunk, there is only one possible answer: I decide whether the actual zip-inventor is on the given hypothesis drunk.

Suppose on the other hand that I have other ways of deciding that Julius is so and so. Now there is the possibility of, as we might put it, "original intelligence" about Julius -- I realize Julius is so and so not by first realizing something else – and/or "variously derivative intelligence" about Julius – intelligence based on other descriptions Julius is thought to satisfy. Now I cannot rest my decision purely and simply on the issue of whether the zip-inventor is on the given hypothesis drunk. On the one hand here is a guy with the same parents as Julius (=Tom), who looks and acts just like Julius and leads a very similar life, and he is drunk. On the other hand here is the guy who invented the zip on the given hypothesis and he is not drunk at all. Is Julius drunk on the given hypothesis or not? It as not as though I have rank-ordered Julius's traits so as to know which way to jump when and if the traits come apart. Unless our grasp of a name has to takes a very particular form, ruling out independent and potentially clashing sources of information about the referent, judgments about counteractual worlds are bound to be problematic.

It makes sense then that the 2-dimensionality of a term should stand or fall with the uniformly <u>derivative</u> character of judgments about the term's referent; they are always reached via the same descriptive route. How though is it to be determined whether judgments about water, heat, and so on are uniformly derivative in that way?

One could try, I suppose, to argue directly that, as noted above, speaking of water feels very different from speaking about an I-know-not-what hidden behind some veil of appearances. But the argument for a 2-dimensional interpretation was never that it feels right. The argument was that it explains a lot and on a more economical basis. This brings us back round to to the modal, epistemic, and conceptual phenomena with which we began. The rigidifier says: look, I can explain these phenomena just as well as Kripke and with a lot less fuss and bother.

Our job as advocates for D&H's skeptical side is to see if this explanatory advantage claim is true. A case can be made that the rigidifier's explanation is oftentimes WORSE than Kripke's, and worse in ways plausibly blameable on the treatment of water-judgments as of their nature done by way of descriptive judgments. There will be three kinds of criticism: (1) You are using a cannon to shoot a mouse; (2) You are hitting a lot besides the mouse. (3) You have missed the mouse .

It's the "mouse is missed" criticism that's the most interesting so let me say briefly why that happens. The rigidifier's interpretation of "actually" makes certain sorts of concept inexpressible. One cannot in the 2D framework express concepts whose extension is tied to what is really actually the case, as opposed to what might be hypothesized to be actually the case.

So, to mention an example that will come up later, we have the concept of a "spitting image" or a "look-alike" of something now under observation.[2] This is used in turn to explain other concepts,

---

[2] Another example: compare the concept audible to the concept plausible. Audible is OK because it does not care which world is really actual. Whatever we can hear on a given hypothesis is on that hypothesis audible (so, light waves are audible on the

e.g., by fool's gold we mean (near enough) whatever looks enough like real gold. It seems crucial here that this means looks to us as we are like gold, not as we might be hypothesized to be. What am I worried about, after all, when I worry that this ring might perhaps be made of fool's gold? First answer: the ring might be a substance like iron pyrites that I as I am cannot tell apart from real gold. Second answer: the ring might be a substance like charcoal that looks to me as I am hypothesized to be like gold because (on the worrying hypothesis) gold looks to me a dull black. The first answer is clearly the right one, but lacking a concept of real actuality – of how the ring looks to me, not as I am hypothesized to be, but as I am -- the 2-dimensionalist cannot express what I am worried about.

At least, he cannot express it directly, by invoking the real actual world as such. There remains the option of picking our world by description. He can try to determine empirically that this world is X, Y, and Z, and then write those features explicitly into the concept, e.g., fool's gold is whatever looks like gold to observers with some particular empirically determined sort of brain. But this just trades one difficulty for another, because now he has changed the subject and is explaining the wrong phenomenon.

Two quick examples of "explaining the wrong phenomenon," the first to do with illusions of possibility. It seems possible that gold could have had a different atomic number. That illusion can be explained in 2D terms if by "different" one means "different from 79." (If a certain world w is actual, gold has an atomic number of

---

hypothesis we can hear light waves). However to suppose that something is widely believed does not at all make it on that hypothesis plausible. What is plausible on a given hypothesis is judged by us as we are, not us as we are hypothesized to be. In the unlikely event that we find Scientology plausible; that just shows we are bad judges of plausibility.

80.) But suppose we don't know gold's atomic number and just think it could have had a different one than it does have. This seeming possibility the 2-dimensionalist cannot explain, for it makes no sense to say that if w is actual, then gold has a different atomic number than it has actually. (Compare the dicto reading of "I thought your yacht was longer than it actually is.") The closest he can come is to explain the illusion it could have had a different atomic number than 79. But that is a different illusion. So there is one example of explaining the wrong phenomenon, what I called missing the mouse.

Second example (but feel free to go straight on to the next section): D&H suggest as the 2D meaning of 'x is red' 'x has that physical property which actually standardly causes produces red* sense data in perceivers' (22). A footnote says "we find at fn 71 [of N&N] …a very clear anticipation of the present suggestion for secondary quality words" (28). However footnote 71 speaks not of whatever conditions might count as normal in hypothetical scenarios, but the conditions that really are normal. Kripke gives no criterion C for normality that works no matter which world is actual, and indeed he questions whether such a criterion is possible::

> If one tries to revise the definition of 'yellow' to be, 'tends to produce such and such visual impressions under circumstances C.'… one will find that the specification of the circumstances C either circularly involves yellowness or plainly makes the alleged definition into a scientific discovery rather than a synonymy (140).

Then how is 'normal' understood by Kripke? I conjecture that the Kripkean notion has a heavy demonstrative element; we presume that these conditions – the ones that really right now obtain -- are normal absent a reason to doubt it. The rigidifier lacking a concept of real-actuality cannot follow Kripke in this. He will have to specify normal conditions descriptively and without reliance on the

concept of yellow. He certainly doesn't know how to do this a priori so he will have to undertake an empirical study of prevalent viewing conditions, including the conditions that obtain inside our heads. But this, as Kripke says, "makes the alleged definition into a scientific discovery." If you want it to be a definition, it's a defiition of shmolor concepts not color concepts. The modal, epistemic, and conceptual phenomena as they arise for our concepts will be left unexplained.

Explaining the *Conceptual* Datum

Both sides agree that it can sometimes be important to the understanding of a name 'n' to realize that the referent should have certain properties. But they offer different explanations of this, according to their different views of meaning. (I assume that understanding is in some sense knowing the meaning). The rigidifier maintains that 'n''s meaning is the same as that of 'the actual G, which comes to the fact that 'n' stands no matter which world is actual for whatever is actually G. It would seem then that

> (2DU) Understanding 'n' is knowing that no matter which world is actual, x is n iff x alone is actually G.

Given this, the rest is a slam dunk. Knowing that 'n' stands for the unique G no matter what is certainly sufficient for knowing that a thing should exhibit G if it wants to be n.

Since for Kripke the meaning is just the referent, understanding for him comes (so I assume) to knowing what the name stands for. This might sound like saying that a person understands provided they know of the appropriate x that 'n' stands for x. But that is not Kripke's view. A couple of passages suggest what more might be involved.

> "if someone else detects heat by some sort of instrument, but is unable to feel it, we might want to say, if we like, that the concept of heat is not the same even though the referent is the same" (131),
>
> "a blind man who uses the term 'light', even though he uses it as a rigid designator for the same phenomenon as we, seems to us to have lost a great deal, perhaps enough for us to declare that he has a different concept" (139).

The Martian has a defective or eccentric understanding of "heat," but why? It is not, I think, that the Martian fails to know of any x that 'heat' stands for x. For we can suppose she senses heat some other way; she can see it, let's say, with her telescopic vision. Just as we know of the condition x that we feel that 'heat' stands for x, she knows of the condition x that she is looking at that 'heat' stands for x. But Kripke would still I think say that "her concept of heat is not the same even though the referent is the same" (131). The Martian's problem is not that she fails to know of the correct x that it is the referent of 'heat'.

By one's idea of heat let's mean whatever it is in one's head that enables one to form thoughts about heat so-described: thoughts of the sort one would express by saying "heat is so and so." The Martian certainly has an idea of heat, for she has thoughts to the effect that "heat looks like a bunch of rapidly vibrating particles." So what is she missing?

Proposal, meant to be in a Kripkean spirit. What sets the Martian apart is that her heat-idea is abnormal. All of our heat-ideas have certain properties in common that the Martian's idea lacks. To know what 'heat' stands for is to know that it stands for heat, where heat is conceived not by any old idea of heat but a normal

idea.[3] I will call this knowing in the normal way that 'heat' stands for heat. Putting this together with the claim about understanding, we get

> (KRU) Understanding 'n' is knowing in the normal way that 'n' stands for n

So for instance I might acquire the word 'Mt Everest' by being told it stands for the world's highest mountain, located somewhere in Asia, or 'the Sun' by being told that it stands for that, the shiniest object in the sky. Something like this is let's assume the normal idea of Mt Everest, or of the Sun. My understanding 'Mt Everest' ('the Sun') is my knowing that it stands for Mt Everest (the sun) as thus normally conceived.

How does this compare to what the rigidifier requires for understanding? Both sides agree that I am expected to know that 'Mt Everest' stands for Mt Everest, conceived as the highest mountain. But (KRU) is content if I know this is true as matters stand. (2DU) says I should know it unconditionally = no matter which world is actual. Do I?

It seems to me that I know Everest is the tallest mountain no matter which world is actual only if my teacher has told me that it is. But she has told me only that being Everest does go with being the tallest mountain. Was the stronger claim perhaps implicit? It would seem not. She would be shocked and horrified to hear me telling my brother, "oh, by the way, if Kanchenjunga should turn out to be tallest, then 'Everest' stands for Kanchenjunga." Her message is this: "presuming I am not greatly mistaken about which mountain is tallest, 'Everest' stands for the tallest mountain." Similar remarks apply to 'the Sun'. No one's

---

[3] Crimmins, "Having Ideas and Having the Concept" (Mind and Language??).

understanding of the sun tells them it is Sirius B if we are massively deluded and that is the star responsible for the appearances by which we identify the Sun.

This shows I think that the 2-D picture of understanding is in one way at least[4] much more demanding than the Kripke picture, and on the face of it more demanding than the truth. The next question is, is any of the additional stuff imputed by the 2D picture actually needed to explain the phenomenon of associated properties?

A reason to doubt it is this. The phenomenon to be explained has to do with necessary conditions on the referent: to be yellowness, a property should look a certain way, to be 100° C a temperature should be the boiling point of water at sea level, to be the Sun a thing should be the shiniest object in the sky. Someone who doesn't expect the referent to have these properties doesn't understand the term as we do. But remember, it is one thing to think if x is the referent it needs to have certain properties, another to think that if x has those properties, it needs to be the referent. The first is a matter of necessary conditions on the referent, the second a matter of sufficient conditions. When the 2-dimensionalist insists that no matter which world is actual, yellowness is whatever feels a certain way, she is talking (at least) about sufficient conditions. She is thus like the imaginary version

---

[4] Less extravagant than KRU, and than the truth, in another way. 2DU doesn't require you to know what 'n' stands for. After all, it's your ignorance of this that's supposed to explain how you can understand 'water' without realizing it stands for H20. But this "ignorance of the referent" is a tendentious redescription of ignorance of some of the referent's essential properties. Why should you need to know the essence of water to know that it's what the word 'water' stands for? If that were the requirement then I don't know what my own name stands for. (Thanks to Brian Weatherson.)

of my teacher who says, I don't care which mountain turns out to be tallest, that's the one we call 'Everest.'

Not only is this extra instruction irrelevant to the explanation of our intuition that to be heat, say, a thing should feel like this, it "explains" an intuition that we don't have, viz. that if, our knowledge to the contrary notwithstanding, it is not fire and soup that feel like this but snow and properly prepared Jell-o, then it is the condition of these latter that we have in mind by 'heat.' When I identify heat as what feels like this, standing before a fire, I mean what does feel like this, given what I know about how various things feel. (If I learn the word from a teacher, her message is not, "heat is whatever presents like so, and now I cast my fate to the winds," it's "heat is what presents like so, presuming here as why shouldn't I that I am not totally misremembering or otherwise mistaking my actual perceptual reactions.")

So far we've had an example of using a cannon to kill a mouse – using a necessary and sufficient connection when the explanation draws only on the necessity -- and an example of hitting some neighboring mice – "explaining" a cast-our-fate-to-the-winds intuition we don't actually have. Next an example of missing the mouse.

Suppose Kripke is right that the Martians have a different concept of heat if they don't feel it as we do. How does the rigidifier propose to explain this? Well, the Martian does not know that 'heat' is what causes heat-sensations, no matter which world is actual. A problem which I won't be discussing is, why can't the Martian know this? It's not as though you need to actually have a feeling to know a fact that alludes to it. (The blind certainly know that light gives rise to visual impressions, and this no doubt plays a role in their understanding of the term.) And anyway, it may be that the Martian does have the feeling, but in response to cold things rather than warm.

The problem I do want to discuss has to do not with the Martians' understanding of 'heat' but our own. If what a person knows whereby they understand 'heat' is that it stands for whatever feels a certain way, then this should be knowable without a prior understanding of 'heat'. But then the feeling by which heat is identified had better have a name other than 'feeling of heat'. Kripke says the following:

> …heat is something we have identified (and fixed the reference of its name) by its giving us a certain sensation, which we call 'the sensation of heat.' We don't have a special name for this sensation other than as a sensation of heat. It's interesting that the language is this way. Whereas you might suppose it, from what I am saying, to have been the other way (131).

(You might indeed.) And later,

> Some philosophers have argued that such terms as 'sensation of yellow,' 'sensation of heat,' 'sensation of pain,' and the like, could not be in the language unless they were identifiable in terms of external observable phenomena, such as heat, yellowness, and associated human behavior. I think that this question is independent of any view argued in the text (140).

How could it be independent, one may wonder? There is indeed a problem here if the role of a reference-fixing description is to specify the referent in prior and independent terms, thereby conferring understanding. But it is only the rigidifier who assumes that understanding is constituted by knowledge of a reference-fixing biconditional. According to Kripke as we are reading him, one understands by

> (1) knowing of the right x that 'n' stands for x, while

(2) conceiving of that x via a normal idea.

A reference-fixing description can contribute in the first connection by specifying in independent terms which x is being referred to; that is how initial baptisms are supposed to work. But it can also contribute in the second connection by reminding us of what counts as a normal idea of x. ("It might here be so important to the concept that its reference is fixed in this way…"(131), "The way the reference is fixed seems overwhelmingly important to us in the case of sensed phenomena…The fact that we identify light in a certain way seems to us to be crucial, even though it is not necessary" (139).) Whether or not it is circular to use the word 'heat' in identifying the referent x of that very word, it is clearly not circular to use the word 'heat' (which we do after all understand) in our account of how we who understand are expected to conceive of its referent.[5]

Explaining the *Epistemic* Datum

Now let's consider the rigidifier's explanation of apriori truths about Neptune, say, or the length a meter. It is a priori, we are told, that a meter, supposing there is such a length (the definiton has not misfired) is the length of this stick. Both sides agree, I think, that the apriority reflects something like immunity to error through misidentification (Jim Pryor, John Campbell in Shoemaker

[5] Lots more to be said about normality, much of it well said by Crimmins. Normality has statistical and normative aspects. It can involve non-representational properties of the idea, e.g., that it calls up certain associations, or that it is triggered by a certain external phenomenon, as our heat-idea is triggered by heat. Sometimes it can be important to be tempted to attribute certain properties to the referent, e.g., indivisibility to atoms or Tarskian properties to truth.

festschrift), so a word about that. Error through misidentification happens when

> one correctly supposes that $\Box\exists x\ x = n$ and that $\Box\exists x\ Gx$
> but one wrongly supposes that Gn.

Our judgment that n is G is immune to error through misidentification if there is no chance at all of this happening; assuming both exist, n is bound to be G. So it is with the reference-fixer's judgment that "a meter is the length of this stick." There is no chance at all that there is a length one meter and something is the length of this stick but a meter is not the length of this stick.. Given how the phrase was introduced, it is this referent or nothing. 'A meter' has no other option if it wants to refer.

That much, it seems, Kripke and the rigidifier can agree on. They disagree as to why 'a meter' has only one option if it wants to refer. What is it about our understanding of 'n' that makes it the case that

> (*) 'n' refers if it does to the G?

The rigidifier's story is based on

> (2DU) To understand 'n' is to know that NMWWIA, 'n' refers to x iff x is the actual G.

This lets her reason her way to (*) as follows:

(1) 'n' refers
(2) 'n' is understandable, say by X
(3) X knows that NNMWWIA, 'n' refers to x iff x is the actual G
(4) NNMWWIA, 'n' refers to x iff x is the actual G
(5) 'n' refers to x iff x is the actual G
(6) 'n' refers to x only if x is the actual G

Conditional proof gives "if (1) then (6)" which then by basic logic allows the rigidifier to derive (*).

Notice how inefficient this explanation is. (3) is the rigidifier's notion of understanding The step from (3) to (4) takes us from 'X knows that A' to A. The step from (4) to (5) take us from 'NNMWWIA, B' to B; the step from (5) to (6) takes us from 'C iff D' to 'C only if D.' And (6) is all the explanation really requires.

This leads us to one wonder if the same explanation might be available at cheaper rates from the Kripkean. This story too will be based on a theory of understanding:

> (KRU) To understand '<u>n</u>' is to know in the normal way (using a normal idea) that it stands for <u>n</u>.

To understand 'a meter', on this theory, is to know in the normal way – using a normal idea -- that it stands for a meter. Now at this point our <u>only</u> idea of a meter is as the length of this stick, so an idea that leaves the connection to stick S out would have to count as abnormal. And now the Kripkean can argue as follows:

(1) '<u>n</u>' refers
(2) '<u>n</u>' is understandable, say by X
(3') X knows that '<u>n</u>' refers to the G
(4') '<u>n</u>' refers to the G

(*) now follows by conditional proof. Relative to the goal of explaining apriority, the surplus content of (2DU) vis a vis (KRU) seems just wasted

That was a "using a cannon to shoot a mouse" type criticism. Now an example of collateral damage, that is, the rigidifier "explains" things that aren't the case. Distinguish two claims:.

(A) if there's such a length as a meter, it's the length of this stick.

This I have agreed is a priori, because 'a meter' refers to the length of stick S if to anything. Second,

(B) if this stick has a length at all, then the length is a meter.

It should be clear the apriority of (B) follows just as easily from the rigidifier's notion of understanding as (A). But is (B) in truth apriori? Recall how Kripke sets the case up:

> There is a certain length which he wants to mark out. He marks it out by an accidental property, namely that there is a stick of that length. Someone else might mark out the reference by another accidental property (55).

Since it is an empirical matter whether stick S is "the length he wants to mark out," we need to ask what happens if he is wrong and it is not. It might be for instance that the stick is a millionth of an inch long but emitting magnification rays that delude us into seeing it as longer. Or maybe the stick is a mile long but much farther away than anyone had realized. I take it that it is no part of the reference-fixer's understanding of 'meter' that it continues to stand for the length of S even if S is much shorter than it appears. Since this cannot be apriori ruled out, we don't know apriori that that the stick is a meter long if it has a length at all.

You can guess the rigidifier's reply:. "That just shows we have not been sufficiently careful about the descriptive condition that defines 'one meter.' The real meaning of 'one meter' is 'the length of that stick, presuming the stick is roughly as long as it looks.' I agree that this is how the answer has got to go.

But at the same time it <u>can't</u> go that way, because of the rigidifier's difficulties about real-actuality. The phrase "as long as it looks"

cannot mean "as long as it looks to our as-if actual selves," for then it is OK if the stick is a mile long in w, provided there are compensating changes in our perceptual system: we have telescopic vision in w so that it takes a mile of stick to make true the experience that a much shorter stick answers to here.[6] The phrase has got to mean "as long as it does or would look to us as we really are." And as we have seen, the 2-dimensionalist has no way of capturing that "really." His only option is to determine empirically that our actual perceptual wiring is X, Y, and Z, and then write X, Y, and Z into the definition of "as long as it looks." That may deliver the right extensional results but at the cost of changing the subject, since our concept of as long as it looks is clueless about the neurophysiology of vision.

Explaining the *Modal* Datum

The third phenomenon for which an explanation is sought is the seeming possibility of things that are in fact impossible. .The 2D

---

[6] I am skating over various subtleties here. Suppose that we have some sort of characterization of my experience in response to this stick: it's a type E experience. Since we don't want to make a "definition into a scientific discovery rather than a synonymy," we must take care that the characterization not help itself to features of real actuality that users of the concept are in no position to know, features that would have to be discovered empirically. This applies I take it to how long in inches a thing must be to answer to my current experience. One reason is that I might not know about inches. I might not yet have any measures of length but perceptual ones. A second reason is that just as I am often surprised by how much bigger a piece of furniture is than I had guessed on the basis of its appearance in the store, I put no great stock in my guess as to the length in inches of a stick that looks to me like this one does now. That an experience is E-type should be silent on the question of how objectively long it represents objects as being.

explanaton of why it seems possible that S is that there are possible worlds w such that if w is actual, then S. It seems like Hesperus could have been distinct from Phosphorus simply because the actual evening-visible planet = the actual morning-visible planet is false on certain hypotheses about which world is actual. I want to argue that depending on how you run it, the 2D style of explanation either explains too much (collateral damage) or doesn't explain enough (misses the mouse).

Recall a key feature of Kripke's approach to illusions of possibility. Kripke doesn't just want to show how someone could fall under the misimpression that, say, Hesperus could have failed to be Phosphorus, by misinterpreting what was in fact a different possibility. That would be easy, since a sufficiently confused person could presumably misinterpret anything as anything. He wants to show that we plausibly do fall under the modal misimpression by misinterpreting a different possibility. It is not just that an intuition of E's possibility could, but that our intuition of its possibility plausibly is, based on the mistaking of one possibility for another. One should be willing to say: oh, I see, once you point out the difference, it's because this really is possible that I supposed that to be possible.

The kind of principle I am relying on here is familiar from psychoanalysis. Here is what in my brief (well….) experience psychoanalysts tell you. "You are under the illusion that nobody loves you. A cruder sort of doctor might say, here is how the illusion arises, take my word for it, now you are cured. But I would never dream of asking you take my word for it. No, the test of my explanation is whether you can be brought to accept the explanation, and to accept that your judgement is to that extent unsupported." The analogy is good enough that I will speak of the

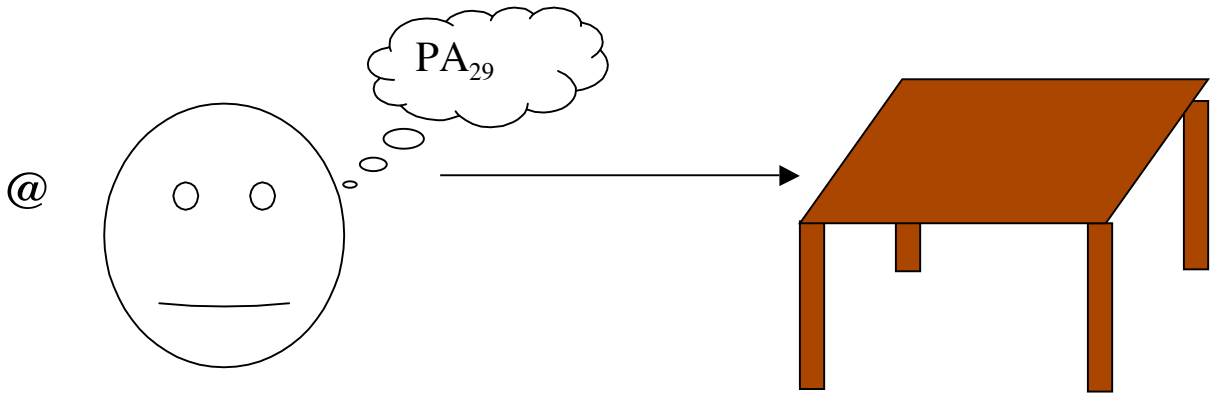
> Psychoanalytic Standard Assuming the conceiver is not too self-deceived or resistant, ◊F explains E's seeming possibility

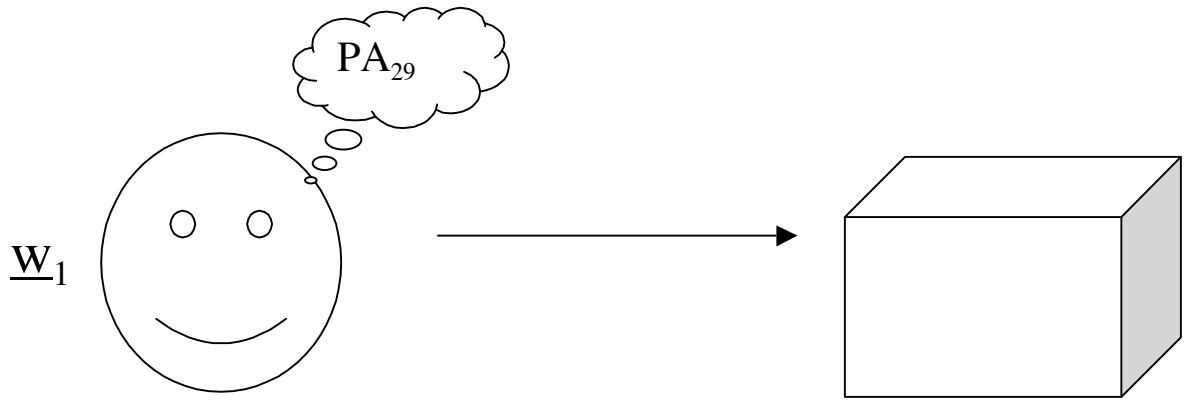
only if he/she does or would accept it as an explanation, and accept that his/her intuition testifies at best to F's possibility, not E's.

This is a high standard, but what makes Kripke's approach so convincing is that this is the standard he tries to meet, and mostly does meet. Philosophers have been telling us for centuries that this or that common impression is false. And we have for centuries been shrugging them off. What makes Kripke special is that he gets you to agree that you are making the mistake he describe. Whether the rigidifier can get you to agree you are making the mistake he describes is not so clear. 0

One way to see the problem is to look at Kripke's explanation of modal illusions in terms of "qualitatively identical epistemic situations"; it seems possible that x is P because its counterpart x* in a qualitatively identical epistemic situation really is P. What does he mean by that phrase qualitatively identical epistemic situation? One obvious thought is is to be in the same epistemic situation as me now is to enjoy the same (narrowly individuated) perceptual appearances: to be, say, in perceptual appearance stae PA_{29}.

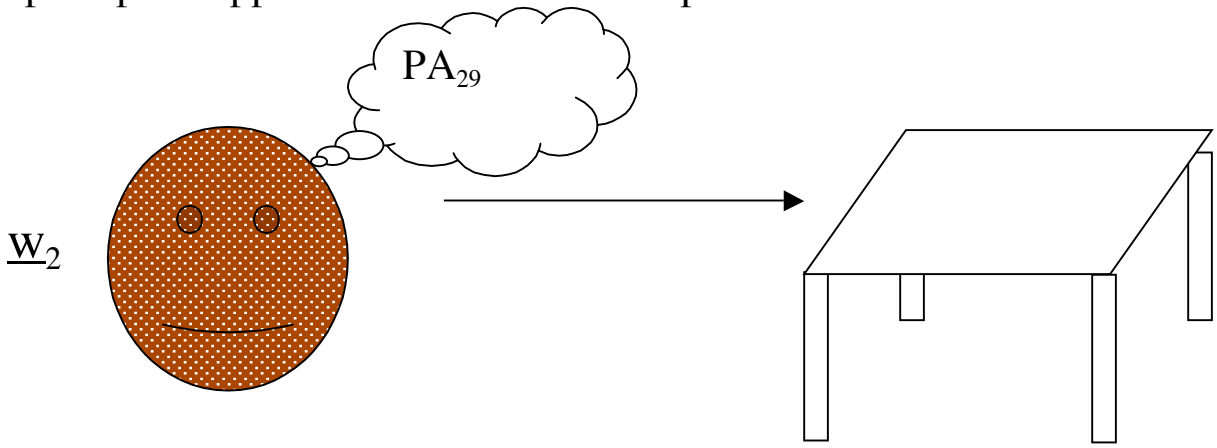
But that seems not to be enough. Take the illusion that this table could have been made of ice. One world I am pretty sure is out there is a world w_1 where Counter-Steve is on drugs so powerful that an ordinary old block of ice looks to him just like this table looks to me.





Does the possibility of a world like that explain why it appears to me that this table could turn out to be made of ice? I take it that it doesn't. There is no temptation to say, "oh now I see why this brown table seems like it could be made of ice, it's because there could be a guy to whom regular ice looked like this."

Well, maybe the problem with that first explanation is that $Steve_1$ is perceptually deluded. The way things appear to him is not how they are. A second idea then is that someone is in my epistemic situation if they enjoy the same (narrowly individuated) perceptual appearances and their experience is veridical. This doesn't work either, I think, because my doppelganger need not be deluded even if he is looking at a visibly icy table. All he needs is to be differently wired so that white things produce in him the same perceptual appearances as brown ones produce in me.

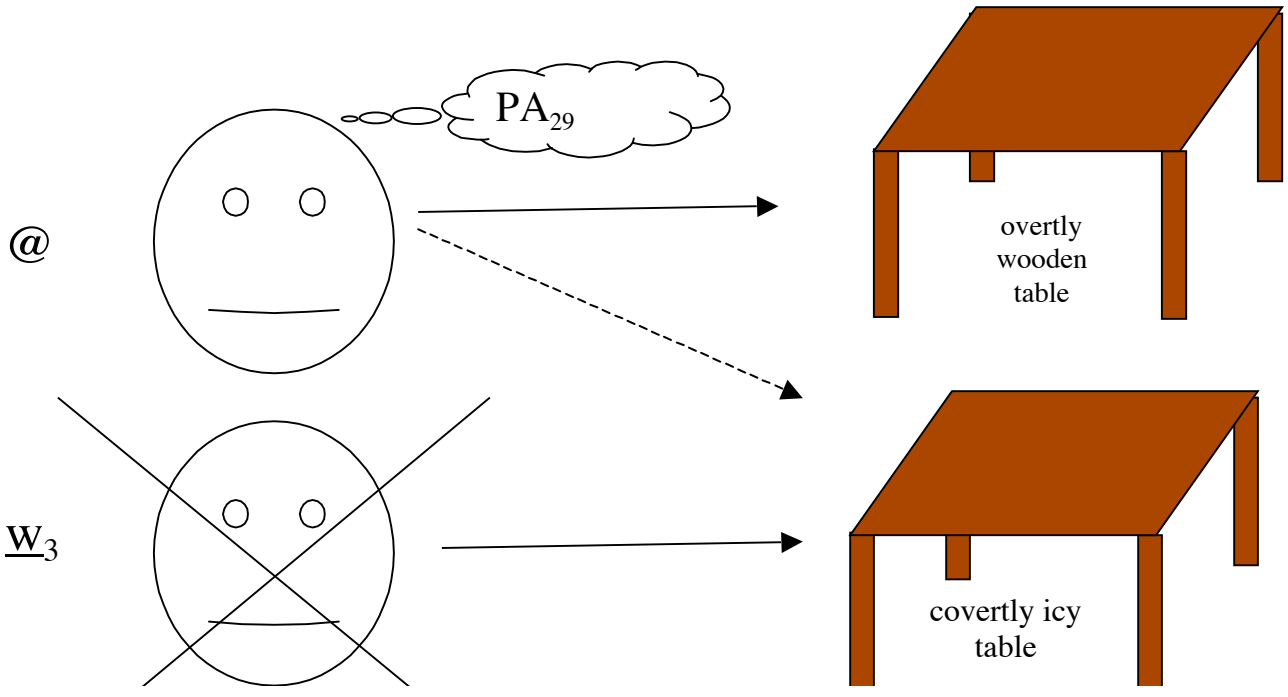


Once again we don't think, "a ha, what seemed like the possibility of this brown-looking table being made of ice was really just the possibility of a

spectrum-inverted Steve$_2$ to whom white things look the way brown ones look to me." I am not thinking, as far as I can tell, I am in the Steve$_2$. situation, because the Steve$_2$ situation is <u>defined</u> by a contrast with mine. (Like worrying that this might be Twin Earth. Dretske zebra analogy)

A third reading, which gets closer, I think, is that someone is in my same epistemic situation if the scene they are experiencing has the same perceptually available properties as this one. (This rules out the <u>w$_2$</u> scenario because brown is after all a perceptible property, and in <u>w$_2$</u> it's missing.) But remember, Kripke also wants to explain in this way the seeming possibility of <u>brown</u> having a different physical nature than it has in fact. That will require a counterpart situation where at least one perceptible property, viz. brown, is changed. So sameness of epistemic situation cannot require sameness of perceptible properties.

I see only one other option and it's this. Someone is in my same epistemic situation if the scene he is experiencing is a dead ringer for the scene I am experiencing, meaning that the two are for me perceptually indistinguishable. If you quickly substituted his situation for mine, keeping my perceptual systems the same, I would be none the wiser; it would not appear that anything had changed. The picture we want then is



Note that the reactions of my as-if actual self are irrelevant on this picture; it's real me to whom the icy table has to look just the wooden table. The feature of w_3 that makes it explanatory -- the table there looks the same to real-me -- is not even expressible in 2-dimensionalist terms. The closest the 2-dimensionalist can come is to say a world it is possible for an icy table to produce PA_{29} in someone whose perceptual system is – plug in here empirically determined features X, Y, and Z of my perceptual system. There is a world like that – it might even be w_3 – but it can't explain my illusion because what seems possible to me is not that an XYZ observer mistakes this icy table for wood but that I am mistaking this icy table for wood.

So we have the following principle: to explain why this, an object of our acquaintance understood to present like so, seems like it could turn out to be Q, one needs a possible scenario in which something superficially indistinguishable from it does turn out to be Q. The counterfactual thing has to look the same, not to the counterfactual folks, but to us. I will call that a facsimile of the actual thing. And I will refer to the principle as the facsimile principle, or the fools' gold principle. If you want to explain in a psychoanalytically satisfying why it seemed possible for gold to be iron pyrites, the explanation should not be that there's this perfectly ordinary brownish hunk of rock ("dunce's gold") not looking like gold to us but looking to the people around it as gold looks to us. Since the 2-dimensionalists cannot express facsimilehood, they drop out of the competition already here.

I said that Kripke respects the psychoanalytic standard and that his explanations often satisfy it. But it seems to me that this is one of those rare cases where the 2D-ist error can be traced back to Kripke. Sometimes not even Kripke has a psychoanalytically satisfying explanation. Sometimes he is forced like the 2D-ist to appeal to "dunce's gold" when it is fool's gold we want.

Here is some heat; is it HMME (high mean molecular energy)? One has to do conduct additional tests. And like any tests, they could come out either way. So there's the appearance that heat could turn out to be HMME, and the appearance that it could turn out to be something else, say LMME. The second appearance is an illusion. How does Kripke propose to explain it away?

> the property by which we identify [heat] originally, that of producing such and such a sensation in us, is not a necessary property but a contingent one. This very phenomenon could have existed, but due to differences in our neural structures and so on, have failed to be felt as heat (NN, 133).

Let's say, to make it definite, that the difference in neural structures had the result that high MME felt cold, and low MME felt hot. Does this explain in a psychoanalytically satisfying way the illusion that it could have been low MME that was heat rather than high? Does it explain the illusion that heat could have turned out to be low MME to point to possible differences in our neural structures?

Here is the worry. With the table, remember, what seemed possible is not just that ice could have paraded itself in front of <u>someone or other</u> who saw it as wood, but that there could have been ice that <u>I with my existing sensory faculties</u> would have seen as wood. To explain that seeming we needed a facsimile of the table – a spitting image of it – that was in fact ice. Likewise what seems possible in the case of low MME is not just that it could have paraded itself in front of <u>someone or other</u> who felt it as hot, but that <u>I</u> with my existing neural structures could have found it hot. To explain that seeming, we need a facsimile of heat that turns out to be low MME. There should be the possibility of fool's heat which turns out on inspection to be low MME. Similarly to explain the seeming possibility of high MME turning out to be cold, we would

need the possibility of fool's cold that was found on inspection to be high MME.

Is there fool's heat of this type, or fool's cold? I don't see how there could be. It may be possible to slip a cleverly disguised icy table in for this wooden one with no change in visual appearance. But it is not possible to slip cleverly disguised low MME in for high MME and have it feel just the same. Having substituted low MME for high, there is no other way to preserve the appearances but to postulate observers with different sensory reactions than ours. But then what we are getting is not really fool's heat but something more like dunce's heat. Because, as discussed, you would have to be pretty confused to see in the possibility of rewiring on your side the explanation of why a switcheroo seems possible on the side of phenomenon you are sensing. Whether fool's heat is absolutely impossible I do not know. But what does seem clearly impossible is for low MME to be fool's heat, because it by hypothesis feels the opposite of hot; it feels cold.

Kripke is right, or anyway I'm not disagreeing, when he says that "the property of producing such and such a sensation in us…is not a necessary property," because we could have been wired differently. High MME could have produced what we call sensations of cold. That is not what I am worried about. What worries me is that the property of interest is not that but producing such and such a sensation in us as we are. And that property does seem to be necessary. There are only three factors in how an external phenomenon is disposed to feel: its condition, our condition, and the conditions of observation. If all these factors are held fixed, as the notion of fool's heat would seem to require, then it is hard to see how the sensory outcome can change.