

# No Fool's Cold: Notes on Illusions of Possibility<sup>1</sup>

S. Yablo, MIT

A lot of philosophers are pessimistic about conceivability evidence. They think it doesn't prove, or even go very far towards justifying, interesting modal conclusions. A number of other philosophers are optimistic; they think it does justify, and perhaps even establish beyond a reasonable doubt, that lots of interesting things are possible. Nothing very surprising there. What is slightly surprising is that both groups can claim to find support their attitude in the work of Saul Kripke.

Pessimists say: Kripke shows that conceivability evidence is highly and systematically fallible. Very often E seems possible, when as a matter of fact, E-worlds cannot be. So it is, for instance, with the seeming possibility of water in the absence of hydrogen, or of Hesperus distinct from Phosphorus, or of this table turning out to be made of ice. Let the pessimistic thesis be

(P) oftentimes E seems possible when it isn't, so conceivability evidence is not to be trusted.

Optimists reply: yes, Kripke finds conceivability evidence to be fallible, but that is only half of the story. The rest of the story is that the failures always take a certain form. A thinker who (mistakenly) conceives E as possible is correctly registering the possibility of something, and

---

<sup>1</sup> This paper was presented at the UNC Greensboro conference on imagination and possibility, with comments by Keith Simmons. Thanks to Keith for exposing various gaps in the argument, not all of which I have been able to deal with here. Thanks to Kit Fine, Tamar Szabó Gendler, Janine Jones, and Saul Kripke for advice and discussion at the conference, and to David Chalmers and Tyler Doggett for extremely helpful written comments provided more recently.

mistaking the possibility of that for the possibility of E. There are illusions of possibility, if you like, but no delusions or hallucinations.

(O) Carefully handled, conceivability evidence can be trusted, for if impossible E seems possible, then something else E is possible, such that we mistake the possibility of E for that of E.

The optimistic thesis (O) represents conceivability evidence as in a sense infallible. If (O) is correct, then that E seems possible, while it may not establish that E is possible, does succeed in establishing the disjunctive conclusion that either E is possible or E is. And indeed in certain cases we can get all the way to the first disjunct, because E is tantamount to E or entails E. This, the optimist continues, is the situation we encounter in the last few pages of Naming and Necessity, where Kripke argues against the identity theory of mind. It seems possible that pain isn't c-fiber firings, and the E that allegedly snookers us into thinking E possible is tantamount to that original E. (I will be questioning that argument in due course.)

It seems likely that both groups are overinterpreting Kripke. Certainly Kripke is not a pessimist, because he closes the book with a positive argument of the sort that pessimists are bound to find fault with. And although this is not as clear, he seems to stop short of outright optimism too. He says (in "Identity and Necessity") that "the only model I can think of for what the illusion might be....does not work in this case" (1977, 101, emphasis added). Others are welcome to argue in favor of some other model that doesn't require a genuinely possible E. Kripke is skeptical, to be sure: "it would have to be a deeper and subtler argument than I can fathom and subtler than ever appeared in any materialist literature that I have read" (IN, 101). But although Kripke has his doubts about the availability of an alternative model, he doesn't rule it entirely out. (Compare Carnap's position in "Empiricism, Semantics, and Ontology: I can't make sense of the question of realism my way, maybe others can find a different way, but it won't be easy.)

So the door is open, technically anyway, to "a deeper and subtler argument" aimed at establishing that some seeming possibilities do not reflect any sort of genuine possibility. Whether this deeper and subtler argument can be given has not been terribly much explored.

\*\*\*\*\*

One idea sometimes encountered is that there are differences in how pains and c-fiber firings are entertained in thought that all by themselves explain why each would seem possible without the other. Thomas Nagel's version of this idea is that C-fiber firings are imagined perceptually – “we put ourselves in a conscious state resembling the state we would be in if we perceived it” – while pain is imagined sympathetically – “we put ourselves in a conscious state resembling the thing itself” (1974, note 11):

the relation between them will appear contingent, even if it is necessary, because of the independence of the disparate types of imagination (ibid).

Chris Hill says in a similar vein that the relation appears contingent because our concept of c-fiber firings is theoretical while our concept of pain is phenomenological. Between concepts like that “there are no substantive a priori ties,” and the absence of such ties allows us to “use the concepts to conceive coherently of situations...in which there are particulars that fall under one of the concepts but do not fall under the other” (1997, 75).

This sort of approach is in one way too broad and in another too narrow. It is too broad in that it threatens to undermine conceivability arguments that most of us find attractive. It certainly seems to me that this cat could have been in severe pain right now. But then so it would, according to Nagel, what with the cat being imagined perceptually and the pain sympathetically.

I agree that the appearance here should not be taken seriously, if it arises in the way Nagel says. That we do take it seriously suggests that the explanation may not be quite so simple. If appearances of contingency resulted just from “disparate types of imagination,” then more would seem possible than in fact does. After all, it is not just the cat that is imagined perceptually but everyday objects in general. Consider the rock that the cat is perched on. All the Nagelian conditions are in place, yet it does not seem that the rock could have been in pain right now. It takes more to tempt us into an illusion of possibility than Nagel supposes.

What about Hill's version of the idea? It seems to me, as I consider this cup of vinegar, that a cup of H<sub>2</sub>O could look just the same. But then so it would, on Hill's view, for looking the same is a phenomenological concept, while our concept of H<sub>2</sub>O is theoretical. Once again, though, this cannot be all there is to it, for there are cases where Hill's conditions are met and the appearance of contingency is lacking. A cup of room-temperature uranium could not look like this.<sup>2</sup>

How is the Nagel-type approach too narrow? By focusing so intently on subjective vs. objective, it just reinforces the impression that Kripke is trying to create, namely that any response to his argument is going to require some kind of special pleading on behalf of the mental. I cannot rule it out, of course, that the proper response does require special pleading. But it would be better if we could identify a general constraint on modal illusions that is independently motivated and that just happens to deliver the desired results when applied to the intuitions supporting mental/physical dualism.

\*\*\*\*\*

I want to explore some of these issues by looking at the role of actuality in modal judgments. Actuality comes in under two separate headings. On the one hand it can figure in the content of a modal judgment. The thing that seems possible -- the condition that seems like it could have obtained -- can have the notion of actuality in it. This is in fact quite common. One says, for instance, "this lemonade is cold but it could have been colder."<sup>3</sup> Colder than what? Colder than it actually is, of course. If C is the "how cold was it?" parameter, then our judgment is roughly this

seems  $\diamond$  (C exceeds C<sub>@</sub>)

---

<sup>2</sup> Tyler Doggett and Daniel Stoljar point out that the Nagel worry also pulls the rug out from under standard objections to behaviorism and functionalism. Given any behavioral property B, we can imagine being in pain without exhibiting B and vice versa. Perhaps the appearance of contingency here is due just to the fact that pain is imagined sympathetically and B perceptually.

<sup>3</sup> Could have been colder as a liquid, I mean. Assume for the sake of the example that so-called frozen lemonade is not really lemonade.

Or perhaps we are doing a puzzle where five irregularly shaped pieces of plastic have to be rearranged into a square. We look the pieces over and it strikes us that the thing can be done. What seems possible, however, is not that the pieces can be made to form a square after being melted down and recast as rectangles; it's that they can be made to form a square with their actual shapes and sizes held fixed. If the shape and size of piece X is S(X), then our judgment is

seems  $\diamond$  (the Xs form a square &  $\forall X (S(X) = S_{@}(X))$ )

A remark attributed to Richard Taylor gives us a third example. "Why are people so sure they could have acted otherwise?" he asks. "After all, nobody ever has." Well, one reason we think this is that it very much seems as though we could have acted otherwise:

seems  $\diamond$  (my action was of a type T incompatible with the type T<sub>@</sub> of the action I really did perform)

To have a schema for judgments of this kind, what seems possible in these cases is that a certain parameter P should have taken a value so and so related to the value it actually takes:

seems  $\diamond$  (.....& P is so and so related to P<sub>@</sub> & .....

That is the first way actuality can come in. It leads pretty directly to a second way. Whether or not it seems possible for some parameter to assume a value so and so related to its actual value is not independent of what we know, or think we know, about what the actual value in fact is, or indeed of other information we possess about actuality. It would not have seemed possible for the pieces to be rigidly rearranged into a pentagon if we had believed each piece to be square, or round. It would not have seemed possible for the room to be colder if it was believed to be at absolute zero already. It might not have seemed possible for us to act otherwise were we convinced that Frankfurt's nefarious neurologist (or an omnipotent variant thereof) stood ready to reprogram our brains if we tried.

There is a temptation, perhaps, to treat this as just more content. But the temptation should be resisted, because it imports more into the content than belongs there. Our judgment is not

seems  $\diamond$  (this lemonade is colder than  $N^\circ$  C).

After all, we may have little positive idea what temperature the lemonade is in degrees centigrade. What seems possible is that the lemonade should be colder than it is, and why it seems possible has to do with the lemonade's felt temperature.<sup>4</sup>

If our sense of the temperature doesn't figure in content, though, what role does it play? It plays what might be called a presuppositional role. The judgment is conditioned on our sense of the temperature's not being too misleading. One thinks, "unless I am very much misled about how cold this liquid is, it could have colder." Besides appearing in the content of a modal judgment, then, actuality can figure in the background to the judgment, that is, the beliefs or presuppositions that allow the seemingly possible thing to seem possible.

\*\*\*\*\*

Back now to the main issue. The optimist says that whenever there's the illusion that E is possible, there's a related hypothesis E that really is possible. For instance, it seems that Hesperus could have been distinct from Phosphorus because there really could have been two planets there, one responsible for Hesperus-appearances and the other for the appearances we enjoy of Phosphorus. I have said a little about E, the content of the (perhaps mistaken) intuition, but nothing about E, the hypothesis that is supposed to really be possible.

Kripke does not even pretend to give us a general strategy for recovering E -- what I will call the underlying possibility -- from E. What he does do is, first, sketch lots of highly convincing examples; second, suggest that at least some of the time, it is good enough to replace names in E with corresponding reference-fixing descriptions; and third, characterize E as the "appropriate corresponding qualitative contingent statement." He explicitly refrains, though, from giving a "general paradigm" for the construction of the proposition whose possibility fools us into thinking E possible.

---

<sup>4</sup> Specifically, with its feeling warmer than lemonade on the verge of freezing feels.

A number of other writers have been bolder. Some say that that there is the illusion that  $\underline{E}$  is possible because the sentence " $\underline{E}$ " could (with its "meaning" in some sense of that word held fixed) have expressed a true proposition, albeit not the proposition it expresses in fact. So,

(a) it could have happened that " $\underline{E}$ " expressed a true proposition.

I myself once conjectured that  $\underline{E}$  seems possible because we could have thought something true with the thought (the internal mental act) whose content in this world is  $\underline{E}$ . So,

(b) it could have happened that thinking the  $\underline{E}$  way was thinking truly.

The best-known suggestion along these lines is that  $\underline{E}$  seems possible because there are worlds such that if (contrary to what we perhaps suppose), they are actual, then  $\underline{E}$ .<sup>5</sup> So a third hypothesis is that

(c) things could have been a way such that, if they actually are that way, then  $\underline{E}$ .

All these proposals are variations on the theme of  $\underline{E}$  seeming possible because what it says is correct, if a certain not-impossible world is actual. Nothing important is lost if we ignore any differences and speak simply of the if-actually account of illusions of possibility.

The if-actually account works extremely well in some cases. The reason it seems possible that the table should turn out to be made of ice is that there are worlds with the property that if they are actual, then it is made of ice. The reason it seems possible that Hesperus should have been other than Phosphorus is that there are worlds with the property that if they are actual, it really is other than Phosphorus. It turns out, though, that the account cannot deal correctly with actuality-based modal contents. I will build up to this slowly.

---

<sup>5</sup> Chalmers's answer to this (personal communication) is that (c) is his explanation of seeming indicative possibility, not subjunctive. The zombie intuition for him is not that we all could have been zombies with no physical change, but that everyone (else) might really be a zombie.

Ivory-billed woodpeckers had been thought extinct; recently, though, a man named David Kullivan reported spotting a pair of them. I happen to believe this report, but not everyone does. Knowing that his word would be doubted, Kullivan was tempted (let's say for purposes of the example) to shoot one of the pair and bring its body back as proof. According to me, believing as I do that ivory-billed woodpeckers exist, had Kullivan shot one, there would have been fewer ivory-billed woodpeckers than there are. To me, then

seems  $\diamond$  (there are fewer ivory-billed woodpeckers than actually)

Now suppose that I am wrong and there are no ivory-billed woodpeckers. Then I am under an illusion of possibility; a smaller number seems possible, but there can't be fewer than none. What explains my illusion? The story would have to be that this seems possible because there's a world such that if it is actual, then there are fewer ivory-billed woodpeckers than there actually are. And that makes no sense.

Of course, there is no peculiarly modal illusion here; where I go wrong is in believing in ivory-billed woodpeckers in the first place. But consider a second example. It seems possible that Hesperus could have turned out to be distinct from Phosphorus. It seems, for example, that Phosphorus could have turned out to be Mars rather than Venus. Another thing that seems possible is for Phosphorus to have turned out to be Xorg, a solar planet over and above the planets that exist in fact. It seems possible, then, that there should have been more planets than actually: all the actual ones, including Hesperus, and then in addition Phosphorus = Xorg.

seems  $\diamond$  (there are extra planets; Hesperus is Venus but Phosphorus is new)

The story would have to be that this seems possible because if we are wrong and the morning-visible planet is "new," then there really are more planets than actually. And once again, that cannot be right. Again, it strikes us that gold could have turned out to have a different chemical makeup. The illusion that gold could have failed to be an element can be explained, notice. But I may not know that gold is an element; my thought is just that it didn't have to turn out with that chemical makeup, whatever the makeup in fact is. This illusion cannot be explained on the



if-actually model, for we would need a world such that gold has a different makeup than it actually does on the supposition that this world is actual.

\*\*\*\*\*

So the if-actually account cannot explain certain illusions of possibility, those in which the hypothesis that seems possible involves a contrast or comparison with actuality.<sup>6</sup> Why should we bother about this? The reason for bothering is that it tells us something about how people are thinking of the modal illusion problem. The if-actually account is exceedingly popular. (I stress that Kripke does not endorse it.) Why, if there is a class of illusions it does not address? It must be that this class of illusions has not been much on people's minds. People have been assuming, implicitly anyway, that the contents of error-prone modal judgments are actuality-neutral in the sense, roughly, facts about which world is actual bear on what the judged hypothesis says. Perhaps to be safer I should just say that there has been a tendency to downplay or underestimate the actuality-based aspects of these contents, and to play up or overestimate their actuality-neutral aspects.

One sort of problem this bias in favor of neutrality leads to has already been seen. But the problem that interests me is not that certain actuality-based illusions will prove difficult to explain, but that certain such illusions will be "explained" too easily. This is how it would happen:

(1) What seems possible is a hypothesis E that is actuality-based.

---

<sup>6</sup> One natural idea about actuality-involving illusions (suggested independently by Robert Stalnaker and David Chalmers) is this: they are to be explained by saying there is a world w such that if w is actual, then the actuality-involving proposition is possible. It seems possible for there to have been fewer ivory-billed woodpeckers because this really is possible on the hypothesis that Kullivan's story is true. But the intuition that Hesperus could have been an additional planet is not based in any factual misinformation of the sort we might try to correct by treating w as actual. The feeling is not that assuming Phosphorus is other than Hesperus, it could have been an additional planet. The feeling is that Phosphorus, although (it turns out) identical to Hesperus, could have been distinct from it in a way that bumped up the number of planets.

- (2) An actuality-neutral (or more neutral) hypothesis  $\underline{E}$ ' is covertly substituted.
- (3) One explains the illusion that  $\diamond \underline{E}$ ' as a subtle misreading of  $\diamond E'$ .
- (4) It would take a very much grosser misreading of  $\diamond E'$  to fall under the illusion that  $\diamond \underline{E}$ .
- (5) One thinks the  $\underline{E}$  illusion has been explained when really it hasn't.

I will give examples in a minute. But first let me link the worry up with what I take to be an important feature of Kripke's procedure.

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  $\underline{E}$ 's possibility could, but that our intuition of its possibility plausibly is, based on the mistaking of one possibility for another.

An example of someone who seems to underestimate the aspiration here is Michael Della Rocca in "Essentialism and Essentialists" (Journal of Philosophy 1996). Say that Lump1 is the lump of clay composing the statue Goliath. It seems possible that Lump1 could have failed to be Goliath, or any other statue; it seems possible, indeed, that Lump1 could have existed in the complete absence of statues.

(a) seems  $\diamond$  (Lump1 exists without any statues)

Della Rocca maintains that this intuition is (or might be for all Kripke has to say about it) explained by the possibility that a lump of clay handled by artisan A at time I should have lacked all these properties

(b) really  $\diamond$  (a lump handled by  $\underline{A}$  at  $\underline{I}$  exists without any statues).

I suppose that (b) might perhaps explain the illusion of someone for whom the reference of "Lump1" was fixed by "the lump of clay handled by  $\underline{A}$  at  $\underline{I}$ ." But "Lump1" in our mouths has its reference fixed by "the lump composing the statue Goliath." (That is how I introduced the term above,

and that is the usual way of introducing it.) So, the genuine possibility needed to explain away our intuition is

(c) really  $\diamond$  (a lump composing the statue Goliath exists without any statues)

But of course there is no such possibility as (c); it cannot happen that a lump both composes a certain statue and fails to coexist with any statues. The scenario that (c) calls possible, and whose possibility would be needed to explain away the intuition that Lumpl could exist without statues, makes no sense.

I seriously doubt, then, whether our actual intuition of Lumpl without statues can be defeated as easily on Della Rocca suggests.<sup>7</sup> The only real possibility in the neighborhood is the one recorded in (b). And there is no way on earth that we are misinterpreting that as the possibility of Lumpl without any statues. The proof that (b) doesn't explain (a) is just that stare at (b) as long as you like, one cannot imagine being so confused as to have been fooled by (b) into supposing that (a). One is not at all tempted 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.<sup>8</sup>

---

<sup>7</sup> Della Rocca brushes up against this problem in a footnote. "One might, perhaps, see some other property as the property in terms of which Lumpl is identified. Even if some other property is the identifying property, the argument that I am about to give would not be affected because I shall show that any property that might plausibly be seen as the property in terms of which Lumpl is identified would be a property that allows a Kripkean reconstrual of our intuition of contingency in this case to go forward" (197). I do not see that he ever shows this. What he does say is that "Lumpl seems to be identified in terms of the designation, 'lump formed by etc', or some similar designator. Any such designator would allow the reconstrual to go through" (197-8). This is false, unless "similar" means "designator H such that there could be an H without Goliath existing." The designator "clay composing Goliath" is an obvious counterexample.

<sup>8</sup> Della Rocca agrees that the (b) possibility is not judged explanatory. It is just that any attempt to justify this judgment winds up begging the question at issue: which modal intuitions are windows on possibility and which are illusions of possibility?

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 impression that nobody loves you. I submit that this is an illusion. A cruder sort of doctor might say, here is how the illusion arises, take my word for it. 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 judgment 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,  $\diamond E$  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  $E$ '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 describes.

\*\*\*\*\*

I said that Kripke "mostly" meets the psychoanalytic standard. This is because I think that with at least some of the illusions he discusses, the standard is not met, and is perhaps unmeetable. Let me start with an example where a psychoanalytically acceptable explanation can be given. I will then argue that a crucial feature of the example goes missing in Kripke's treatment of certain other examples.

Kripke says: "...though we can imagine making a table out of another block of wood or even from ice, identical in appearance to this, and though we could have put it in this very position in the room, it seems to me that this is not to imagine this table as made of wood or ice, but rather it is to imagine another table, resembling this one in all external details, made of another block of wood, or even of ice" (1980, 114, emphasis added).

Imagine someone, call them Schmipke, expressing puzzlement about Kripke's procedure: "Hasn't Kripke gone to a lot of unnecessary trouble here? Why does he impose this condition of identical in appearance with the actual table? "Identical in appearance" suggests that the otherworldly table looks just like the real one to us: if both of them were sitting here side by side, we could not tell them apart. This is suggested as well by the language he uses in "Identity and Necessity": "I could find out that an ingenious trick has been played on me and that, in fact, this lectern is made of ice" (1977, 88). The ice has to be "cleverly hardened" in the shape of a table, and presumably painted too. Otherwise it would not be a spitting image of our actual table, as Kripke clearly intends. Is any of this really necessary? Why does Kripke ask w to satisfy the actuality-based condition that its table looks or would look just the same to us? What is wrong with the neutral condition of, not identical in appearance, but simply: identical appearances?<sup>9</sup>

This seems a fair question, so let's try it. Until further notice, all we require from w is that there's an icy table there, and that the people looking at it (perhaps counterfactual versions of ourselves) have the same experiences qualitatively speaking as we do looking at our table. It is of course compatible with this that the tables look to us very different. But then our reason for thinking of the icy table in w as "in disguise," cleverly tricked up to look like wood, no longer applies. Now that we have dropped the identical-in-appearance requirement, the icy table can be made any number of ways. Let it be, say, a table-shaped, table-sized, but otherwise perfectly ordinary frosty white block of ice. Of course, it needs to be added that the observers in w are spectrum-inverted with respect to observers here, so that the qualitative appearances they enjoy in front of a frosty white object are just like the ones we enjoy when looking at an otherwise similar brown object. But if both of those changes are made at once, then the experience of observers there looking at their table is just like the experience we enjoy looking at ours.<sup>10</sup>

---

<sup>9</sup> Or, if that isn't neutral enough, let the condition be not that observers in w enjoy qualitatively identical appearances, but that they enjoy qualitative appearances XYZ. I will ignore this complication.

<sup>10</sup> Shmipke concedes the possibility of spectrum inversion.

Note that there is some slight support for Schmipke's position in the text. Kripke says that what the icy table intuition comes to is that "I (or some conscious being) could have been qualitatively in the same epistemic situation that in fact obtains, etc." He doesn't say the conscious being has to resemble me in any important respect. The counterfactual being's brain might be wired so that it is in the same qualitative state standing in front of an icy table as I am standing in front of a wooden one. So contrary to what we said above, it could be that Kripke is imposing only the neutral condition of icy table, appearances XYZ.

The question is, does the revised explanation meet the psychoanalytic standard? Does it explain our illusion that this table could have turned out to be made of ice, to point out that had our brains been different, a regular icy table would have caused in us the same qualitative state that a wooden table does cause in us? I tend to think it does not. Because what seems possible is that this table with relevant perceptible properties held fixed could have turned out to be ice. No one is going to be tempted into thinking that possible by reflection on the possibility that we see a regular icy table as brown, because in that scenario the perceptible properties change. The color of the table goes from brown to white.<sup>11</sup>

It may help to consider an analogy. Say that I am under the impression that that animal there [pointing] is a zebra, when really it is a horse. Dretske's explanation is "The horse is painted to look just like a zebra. When two things look just the same, the one is easily mistaken for the other. It makes sense then that you would take this horse for a zebra." That corresponds to the Kripkean explanation of the "could have turned out to be ice" illusion. Because the table's appearance is indistinguishable from that of disguised ice, one naturally concludes that it could be, or have been, disguised ice.

Imagine now a second, Schmetzkean explanation of my zebra illusion. "The horse is not painted at all. And you're enjoying ordinary horsy phenomenology. But there is a guy counter-Steve, a counterfactual

---

<sup>11</sup> A property is perceptible iff when an object perceptually appears to have it and doesn't, we have misperceived. Not all properties figuring in the content of a perceptual state are perceptible in this sense. Our experience may represent the table as wooden, but it's not as if our eyes are playing tricks on us if it's well-disguised ice.

variant of yourself, who has zebraish phenomenology when looking at a horse, and horsy phenomenology when looking at a zebra. Because your phenomenology is indistinguishable from that of counter-Steve looking at a zebra, it makes sense that you would take this horse for a zebra." That corresponds to the Sch mipkean explanation of the "could have turned out to be ice" illusion. Because my actual table phenomenology is indistinguishable from my alter ego's ice phenomenology, I am led to suppose that this table could be, or have been, a regular old hunk of ice.

Is it just me, or does the first pair of explanations work better than the second? "I am liable to confuse A with B because they look the same to me" sounds quite plausible. If things look the same, then one is indeed liable to confuse them. "I am liable to confuse A with B because the same looks result if it is me looking at A or counter-Steve looking at B." There is no chance at all that I am confusing myself with counter-Steve, even if his phenomenology is just the same. Counter-Steve is by definition a person who sees things differently than I do. (One might just as well worry that our planet has all along been Twin-Earth, making water not H<sub>2</sub>O but XYZ.)

So we have the following principle: to explain why this, 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 or fool's gold principle.

\*\*\*\*\*

Kripke gives two models for the explaining-away of the intuition that A could be Q. First is the reference-fixer model:

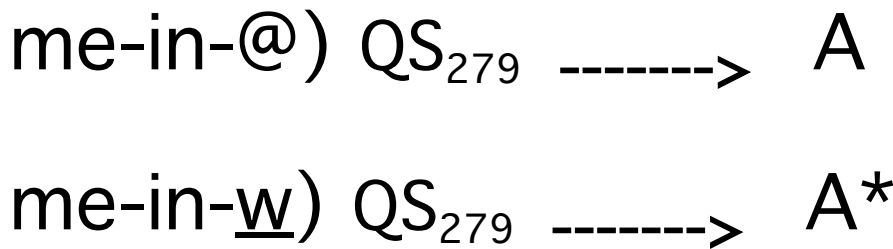
(RF) it seems possible for A to be Q because it really is possible that the so and so is Q, where "the so and so" is a descriptive condition fixing "A"'s reference.

Then there is the epistemic counterpart model:

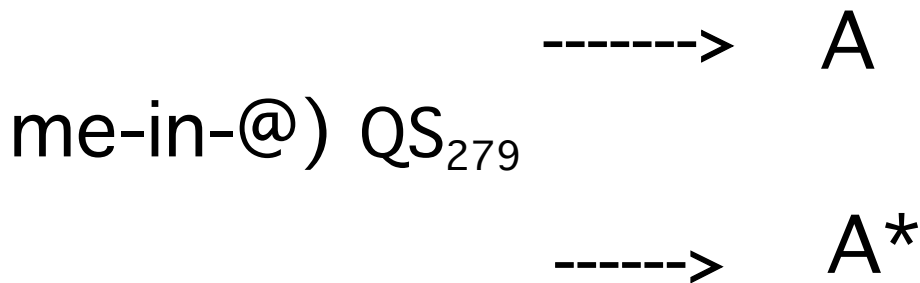
(EC) it seems possible for A to be Q because it really is possible for A\* to be Q, where A\* is a facsimile of A.

The epistemic counterpart model might seem the more accommodating of the two, because it does not require anything in the way of reference-fixing descriptions. But there is a respect in which the reference-fixing model is more accommodating and indeed too accommodating.

The epistemic counterpart model requires an A\* indiscernible in relevant respects from A, what we have called a facsimile of A. Can this requirement be enforced by asking A\* to satisfy some carefully constructed reference-fixing description D? It is not at all obvious that a suitable D can be found. One obvious possibility is "the thing that puts me into qualitative state 279." The picture this gives is:



Here we have dissimilar observers in distinct worlds confronting two (perhaps readily distinguishable) objects and reacting the same way. (EC) by contrast envisages a single observer confronting two objects to which she responds identically:



Perhaps we can arrange for the second picture by letting D be the "the thing that puts me as I actually am into qualitative state 279." But this



forgets that "the thing that actually puts me in state 279" stands in counter-Steve's mouth for A\*. We are left again with the first picture.

One could try to force the second picture by letting D be "the thing that in  $\alpha$  puts me into state 279," where  $\alpha$  is a stable designator of actuality; it picks out our world @ no matter in which actual or counterfactual context it is uttered. But the point of a reference fixing description is that it is supposed to be a piece of language that directs us to the referent across a range of counterfactual situations. And the term "whatever in  $\alpha$  puts me into state 279" is not even understandable in counterfactual situations. Had things been different, we would not have been thinking, "too bad things are so different here, how much better to live in a non-counterfactual world like  $\alpha$ ."

Two pictures have been sketched of how to explain away modal illusions. Which of the two is meant to apply in the case of the icy table? Passages like "I (or some conscious observer) could have been in qualitatively the same epistemic situation" (1980, 142, emphasis added) suggest the first picture. But there are also passages like this:

.... it seems to me that this is not to imagine this table as made of wood or ice, but rather it is to imagine another table, resembling this one in all external details, made of another block of wood, or even of ice (1980, 114, emphasis added).

"Resembling in all external details" means, I take it, that we would not notice if the one table were instantaneously substituted for the other. And that is the second picture. The reason this matters is, once again, that the first picture fails to explain the illusion. It defies credulity that my feeling that this table could have been made of ice is based on the fact that my brain could have been such that suitably carved ice elicited in me the present sort of appearances.

\*\*\*\*\*

But let us not dwell too long on the icy table example, since Kripke uses it mainly for illustration. His real interest is in the kind of modal illusion that arises in science. Here is some heat; is it some type of molecular

energy?<sup>12</sup> One has to do conduct further tests, and like any tests, they could come out either way. So there is the appearance that heat could turn out to be a certain type of molecular energy, and the appearance that it could turn out to be something else. The second appearance is an illusion. How does Kripke propose to account for it?

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 (1980, 133).

It might be, for instance, that due to differences in our neural structures high mean molecular energy -- henceforth HME -- felt cold, and low mean molecular energy -- henceforth LME -- felt hot. Does this explain in a psychoanalytically satisfying way our feeling that it could have been LME that was heat rather than high? Does pointing to possible differences in our neural structures explain why this cold seems like it could have turned out to be HME?

Here is the worry. With the table, remember, what seemed possible is not only that ice could have paraded itself in front of someone or other who saw it as wood, but that there could have been ice that I with my existing sensory faculties 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 LME is not just that it could have paraded itself in front of someone or other who felt it as hot, but that I with my existing neural structures could have found it to be hot. To explain that seeming, we need a counterfactual facsimile of heat that turns out on closer inspection to be LME. There should in other words be the possibility of LME-type fool's heat. Similarly, to explain the seeming possibility of cold turning out to be HME, we would need the possibility of fool's cold that was found by scientists to be HME.

Is there fool's heat of this type, or fool's cold? I do not 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 LE in for HE and have it feel just the same. Having substituted low ME for high, there is no way to preserve

---

<sup>12</sup> Like Kripke I will run heat together with temperature.

the appearances but to postulate observers who react differently than ourselves to the same external phenomena. But then what we are getting is not really fool's heat but something more like dunce's heat. 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 don't know. But what does seem clearly impossible is for LME to be fool's heat, because it by hypothesis feels the opposite of hot; it feels cold.

\*\*\*\*\*

Kripke is right, or anyway I am 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. LME could, it seems, 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 this property is, I suspect, necessary. There would seem to be 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.

Someone might say: that LME can't be fool's heat doesn't show that there can't be fool's heat at all. Surely there is something in some faraway world that although not HME feels or would feel hot to us as we are. Suppose that is so (Kripke actually questions it<sup>13</sup>), and call the

---

<sup>13</sup> Kripke actually discusses something like this in NN.

"Some people have been inclined to argue that although certainly we cannot say that sound waves 'would have been heat' if they had been felt by the sensation which we feel when we feel heat, the situation is different with respect to a possible phenomenon, not present in the actual world, and distinct from molecular motion. Perhaps, it is suggested, there might be another form of heat other than 'our heat', which was not molecular motion; though no actual

something ABC ("alien basis caliente"). ABC is all you need to explain the illusion that heat could have been other than HME in the approved Kripkean fashion, that is, in terms of a genuine underlying possibility.

But, granted that one can explain, or try to explain, the illusion in this way, would the explanation be correct? I am not sure that it would, for the following reason. Our feeling that heat could have turned out to be something else is indifferent to whether the something else is alien ABC or actual LME. It would be very surprising if the feeling had two radically different explanations depending on the precise form of the something else. The LME form of the illusion cannot be explained by pointing to a possible facsimile of heat that really is LME. (Whether LME can be fool's heat is a factual question and the answer is that it can be at best dunce's heat.) Therefore the ABC form of the illusion ought not to be explained with a possible facsimile either.

I have been arguing that strong epistemic counterparts, or facsimiles, are needed to explain illusions of possibility. However there are some illusions to which epistemic counterparts, strong or weak, might seem altogether irrelevant. It seems possible not only that heat could have failed to be HME, but also that HME could have failed to be heat. Kripke treats the latter illusion as reflecting the genuine possibility that HME might not have felt hot. Given that epistemic counterparts don't figure here at all, the insistence that any epistemic counterparts should be strong may seem to leave Kripke's explanation untouched.

Once again, I appeal to the principle that similar intuitions should receive similar explanations. Our intuition that HME could have turned out to be something other than heat differs only in specificity from the intuition that it could have turned out to be cold. Weak epistemic counterparts of cold are of no use in explaining the latter illusion; it doesn't matter what "those people" (the residents of w) think. But if otherworldly

---

phenomenon other than molecular motion, such as sound, would qualify. Although I am disinclined to accept these views, they would make relatively little difference to the substance of the present lectures. Someone who is inclined to hold these views can simply replace the term... 'heat' with ... 'our heat'... (p.130, note 68)

observers are irrelevant here, then they are irrelevant to the unspecific intuition as well.

The upshot is that if S is a sensed phenomenon like heat, and P is a physical phenomenon like LME, then otherworldly observers are no use in explaining either why S seems like it could have been other than P, or why P seems like it could have been other than S. Since, as we have seen, actual observers cannot explain these apparent contingencies either, it seems that there is no psychoanalytically satisfying explanation in Kripke for the appearance that S is only contingently related to P.

But, someone might say, this just shows we have been going about it the wrong way around. Rather than looking for a strong epistemic counterpart of heat that is LME, we should be looking for a strong epistemic counterpart of me to whom LME feels hot.

I do not deny that such a person is possible; the question is what he can do for us. It seems not an accident that the intuitions explained by facsimiles of the table are intuitions about what is possible for the table. Likewise the intuitions explained by gold-facsimiles are intuitions about gold, e.g., that it could have turned out to be iron pyrites. One would expect, then, that the intuitions explainable by reference to me-facsimiles are in the first instance intuitions about me. Am I the sort of person who has heat sensations in response to HME, or the sort of person to whom LME feels hot? There is the feeling (suppose for argument's sake that it is an illusion) that I could have been the second sort of person. How does this feeling arise? Well, a possible strong epistemic counterpart of mine does have heat sensations in response to LME.

But it is one thing to explain apparent de re possibilities for ourselves, another to explain apparent de re possibilities for heat. When we ask, "did heat have to be HME or could it have been LME?", and answer that it could have turned out either way, we are caught between two seeming possibilities for heat. The proof of this that the seeming possibility of heat being LME does not depend in the least on there being Steve-like beings around to whom LME feels hot. (It seems possible indeed that heat's being LME creates conditions inhospitable to life.) The intuition that heat could have been LME although there was no one around to realize it cannot be explained by pointing to a possible me-facsimile reacting differently to LME, simply because it is stipulated in the intuition that no observers are present.

\*\*\*\*\*

Here is the position so far. It is not hard to disguise a genuinely icy table so that it looks wooden. So if Kripke wants to explain the seeming possibility of this table A being made of ice, he has at his disposal a facsimile A\* of the table that really is made of ice. Sometimes, though, the appearance is closer to the reality, and facsimiles of A are no more capable of possessing the seemingly possible property Q than A is itself. How the second sort of illusion arises is an interesting question, but a question for another paper.<sup>14</sup> The claim for now is just that we cannot explain the second sort of illusion by pointing to a world where an A-facsimile really is Q, because such a world is not possible.

Kripke says, "perhaps we can imagine that, by some miracle, sound waves somehow enabled some creature to see. I mean, they gave him visual impressions just as we have, maybe exactly the same color sense. We can also imagine the same creature to be completely insensitive to light (photons). Who knows what subtle undreamt of possibilities there may be?" (NN, 130). He asks, "Would we say that in such a possible world, it was sound which was light, that these wave motions in the air were light?" He says no, "given our concept of light, we should describe the situation differently" (1980, 130).

I agree. The indicated world does not testify to the genuine possibility of light being pressure waves in the air. But now let us ask a slightly different question. Does it explain the seeming possibility of light having turned out to be waves in the air? Again the answer is no. For that you would need sound to be a facsimile of light. And it isn't, for the obvious reason that airwaves do not look the least bit like light. But then what does explain the seeming possibility of light turning out to be compression waves in the air? I am not going to comment on that. What we do know is that the explanation is not in terms of a genuinely possible strong epistemic counterpart.

One further example, the first not taken from Kripke. Suppose that Q is a broadly geometrical property our concept of which is recognitional. Q

---

<sup>14</sup> I suspect that the explanation is often as simple as this: there is a facsimile of A that might for all we know a priori be Q.

might be the property of being jagged, or loopy, or jumbled. It might be the property of "leftiness," which we recognize by asking if the figure in question appears to be facing left (in the manner of 'J' and '3'), or right (in the manner of 'C' and '5'). I will focus for no particular reason on the property of being oval. Everyone knows how to recognize ovals, but nobody knows the formula (for there is no formula to know). The one and only way to tell whether something is oval is to lay eyes on it and see how it looks. A thing is judged oval iff it looks more or less the shape of an egg.

Now suppose I tell you that cassinis are the plane figures, whatever they may be, defined by the equation  $(x^2 + y^2)^2 - (x^2 - y^2) = 5$ . Is being a cassini a way of being oval? I take it that until you do the experiment, this is an empirically open question. Cassinis could turn out to be oval or they could turn out not to be. You need to draw the figure and see how it strikes you.<sup>15</sup>

This seems not too different, intuitively, from the way LME needs to be sampled to determine whether or not it is heat. Presumably the Kripkean will want to give the same sort of explanation. Just as there are worlds where HME feels hot and worlds where it feels cold, there are worlds where cassinis look egg-shaped and worlds where they look to be shaped like bunny ears or figure-8s.

But this is all a mistake, since for cassinis to look other than egg-shaped to us as we are is impossible. There may perhaps be counterfactual observers who due to their greater visual acuity are bothered by departures from the exact profile of an egg that we ourselves hardly notice. To them, cassinis do not look egg-shaped. But those observers can no more explain the seeming possibility of cassinis' turning out not to be oval, than spectrum-inverted observers can explain the seeming possibility of the table's being made of ice. This is because what seems possible (until we do the experiment) is that cassinis look other than egg-shaped to us as we are, with our existing sensory endowment.<sup>16</sup>

---

<sup>15</sup> Cassinis as I have defined them are oval. (They belong to the class of "cassinian ovals" -- lots of whose other members are, however, not egg-shaped at all.)

<sup>16</sup> It is not as easy as one might think to throw the facsimile requirement over as too onerous. If the appearance that A could be Q is sufficiently

\*\*\*\*\*

What is the bearing of all this on Kripke's arguments against the mind-body identity theory? Kripke holds that any supposed identities between mental states and physical ones "cannot be interpreted as analogous to that of the scientific identification of the usual sort, as exemplified by the identity of heat and molecular motion" (1980, 150). This is because the model that explains away contrary appearances in the scientific case is powerless against the appearance that pain can come apart from c-fiber firings. Which is more plausible, that the model should suddenly meet its match in illusions about pain and c-fiber firings, or that the model fails to explain away anti-materialist intuitions because those intuitions are correct?

This argument rests on a false assumption, namely that dualist intuitions, if mistaken, would be the sole holdouts against the epistemic counterpart model of illusions of possibility. The model breaks down already in scientific cases like the illusion that this heat could exist without HME (and vice versa).<sup>17</sup> One needn't know how exactly the scientific illusion arises to suspect that a similar mechanism might be behind the corresponding illusion about pain.

I do not say the cases are analogous in every respect. The disanalogy stressed by Kripke is this: Identity theorists about heat can concede the existence of a world v where HME gives rise to sensations of cold. Materialists cannot, however, concede the existence of a world w where

---

explained by noting that dunce's A can be Q, then more ought to seem possible than in fact does. It should seem, not only that this brown table could have turned out to be icy, but that it could have turned out to be icy-looking, that is, white – for there is (we're assuming) a world where white tables cause the same sort of experience as this brown table causes in me. Similarly the Eiffel Tower should seem like it could have turned out to be three feet in height. For again, a reduced Tower should present to similarly scaled-down observers the same narrow appearances as I enjoy of the real Tower here.

<sup>17</sup> One doesn't notice this because Kripke lowers the bar, dropping the facsimile requirement at precisely the point that it threatens to make a counterpart-style explanation unavailable.



c-fiber firings are not felt as pain, because not to be felt as pain is not to be pain.

But this puts the materialist at a disadvantage only if we assume that v is what it takes to explain why this cold seems like it could have been HME, and w is what it takes to explain why this non-pain – this pleasure, say -- seems like it could have had been c-fiber firings. And my claim has been that intuitions like this cannot be explained by v and w at all -- unless their HME and c-fiber firings are such as to feel the relevant ways to us as we are.<sup>18</sup>

The materialist may seem still at a disadvantage, for the following reason. How otherworldly HME feels, we know. It feels hot. But whether otherworldly c-fiber firings are bound to present as pain is not clear. Certainly if they are pain, then insofar as it is essential to pain to feel a certain way, that is how c-fiber firings are bound to feel. But what if we suppose with the dualist that c-fiber firings are not identical to mental states but cause them? The c-fiber firings in w might affect minds (ours included) differently than the c-fiber firings here.

I think we should grant Kripke that a world like w, if it existed, would explain the dualist intuition, at the same time as it verified that intuition. But that is just to say that the intuition would be well explained by w if it were correct, which does nothing to show that it is correct. The premise Kripke needs is that we still find ourselves with reason to postulate w even if we suppose for reductio that it is the identity theory that is correct; this is what supposedly makes materialism a self-undermining position. But the stronger premise, we have seen, is false. This suggests to me that Kripke's argument is not in the end successful.

Does this make me a pessimist about conceivability evidence? It certainly puts me at odds with

(O) carefully handled, conceivability evidence can be trusted, for if impossible E seems possible, then something else E is possible, such that we mistake the possibility of E for that of E.

---

<sup>18</sup> Of course there may be other reasons to think v exists, e.g., the well-attested phenomenon of the same stimulus causing different perceptual reactions in different perceivers. There are not to my knowledge any well-attested phenomena to suggest the possibility of a world like w.

But although this was called the optimistic thesis above, a better term might have been super-optimistic or Pollyannaish -- because for a type of evidence to never mislead about its proper object (the real possibility confusedly glimpsed, in this case) is exceedingly unusual and arguably unprecedented.<sup>19</sup> The thesis we want, I think, is that

(O') carefully handled, conceivability evidence can be trusted, for when impossible E seems possible, that will generally be because of distorting factors that we can discover and control for.

---

<sup>19</sup> Berkeley suggests a similarly Pollyannaish thesis about perception in Three Dialogues Between Hylas and Philonous.

Hylas: What say you to this? Since, according to you, men judge of the reality of things by their senses, how can a man be mistaken in thinking the moon a plain lucid surface, about a foot in diameter; or a square tower, seen at a distance, round; or an oar, with one end in the water.

Philonous: He is not mistaken with regard to the ideas he actually perceives; but in the inferences he makes from his present perceptions. Thus in the case of the oar, what he immediately perceives by site is certainly crooked; and so far he is in the right. But if he thence conclude, that upon taking the oar out of the water he shall perceive the same crookedness...he is mistaken...his mistake lies not in what he perceives immediately and at present, (it being a manifest contradiction to suppose he should err in respect of that) but in the wrong judgment he makes concerning the ideas he apprehends to be connected with those immediately perceived (3<sup>rd</sup> Dialogue).

Where the Kripkean super-optimist treats seeming failures of imagination as failures of interpretation, the Berkeleyan one shifts the blame rather from experience to inference. The insistence that there are severe, a priori discoverable, limits on our liability to mistakes about a subject matter often goes hand in hand with idealism about that subject matter. This seems to me a further reason not to associate Kripke with the super-optimistic thesis (O).

Kripke's first great contribution to conceivability studies was to have seen the need for a technology of modal error detection in the first place. His second great contribution was to have made a start at developing this technology. There is no need to foist on him a third contribution of identifying the one and only way modal illusions can arise.

## Bibliography

Berkeley, George (1979). Three Dialogues between Hylas and Philonous (Indianapolis, Indiana: Hackett Publishing Company)

Della Rocca, Michael (1996). "Essentialists and Essentialism," Journal of Philosophy 93: 186-202

Della Rocca, Michael (2002). "Essentialism versus Essentialism," in Tamar Szabó Gendler and John Hawthorne, ed., Conceivability and Possibility (Oxford: Oxford University Press): 223-252

Hill, Chris (1997). "Imaginability, Conceivability, and the Mind-Body Problem," Philosophical Studies 87: 61-85

Kripke, Saul (1977). "Identity and Necessity," in Stephen P. Schwartz, ed. Naming, Necessity, and Natural Kinds (Ithaca: Cornell University Press)

Kripke, Saul (1980). Naming and Necessity (Cambridge, MA: Harvard University Press)

Nagel, Thomas (1974). "What Is It like to Be a Bat?," Philosophical Review 83: 435-450

Nagel, Thomas (1986). The View from Nowhere (Oxford: Oxford University Press)

Yablo, Stephen (1993). "Is Conceivability a Guide to Possibility?," Philosophy and Phenomenological Research 53: 1-42

Yablo, Stephen (2002). "Coulda, Woulda, Shoulda," in Tamar Szabó Gendler and John Hawthorne, ed., Conceivability and Possibility (Oxford: Oxford University Press): 441-492