

## Thinking about Water

James Pryor  
Princeton University  
<[jimpryor@princeton.edu](mailto:jimpryor@princeton.edu)>  
Old Draft (10/16/00)

### Chapter 1 Introduction

You know the story. There's this place Twin Earth which is just like here except that, there, the clear drinkable liquid that falls from the sky and fills the lakes and rivers is made of XYZ rather than H<sub>2</sub>O. According to robust and widely-shared intuitions, when you and your twin have beliefs you'd each express by saying "Water puts out fires," your beliefs are about different substances. Your belief concerns water, which is essentially made of H<sub>2</sub>O. Your twin's belief concerns the look-alike substance on Twin Earth. Hence the belief you express by saying "Water puts out fires" has a **wide** or **Twin-Earthable** content: it's possible for someone to be just like you intrinsically but fail to have beliefs with that content.

This story stirs up a hornet's nest of philosophical problems. How pervasive are these wide or Twin-Earthable contents? For instance, do the beliefs you express using color terms like "red" also have Twin-Earthable contents? Do your perceptual experiences have Twin-Earthable contents? Can Twin-Earthable contents play a genuine role in the explanation of your behavior? What happens when you attempt to form a thought with a Twin-Earthable content but there's nothing in your environment of the appropriate sort for your thought to be about?

Those are all good questions. But one can only do so much at one time. Our focus in this inquiry will be on the relations between Twin-Earthable contents and introspective self-knowledge. Our target will be questions like these: If the contents of your thoughts are Twin-Earthable, then how are you able to tell what you're thinking by introspective reflection alone? Doesn't what you're thinking depend on what your environment is like? And how could you know such matters by introspection alone? It's hard to see how introspection could even enable you to know that two of your thoughts have *the same* content, if those contents are Twin-Earthable. For instance, how could you tell that the

thoughts you expressed in your childhood using “water” have the same content as the thoughts you now express using “water”? How do you know you haven’t been kidnapped and taken to Twin Earth, where your thoughts slowly shifted in content to match the thoughts of Twin Earth’s native inhabitants?

One way of answering these questions just concedes that you would not have any privileged or introspective access to the Twin-Earthable contents of your thoughts. According to this view, you can only have introspective access to **narrow** facts about your thoughts, facts which supervene on what you’re like intrinsically. This is an **Incompatibilist** view about the relation between Twin-Earthable contents and introspective self-knowledge. Another approach to these questions, suggested by Putnam in Ch. 1 of *Reason, Truth, and History*, insists that you *can* have introspective access to the contents of your thoughts, even when those contents are Twin-Earthable—and *what’s more*, this view continues, since philosophy tells us you can only have thoughts with those contents when your external environment is configured in certain ways, it follows that you can ascertain on the basis of introspection and *a priori* philosophical reasoning alone that your environment is configured in the relevant ways. For instance, since having thoughts about vats requires that you not always have been a brain in a vat, and since you do have thoughts about vats, a little bit of introspection and philosophy are all that’s needed to ascertain that you have not always been a brain in a vat. No *perceptual* knowledge is required to reach this conclusion; it follows just from philosophical reasoning and introspective knowledge of what thoughts you’re having.

The Incompatibilist views about Twin-Earthable contents and introspective self-knowledge is counter-intuitive, and so is the Putnam-inspired view. Intuitively, we’d like a story intermediate between those two views. We’d like a story which gives us introspective knowledge that we’re thinking about water, but which denies that we can know what our external environment is like by introspection and *a priori* reasoning alone. It remains to be seen whether any such intermediate view is really available.

These problems about the relation between Twin-Earthable contents and introspective self-knowledge, together with the problems mentioned earlier, prompt some philosophers to postulate a range of narrow facts about our thoughts. They say it is these narrow facts which really do the work in explaining our behavior, and which are the

proper objects of our introspective self-knowledge. Putting their view crudely, what you *really* believe when you say things like “Water puts out fires” is something narrow, something along the lines of “The clear drinkable liquid around here puts out fires.” That’s what your twin on Twin Earth also believes, *really*. It may be convenient for semantics and some other philosophical purposes to talk about differences between your beliefs—e.g., the fact that your beliefs stand in certain causal relations to distinct chemical substances—but it’s really the common, narrow characterization of your beliefs which is important for explaining your behavior, and which is what you’re aware of through introspective reflection.

This kind of view is often supported by drawing an analogy between Twin-Earthable thoughts and indexical thoughts. Suppose you and I each have beliefs we’d express by saying “My pants are on fire.” In one sense our beliefs have different contents: they are beliefs about different people. That may prompt us to call them different beliefs. But there’s another sense in which our beliefs seem to be the *same* belief. There’s a sense in which the way the world seems to me, when I believe my pants are on fire, is the same way that the world seems to you, when you believe that your pants are on fire. Plausibly, it’s the fact that our beliefs are “the same” in this second sense which is important for explaining why we each act in the way we do. Now consider again you and your twin on Twin Earth, who each have beliefs you’d express by saying “Water puts out fires.” Here too, there’s *a* sense in which your beliefs seem to be different. They are beliefs about different chemical substances. But perhaps there’s also another sense in which your beliefs are the *same* belief, a sense in which you’re both taking the world to be the same way. If *there is* a sense in which you and your twin express the same beliefs with “water,” then it will be narrow facts about your beliefs which make them “the same” in that sense. It is those narrow facts which really characterize the way the world seems to you, from the inside.

Jackson and Chalmers have recently advocated a view about belief which is inspired by this kind of reasoning. Jackson and Chalmers like the idea that what you *really* believe when you say things like “Water puts out fires” is something like “The clear drinkable liquid around here puts out fires.” They would welcome their view’s ability to avoid the problems we raised about self-knowledge. However, as we’ll see, they

don't directly argue for their view from those problems. On the face of it, their arguments come from a quite different direction.

Stories about Twin Earth are used to teach two sorts of lessons. One lesson I've already mentioned: it's that the contents of some thoughts and assertions can vary between twins who inhabit different environments. In other words, those contents are Twin-Earthable. You think about water, but your twin's thoughts are not about water, but about a different substance. The second lesson is that there can be necessary truths which are knowable only *a posteriori*. The reason your twin isn't thinking about water is that the stuff in his environment is made of XYZ rather than H<sub>2</sub>O. According to robust and widely-shared intuitions, it's *necessary* that something be made of H<sub>2</sub>O if it's to count as water. But this fact is of course knowable only *a posteriori*.

In Jackson and Chalmers' eyes, the existence of necessary *a posteriori* truths like these are deeply puzzling.

In part this puzzlement arises from their conviction that all necessity is ultimately grounded in conceptual entailments, and all knowledge of necessity grounded in something like conceptual analysis. How could *empirical* methods of investigation give us insight into what's necessarily the case? It's not as if our senses enable us to peer into all the other possible worlds. To be sure, we might learn some necessary truths on the basis of other people's testimony, or other *a posteriori* methods of that sort. But it seems like, if something is necessarily so, and it's knowable at all, then it would *ultimately* have to be *a priori* methods of investigation which enable us to know it.

Another cause of puzzlement is the fact that we can conceive of situations in which water is *not* made of H<sub>2</sub>O. This makes the connection between water and H<sub>2</sub>O *seem* contingent. It seems like water might not have been made of H<sub>2</sub>O. Even if we *agree* that the connection between water and H<sub>2</sub>O is in fact necessary, this appearance of contingency persists. It's a kind of illusion. Our ability to conceive of situations in which water is not made of H<sub>2</sub>O in some way tempts us to (falsely) think of the connection between water and H<sub>2</sub>O as contingent. How is this illusion to be explained? And how are we supposed to tell, when we find ourselves conceiving some counterfactual scenario, whether that scenario is genuinely possible or not?

Jackson and Chalmers use these and other puzzles about the necessary *a posteriori* to support their view about belief. As we'll see, their view makes precise the crude idea that what you really believe when you say things like "Water puts out fires" is something like "The clear drinkable liquid around here puts out fires."

Part I of this essay will introduce Jackson and Chalmers's framework for explaining the necessary *a posteriori*, and will examine their reasons for adopting it.

Part II will argue that the fundamental philosophical assumptions of that framework are not sustainable. So we need a different account of what goes on when one has, and knows oneself to have, a thought with a Twin-Earthable content.

Although we end up rejecting Jackson and Chalmers's framework, our study of it and the reasoning which motivates it will help us better understand the nature of these Twin-Earthable thoughts, and the relations between necessity and epistemic notions like conceivability and *a posteriori*. Most importantly, we will see that much of the false reasoning that drives Jackson and Chalmers's framework also drives the Incompatibilist and Putnam-inspired views about Twin-Earthable contents and introspective self-knowledge. The arguments for these views start in different places, but in the end it is a common way of thinking about Twin-Earthable contents which keeps them alive. Seeing what's wrong with that way of thinking is essential if we're to develop a sober and modest account of Twin-Earthable thoughts and our introspective access to them. This will be the focus of Part III.

**PART I**  
**Kripke's Strategy and the 2D Framework**

## Chapter 2 Explaining Illusions of Contingency

You can readily conceive of situations in which water is made of something other than H<sub>2</sub>O. Clearly, you could conceive of such situations *before* you knew that water *is* made of H<sub>2</sub>O. But I think that even *after* learning that water is made of H<sub>2</sub>O, and that it's *necessarily* made of H<sub>2</sub>O, you can continue to conceive of its being made of something else. Let me explain.

Some of the things you know, you know only defeasibly. Your knowledge can be defeated by additional evidence. That is, you might start off with enough justification to count as knowing *p*, and then go on to acquire additional evidence which leaves you *unjustified* in believing *p*. Hence, you'd no longer be in a position to know *p*.

For instance, skeptical arguments notwithstanding, you really do know that you have hands. But this knowledge is defeasible. Even though you know you have hands, you could wake up tomorrow and start acquiring evidence that makes it reasonable for you to believe (falsely) that you are a handless brain in a vat. A ticker tape might start running across the bottom of your visual field with the words "You are a handless brain in a vat..." Or you might hear what seem to be the disembodied voices of the scientists who are manipulating your brain. As a matter of fact, you are *not* a brain in a vat; these experiences are just the product of some systematic hallucination. But they might nonetheless defeat your justification for believing that you have hands. Such is the fate of defeasible knowledge.

Your knowledge that water is made of H<sub>2</sub>O is also defeasible in this way. Even if you know that water is made of H<sub>2</sub>O, and that it's *necessarily* made of H<sub>2</sub>O, this knowledge is defeasible. There are possible courses of experience you could have, starting tomorrow, which would make it reasonable for you to give up your (true) belief that water is made of H<sub>2</sub>O. You may not have any firm or definite idea of what form that evidence would take. But insofar as you are rational, you will be "sensitive" to the defeasible nature of your evidence for believing that water is made of H<sub>2</sub>O, and prepared to revise this belief if future evidence should so demand. The ability to *conceive of* a situation in which water is not made of H<sub>2</sub>O is a natural accompaniment to that kind of sensitivity.

So even if your evidence enables you to know that water *is* made of H<sub>2</sub>O, and that it's necessarily made of H<sub>2</sub>O, there's still a sense in which your evidence "leaves open" the possibility that it's made of something else.<sup>1</sup> That makes the connection between water and H<sub>2</sub>O *seem* contingent, in some sense—even though you know it's not contingent. We have here an "illusion of contingency." It seems like things could turn out either way. Of course, you need not be taken in by this illusion. But its mere existence is a puzzle. How are such illusions of contingency to be explained? What is the difference between an illusion of contingency, of this sort, and a genuine modal intuition?

---

<sup>1</sup> Some would say these are cases where it's *epistemically possible* that water is made of something other than H<sub>2</sub>O. (This is only one possible construal of "epistemic possibility"; there are others.) For a number of reasons, I prefer to avoid the use of possible worlds and related modal apparatus to characterize evidential situations. So I will not use locutions like "epistemic possibility."

Notice that nothing in my account of "conceivability" turns on the fact that your knowledge that water is made of H<sub>2</sub>O is *a posteriori*; it only turns on the fact that your knowledge is defeasible. If your *a priori* knowledge of some matters is also defeasible, then the contraries of what you know in those cases could also be "conceivable," in the sense I've explained. For instance, when Wiles first thought he had a proof of Fermat's Last Theorem, his grasp of the "proof" gave him some *a priori* justification for believing that Fermat's Last Theorem is provable. (Of course, he also had *a posteriori* justification for believing that the Theorem is provable, based on his partly empirical justification for believing that he had himself proven it. But it's his *a priori* justification for believing that the Theorem is provable that we're concerned with here.) Then Wiles discovered a flaw in his "proof." At this point, his original justification for believing that Fermat's Last Theorem is provable was defeated. This shows that, even before the flaw was discovered, it had to be conceivable to Wiles that Fermat's Last Theorem is not provable, in the sense I've explained.

In discussions of the necessary *a posteriori*, one often sees a more restricted notion of conceivability being employed. That notion might be called "ideal conceivability." Something counts as ideally conceivable iff it's conceivable in the sense I've already articulated, and it would remain conceivable in that sense no matter how much additional (good) *a priori* reasoning you engaged in. This seems to be the notion of "conceivability" that Chalmers is employing, for instance. It would be ideally conceivable in this sense that Fermat's Last Theorem is unprovable only if a fully-informed ideal *a priori* reasoner would still regard it as an open possibility that the Theorem is unprovable. Since there *is*, after all, an *a priori* proof of Fermat's Last Theorem, the ideal reasoner would know that the Theorem is provable. So it would *not* be ideally conceivable that Fermat's Last Theorem is unprovable (even for us, less-than-ideal reasoners). It is not clear to me that this notion of ideal conceivability has much theoretical utility.

[We have to be careful in characterizing the notion of ideal conceivability. I believe that *Gone With the Wind* won the Oscar for Best Picture for 1939. This is a straightforwardly empirical belief, and its denial should presumably count as both conceivable in my sense and ideally conceivable. But imagine a bizarre epistemology, which allows no future evidence to count against one's current empirical beliefs. Philosophers have an uncanny knack for arriving at all sorts of bizarre views by *a priori* reasoning. So suppose I reasoned myself into accepting that epistemology. *If I did so*, I would deprive myself of the ability to conceive of its turning out that *Gone With the Wind* did not win Best Picture for 1939. And on the characterization of "ideal conceivability" I gave, the mere possibility that I might do that shows that it's *now* ideally inconceivable that *Gone With the Wind* not have won Best Picture for 1939, just as it's ideally inconceivable that Fermat's Last Theorem not be provable. This is an unfortunate result. Presumably facts about what I would and wouldn't be able to conceive *after a priori* reasoning myself into accepting bizarre epistemologies should be irrelevant to the question what's *now* conceivable for me. I do not know what a better characterization of ideal conceivability would look like.]

Kripke had a Strategy for explaining these illusions of contingency. Consider the apparent intuition that Hesperus could have turned out to be distinct from Phosphorus. Kripke says that you do have a genuine modal intuition here, but it's not really an intuition that Hesperus could have been distinct from Phosphorus. Rather, it's an intuition that it's possible for someone to *have the evidence you have* concerning Hesperus and Phosphorus, and be related thereby to two distinct objects. That really is possible. We just somehow mistake an intuition of that possibility for an intuition that Hesperus itself could have turned out to be distinct from Phosphorus.<sup>2</sup> Kripke writes:

There certainly is a possible world in which a man should have seen a certain star at a certain position in the evening and called it "Hesperus" and a certain star in the morning and called it "Phosphorus"; and should have concluded—should have found out by empirical investigation—that he names two different stars, or two different heavenly bodies... And so it's true that given the evidence that someone has antecedent to his empirical investigation, he can be placed in a sense in exactly the same situation, that is a qualitatively identical epistemic situation, and call two heavenly bodies "Hesperus" and "Phosphorus," without their being identical. So in that sense we can say that it might have turned out either way...<sup>3</sup>

What is it for a subject to be in an epistemic situation qualitatively the same as yours? Kripke's general story seems to be that your counterpart has *all the same sensory evidence* as you have, although his evidence derives from objects that are different than the objects in your environment from which your evidence derives. Sometimes Kripke works with a simpler and more specific model of your counterpart's epistemic situation. This more specific model doesn't require your counterpart to share all your sensory evidence, but only the evidence which *leads you (??) to fix the reference* of terms like

---

<sup>2</sup> As I describe matters, we *are* genuinely able to *conceive of* a situation in which Hesperus is distinct from Phosphorus. That's a result of the fact that our evidence for believing that Hesperus is (necessarily) Phosphorus is defeasible. So we're not mistaken about the content of *our conception*. It's just that, in virtue of being able to conceive that, *we seem to have a modal intuition* that a certain situation is metaphysically possible. This is where Kripke says we go awry. On his account, we do have a genuine modal intuition, but we're mistaken about *what situation* we're intuiting to be possible.

Kripke mentions two ways of using "It could have turned out that P." The first use means, roughly, that P is conceivable (given one's evidence); that, of course, is compatible with P's being metaphysically impossible. (See Kripke 1980, p. 103; compare p. 143n72.) On the second use, "It could have turned out that P" entails that P genuinely *is* possible. (See Kripke 1980, pp. 141-2.) Kripke primarily uses "It could have turned out that P" in the second way; I will follow him in this.

<sup>3</sup> Kripke 1980, pp. 103-4; see also pp. 141-4, 150-1; and Kripke 1971, p. 93n15, 97-8, and 100n18.

“Hesperus” and “Phosphorus” in the way you do. [The evidence which justifies you in believing that the reference-fixing description is satisfied at all, and any evidence you have concerning which particular objects satisfy it.] In the counterpart’s situation, it turns out that distinct objects satisfy the descriptions he uses to fix those names’ reference. Thus, according to Kripke, when you seem to intuit the possibility that Hesperus might be distinct from Phosphorus, what you’re really intuiting is that some object might satisfy the description you (in fact) use to fix the reference of “Hesperus,” and some distinct object satisfy the description you (in fact) use to fix the reference of “Phosphorus.” That genuinely is possible. Kripke writes:

Our general paradigm [was] to redescribe both the prior evidence and the statement qualitatively and claim that they are only contingently related. In the case of identities, using two rigid designators, such as the Hesperus-Phosphorus cases above, there is a simpler paradigm which is often usable to at least approximately the same effect. Let “ $R_1$ ” and “ $R_2$ ” be the two rigid designators which flank the identity sign. Then “ $R_1 = R_2$ ” is necessary if true. The references of “ $R_1$ ” and “ $R_2$ ,” respectively, may well be fixed by nonrigid designators “ $D_1$ ” and “ $D_2$ ,” in the Hesperus and Phosphorus case these have the form “the heavenly body in such-and-such position in the sky in the evening (morning).” Then although “ $R_1 = R_2$ ” is necessary, “ $D_1 = D_2$ ” may well be contingent, and this is often what leads to the erroneous view that “ $R_1 = R_2$ ” might have turned out otherwise.<sup>4</sup>

It genuinely is possible for different objects to satisfy the two reference-fixing descriptions  $D_1$  and  $D_2$ . It’s just that we mistake an intuition that *that’s* possible for an intuition that it’s possible for Hesperus to be distinct from Phosphorus. That is not possible. Kripke tries to persuade us that, on reflection, we don’t have a genuine intuition that *it’s* possible. We mischaracterize what it is we really intuit.

Likewise, suppose you have fixed the reference of “water” to be the clear drinkable liquid you’ve predominantly interacted with. It is genuinely possible for stuffs other than water, stuffs not made of  $H_2O$ , to satisfy that reference-fixing condition. We might mistake an intuition that *that’s* possible for an intuition that *water itself* might not be made of  $H_2O$ . But it’s not really possible for water to fail to be made of  $H_2O$ . “Water” rigidly names a stuff which is always and essentially made of  $H_2O$ .<sup>5</sup> On Kripke’s account,

---

<sup>4</sup> Kripke 1980, p. 143-4.

<sup>5</sup> In this essay, I will assume that “water” is a rigid name of a certain kind of stuff. I *don’t* mean it’s a name for the concrete fusion of all *actual samples* of water. Rather, I mean it’s an abstract singular term

we don't have a genuine intuition that it's possible for *water* to fail to be made of H<sub>2</sub>O. We mischaracterize what it is we really intuit.

Now, just as we seem able to imagine water turning out not to be made of H<sub>2</sub>O, and heat turning out not to be molecular motion, we also seem able to imagine pain turning out not to be identical with any brain state, like C-fiber firing. We seem able to imagine pain occurring while C-fiber firing is absent, and able to imagine C-fibers firing without anyone's being in pain. Kripke thinks that it's *not* possible to explain away these apparent possibilities in the same way that he's explained such apparent possibilities as heat's turning out not to be molecular motion. He writes:

In the case of the identity of heat with molecular motion the important consideration was that although "heat" is a rigid designator, the reference of that designator was determined by an accidental property of the referent, namely the property of producing in us the sensation S. It is thus possible that a phenomenon should have been rigidly designated in the same way as a phenomenon of heat, with its reference also picked out by means of the sensation S, without that phenomenon being heat and therefore without its being molecular motion. Pain, on the other hand, is not picked out by one of its accidental properties; rather it is picked out by the property of being pain itself, by its immediate phenomenological

---

which rigidly designates the *kind* water. Natural kind terms do sometimes function as names of kinds, as in "Water is the liquid I most like to drink," and "*Homo sapiens* is the species most likely to colonize the moon." More often, natural kind terms function as mass predicates, as in "This stuff is water," "There's water in the basement," and "(All) water contains hydrogen." For *simplicity*, I will pretend that these mass predicate uses can all be understood as covert uses of the relevant abstract singular term. Thus: "This stuff is (a quantity of the stuff) water," "There's (some quantity of the stuff) water in the basement," and "(All quantities of the stuff) water contain hydrogen." I *don't* believe this is really the proper philosophical account of those mass predicate uses. However, it will greatly simplify our discussion to assume this, and I believe that, in the end, little of what I have to say turns on the proper resolution of such issues. [refer to Koslicki 1999, Scott?]

One important difference is how to understand the notion of rigidity. What it takes for a *predicate* to be rigid has until now never been satisfactorily explained. See Soames for discussion. But when we're dealing with abstract singular terms, we can just operate with the familiar notion of rigidity. A singular term designating a kind will be rigid just in case it refers to that same kind wrt every possible world in which that kind exists, and doesn't refer wrt any possible world to anything else. (Note that this sort of rigidity won't be a special property of names of *natural* kinds. The abstract singular term "sofa," for instance, will also refer wrt every possible world to the same artifactual kind. And the same holds for any proper name of a kind. Descriptions of kinds, like "Alex's favorite kind of furniture," may still turn out non-rigid.)

It is unclear to me what the relation is between the kind water and the kind H<sub>2</sub>O. So I want to leave this an open question. I won't ever say that water is *identical* to H<sub>2</sub>O. Rather, I'll say that water is necessarily *made of* H<sub>2</sub>O. This is my attempt to be neutral, for present purposes, about what precise metaphysical relation the kinds stand in to each other. Sometimes, in the interests of brevity, I may say that "water" refers to H<sub>2</sub>O. What I mean is that "water" refers to a stuff which is made of H<sub>2</sub>O.

quality. Thus pain, unlike heat, is not only rigidly designated by “pain” but the reference of the designator is determined by an essential property of the referent. Thus it is not possible to say that although pain is necessarily identical with a certain physical state, a certain phenomenon can be picked out in the same way we pick out pain without being correlated with that physical state. If any phenomenon is picked out in exactly the same way that we pick out pain, then that phenomenon *is* pain.<sup>6</sup>

That is, anything which had the properties we use to fix the reference of “pain” would thereby itself have to be pain. Pain is the only thing which can possibly have those properties. More generally:

Someone can be in the same epistemic situation as he would be in if there were heat, even in the absence of heat, simply by feeling the sensation of heat; and even in the presence of heat, he can have the same evidence as he would have in the absence of heat simply by lacking the sensation S. No such possibility exists in the case of pain and other mental phenomena. To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain; to be in the same epistemic situation that would obtain in the absence of pain *is* not to have a pain. The apparent contingency of the connection between the mental state and the corresponding brain state thus cannot be explained by some sort of qualitative analogue as in the case of heat.<sup>7</sup>

Any epistemic counterpart of pain would have to be *pain* itself. Hence, it can't be the case that we're intuiting a possible situation where some epistemic counterpart of pain, distinct from pain, occurs in the absence of C-fiber firing. Our apparent intuition that pain could turn out to be independent of C-fiber firing can not be explained away in that manner.<sup>8</sup>

---

<sup>6</sup> Kripke 1980, pp. 152-3.

<sup>7</sup> Kripke 1980, p. 151; see also Kripke 1971, pp. 98-101.

<sup>8</sup> As is often remarked (see Boyd 1980), Kripke neglects the possibility that one could be in the same epistemic situation that would obtain if there were C-fiber firing, without there being any C-fiber firing. He seems to be assuming that “C-fiber firing,” unlike “heat,” is not Twin-Earthable. Anything which counted as evidence for me that there was C-fiber firing would also count as evidence for my twins on Twin Earth that there was C-fiber firing. (At Kripke 1971, pp. 99-100, Kripke says that we “pick out” our brain states by properties that “pick out their object essentially.” What he means by this is that those properties cannot be possessed by different things in different worlds.)

However, it is controversial whether terms like “C-fiber firing” are Twin-Earthable. In addition to Boyd, see also Hirsch 1986's discussion of “structural realism”; Jackson 1998, pp. 23-4 on “Kantian physicalism”; and Chalmers 1996, pp. 134-6 on “protophenomenalism.”

It's important to recognize that Kripke's argument here threatens more than just mind-brain *identity theories*. As we just saw, if Kripke's argument is going to work at all, he needs the assumption that terms like “C-fiber firing” aren't Twin-Earthable. But with that assumption in place, his argument also tells against views which say that C-fiber firing is *metaphysically sufficient* for pain. A *prima facie* problem for

Kripke's Strategy for explaining away illusions of contingency prompts several questions. First, do we really *need* to give illusions of contingency the substantive kinds of explanation that Kripke attempts to give them? Some philosophers feel it's enough to say this, instead:

Look, we have a tendency to confuse epistemic matters with metaphysical matters. Once we see that these are different, we should recognize that conceivability is no guarantee of possibility. Of course we can *conceive of* situations in which water is not made of H<sub>2</sub>O; but it doesn't follow from that that such situations are metaphysically possible. To the extent that the illusion that they *are* possible persists, that's just a matter of our tending to fall back into our old tendencies. The flesh is weak.

This seems too short an answer. Obviously, there's *some* relation between conceivability and possibility. Otherwise, how could we come to have *any* justified beliefs about possibility? Given that there's some relation between them, we'd like to have an account of when conceivability can be relied on as a guide to possibility, and when it can't. Kripke's desire to explain away illusions of contingency looks like a laudable first step towards such an account.

But it's a further question whether the kinds of explanation Kripke offers really are good explanations. Kripke needs us to accept that we're systematically mistaken about the nature of some of our modal intuitions. This is hard to believe. What's more it raises a methodological worry. Kripke *himself* often relies on modal intuitions—e.g., the intuition that Aristotle might not have been a philosopher. If we're so liable to be mistaken about the content of our modal intuitions, then how does Kripke know that *his* intuition is really an intuition about what's possible for Aristotle?

I'm not sure what to say about these matters. It does seem to me that Kripke owes us more of a story about when we should accept modal intuitions as genuine and when it's legitimate to rephrase them or explain them away. For the time being, though, I will tentatively accept Kripke's explanations. That is, I will accept that sometimes our

---

such views is that it seems possible to us for C-fibers to fire without anyone being in pain. If neither of "C-fiber firing" and "pain" are Twin-Earthable then what we're intuiting here *can't* be the possible painless occurrence of a brain state that stands to C-fiber firing in the way that XYZ stands to water. If "C-fiber firing" is not Twin-Earthable, there is no such state. (See Kripke 1980, p. 145n74 for some remarks in this direction.)

seeming intuition that such-and-such is a genuine possibility for *Hesperus* is really an intuition about *possible bearers of the properties we use to fix the reference of "Hesperus."* This will explain some of our illusions of contingency concerning Hesperus.

But that leaves us with the question whether illusions of contingency are *always* to be explained in this way. Perhaps some *other* explanation will sometimes be appropriate; and perhaps this *other* explanation will be able to explain why mind/body connections falsely appear to us to be contingent. We're really not yet in a position to say. Some philosophers have started to develop such alternative explanatory strategies (Chris Hill). Kripke has not shown that no such alternative strategy can work.

I think we should view Kripke's argument against materialism as just posing a challenge, rather than purporting to be a decisive argument. The challenge for the materialist is to find some other good explanation of the appearance of contingency for mind/body connections. Perhaps this can be done. We'll have to wait and see.

The Strategy Kripke uses to explain away illusions of contingency is flexible and modest. It allows that, in different cases, we might want to tell different stories about what makes a subject's epistemic situation "qualitatively the same" as our own. And it doesn't claim to be the *only* way by which illusions of contingency might be explained.

As an exercise, let's try to systematize Kripke's strategy, and to strengthen its claims to comprehensiveness. We'll call the result the **Strengthened Kripke Strategy**.

The Strengthened Kripke Strategy will employ Kripke's more specific account of what makes a subject's epistemic situation with respect to Hesperus "qualitatively the same" as ours. That's the account according to which some object, possibly distinct from Hesperus, satisfies and gives evidence of satisfying the description we used to fix the reference of our term "Hesperus." (Assuming *we did* fix the reference of "Hesperus" with some description. We'll return to this assumption later.)

Let's take a moment to get clear on what it means to talk about the description we use to fix the reference of "Hesperus." We'll understand the notion of a description's **fixing the reference** of a term as follows:

- (i) the term is *stipulated* to refer, with respect to every possible world, to whatever object it is which, in the actual world, uniquely satisfies that description; and

- (ii) *competence with the term requires one to be in a position to know, just by reflection, that the term is governed by that stipulation*

This is my best attempt at making precise the notion Kripke introduced in *Naming and Necessity*.<sup>9</sup> In the literature, names which have their reference fixed in this way are called “**descriptive names**.”

There’s an important difference between *it’s now being true* that a description fixes some term’s reference and a description having been *originally* used to fix the term’s reference. Standardly, when a description is used to fix a term’s reference, the term gets passed on to later speakers in such a way that competence *no longer* requires them to know of the original reference-fixing stipulation. As we’re understanding the notion of reference-fixing, a description *currently* fixes the reference of a term only if the two conditions I described are *currently* satisfied. As Kripke argued, it is unlikely that very many names in English currently have their reference fixed by any description, in this sense. For many names, you can be a competent user of the name even if you don’t associate enough descriptive information with the name to pick out its referent uniquely, and even if the descriptive information you do associate with the name is inaccurate.<sup>10</sup>

The claim that a word has a given reference-fixing description needs to be clearly distinguished from two other claims: on the one hand, the claim that there is a certain

---

<sup>9</sup> See Kripke 1980, pp. 54-58 and 78-80. I do not mean my talk of “stipulations governing a term” to be hostage to historical questions about whether the term was introduced into the language by any stipulative act. (If you like, treat my talk of “linguistic stipulations” operationally: do terms have the semantic properties they would have had, and do they require competent speakers to know the things they would have been required to know, if the terms were introduced and governed by explicit stipulations of the sort I describe?)

In some cases, one might have a certain object in mind, and introduce a term to refer to that object with a “referential” use of a description (see Donnellan 1966). In such cases, the object one has in mind need not uniquely satisfy the description. Perhaps introducing a term in this way would make the term refer to the object one has in mind, rather than the object which uniquely satisfies the description. If so, such cases would *not* count as cases where the description “fixes the reference” of the term, in the sense defined here. (See the discussion of such cases in Kripke 1980, at pp. 24-26, 80n34, and 87n37; and see Kripke 1979, esp. §2c.)

Kripke says at several places that if one fixes the reference of a term “N” with a description “the F,” then one is in a position to know *a priori* that N is the F: see esp. pp. 56-7, 63, 78, and 79n33. I am treating this as a substantive epistemological claim, and not as part of the definition of “reference-fixing.” I will argue against this epistemological claim in Part II of the essay.

<sup>10</sup> Condition (ii) in the definition might be weakened, so that only certain *experts* need to be in a position to recognize the stipulation described in (i), and other speakers’ competence can be parasitic on the experts’ competence. I will ignore such complications.

descriptive *stereotype* associated with the word, and on the other hand, the claim that the word has a descriptive *content* or descriptive *mode of presentation*.

Putnam's notion of a *stereotype* is meant to capture the descriptive information that competent users of a linguistic expression are required to associate with that expression. For example, the stereotype associated with "lemon" might include the properties of having a yellow peel, a tart taste, and so on.<sup>11</sup> For many terms, like "water," we do seem to have robust stereotypes. But for many other terms we do not.<sup>12</sup> What is important for present purposes is that the information contained in a stereotype need not be uniquely identifying (there may be a number of fruits with yellow peels and a tart taste) and need not be accurate (lemons would still be lemons, even if it turned out that their peels only falsely appeared to be yellow). So stereotypes should not be conflated with reference-fixing descriptions. A reference-fixing description *does* have to be uniquely satisfied by an object, in order for that object to count as the term's referent.<sup>13</sup>

Reference-fixing has to do with how a term *gets assigned* the meaning it has, and with what's required to *be competent* in the use of the term. There is no straightforward route from such considerations to conclusions about what the term's *content* is. For example, it is plausible that competence with "you" requires knowing that it refers, relative to a context, to the person who is being addressed in that context. Yet on many views, the content of "you" has nothing to do with being addressed. What I believe when I say "You are tall" is something that Joe can also believe so long as he believes, of the person I am addressing, that he is tall. Joe does not in addition have to believe that this person is addressed by me. Joe does not even have to have the concept of *being addressed*, or to have mastered the use of second-person pronouns. Facts about who is addressed play a role in determining the reference of "you" relative to a given context, but the property of being addressed is not part of *the content* of "you." This illustrates the

---

<sup>11</sup> See Putnam 1975, pp. 247-52, 269; also "Is Semantics Possible?" ??

<sup>12</sup> E.g., consider Putnam's "elm/beechn" example. Putnam is a competent user of these terms, yet associates only very rough descriptive information with each.

<sup>13</sup> There is the following relation between stereotypes and reference-fixing descriptions: When a term's reference *is* fixed by a description, competent speakers are required to know that; and so it could be argued that the reference-fixing properties thereby get to be *part of* the term's stereotype. But there is no route from being part of the stereotype to playing a reference-fixing role.

way in which something can be part of what we need to know to be competent with a word, and yet fail to make any contribution to that word's content.<sup>14</sup> We will be examining the relation between claims about competence and claims about content later.

Our next step is to introduce the notion of a **reference-fixing transform** of an expression. That's what you get when you replace all the Twin-Earthable terms in that expression, like "water," with the descriptions which fix their reference. (For the time being we assume that there *are* such reference-fixing descriptions. More on this later.) For example, suppose you fix the reference of "water" to be the clear drinkable liquid humans have predominantly interacted with. Then the reference-fixing transform of "water" will be "the clear drinkable liquid humans have predominantly interacted with," and the reference-fixing transform of the sentence "Water contains hydrogen" will be "The clear drinkable liquid humans have predominantly interacted with contains hydrogen."

Kripke's Strategy says that if it falsely seems to us that water could fail to contain hydrogen, that's because we have a genuine modal intuition that the clear drinkable liquid humans have predominantly interacted with could fail to contain hydrogen. That genuinely *is* possible, because it is possible for humans to have predominantly interacted with a clear drinkable liquid other than water. In other words, we falsely seem to intuit the possibility of what's expressed by "Water contains no hydrogen"; but what we're really intuiting is the possibility expressed by the reference-fixing transform of that sentence, namely, the possibility expressed by "The clear drinkable liquid humans have predominantly interacted with contains no hydrogen." So it's the *genuine contingency* expressed by the reference-fixing transform which explains the *seeming* contingency of what's expressed by the original sentence.

---

<sup>14</sup> On the flip side, the fact that (i) certain descriptive properties are part of a name's content will not entail that (ii) they fix the name's referent. Nor will the fact that (iii) those descriptive properties are part of a mode of presentation you associate with the name. To fix the name's referent, the descriptive properties must at least be *true* of the referent and *uniquely identifying*. Neither of these conditions are obviously guaranteed by (i) or (iii). For instance, you might hold that the descriptive property of being human (and Roman?) is part of the content of "Cicero." Or, even if it was Schmidt who proved the incompleteness of arithmetic, you might still hold that the descriptive property of proving that result is part of the mode of presentation you associate with "Gödel." In neither case do the relevant descriptive properties fix the name's referent.

So far, so good. However, we have to complicate this story in two respects. First, it will often happen that a reference-fixing description contains terms which are themselves Twin-Earthable. If so, the reference-fixing transforms we get by simply substituting in that reference-fixing description won't be able to account for all our illusions of contingency. An example will help make this clear.

Say that a liquid is *continuous* iff it's liquid "all the way down." That is, iff any part of the liquid has proper parts which are themselves liquid. Say that a liquid is *particulate* iff it's not continuous. Now, it is an *a posteriori* truth about actually-occurring liquids that they're all particulate. They all have proper parts which are themselves too small to be in a liquid state. It could be argued that this is a necessary truth about liquidity: given that all liquids are in fact particulate, it is necessary for liquids to be particulate. Let's suppose for the purposes of this example that this is necessary. So it's necessary and *a posteriori* that liquids are particulate.

Now, we can *imagine* it having turned out that water is continuous. But in fact that is not really possible. Water is essentially made up of hydrogen particles and oxygen particles, which aren't themselves liquids. So our intuition that water could have turned out to be continuous is an illusion. How are we to explain this illusion? It won't do to say that what we're *really* intuiting is the possibility expressed by "The clear drinkable liquid humans have predominantly interacted with is continuous." For we just said that it's not possible for *any* liquid to be continuous. So there are no possible situations in which humans interact with a continuous liquid. Hence, our initial reference-fixing transform can't do the job we want it to.

The fix is to iterate the process by which we construct the reference-fixing transform. First, we replace "water" with "the clear drinkable liquid which humans have predominantly interacted with." Then we replace "liquid" with whatever description we used to fix *its* reference. (Assuming there is one; more on this later.) And we keep on going, until we find a sentence which does express a genuine possibility. It will be the genuine possibility expressed by that last sentence which explains our illusion of intuiting that water could have turned out to be continuous.

From here on, we'll assume that reference-fixing transforms are the result of carrying out this process on every Twin-Earthable term in an expression, until no Twin-

Earthable term remains, other than explicit indexicals like “here,” “me,” and “that.” We will discuss these in a moment.<sup>15</sup> We might come upon Twin-Earthable terms which *have no* reference-fixing description; so there is no guarantee that a reference-fixing transform can always be constructed for any given sentence. But for the time being, let’s assume that we’re always able to come up with the reference-fixing descriptions we need. According to the Strengthened Kripke Strategy, when a sentence S falsely seems to us to express something contingent, that’s because we do have a genuine modal intuition of the possibility expressed by S’s reference-fixing transform.

Our second complication concerns indexicals. So far, we’ve been given no guidance about how to handle reference-fixing transforms when they contain indexicals. Let’s take an example. Suppose I fix the reference of “Chico” by saying:

I hereby stipulate that “Chico” is to refer wrt every possible world to the (actual) truth value of the claim that *here* = Chicago.

Suppose that I am in Chicago as I do this. Hence, “Chico” rigidly designates the truth-value Truth. So it is necessary that Chico is Truth. Of course, it is also *a posteriori* that Chico is Truth. *A priori* reasoning alone won’t tell me what city I’m in as I perform my stipulative act. For that, I need empirical evidence. Hence it is necessary and *a posteriori* that Chico is Truth. Because it is *a posteriori*, we can imagine it having turned out that Chico is not Truth. So it will seem to us possible that Chico might not have been Truth. But this is an illusion. Chico is necessarily Truth. How is the illusion to be explained?

Our procedure so far tells us to take the sentence “Chico is not Truth” and consider its reference-fixing transform. That will be “The truth-value of the claim that *here* = Chicago is not Truth.” Now our illusion of contingency is supposed to be explained by this latter sentence’s expressing something which genuinely *is* possible. But this latter sentence contains an indexical, “here.” So considered out of context, it doesn’t express anything at all. The obvious thing to try is to consider it relative to my context, as I stand in Chicago, having just introduced the term “Chico.” But that won’t give us the right result. Considered relative to the context just described, “The truth-value of the

---

<sup>15</sup> Sometimes it’s claimed that Twin-Earthable terms like “water” are *themselves* indexical terms. We will investigate this claim in Chapter 3 and find that it rests on a confusion. For the time being, just assume that we form a reference-fixing transform by repeatedly replacing terms like “water” with their reference-fixing descriptions, and leaving explicit indexicals like “here,” “me” and “that” alone.

claim that *here* = Chicago is not Truth” expresses the proposition that the truth-value of the claim that *Chicago* = Chicago is not Truth. That proposition is necessarily false. So it can’t be *its* possibility which explains the seeming possibility of Chico’s not being Truth.

Our mistake was in thinking that it’s *the proposition expressed by* the reference-fixing transform which accounts for seeming possibility of Chico’s not being Truth. Rather, it’s the *character* of the reference-fixing transform which is important. Relative to some possible contexts, the reference-fixing transform expresses a proposition which is, in those contexts, true. Relative to other possible contexts (in other places and on other possible worlds), the reference-fixing transform expresses a proposition which is, in those contexts, false. It’s because of this that the original claim is *a posteriori*. We seem to intuit the possibility of Chico’s not being Truth, but what we’re really intuiting is the possibility of someone being in a context, perhaps different from ours, where some other truth value is picked out by the reference-fixing description “the truth-value of the claim that *here* = Chicago.”

Of course, it is always possible for *a description*, a piece of language, to pick out different objects; for it’s always possible for speakers to use that piece of language differently, and to assign it a different meaning. What concerns us is whether the *character* that the description actually has, as we understand it, picks out different objects with respect to different contexts. So instead of continuing to talk of reference-fixing transforms, which are *sentences*, let’s talk about the characters we understand those reference-fixing transforms to have. Let me introduce a new notion to highlight exactly what aspect of their character we will be concerned with. I’ll call this new notion a **reference-fixing profile** (or **RFP**). Here is how you get a RFP. You start with an expression T. You form its reference-fixing transform T\*. You consider the character C that T\* actually has, as you understand T\*. C gives us a function from contexts to intensions, which are in turn functions from worlds to extensions. Now take the function which maps every context *c* to the extension of C(*c*) in the world of that context *c*. That function will be the RFP of T. So for instance, the RFP of “my spouse” is the function which maps every context *c* to the person who is, in the world of that context, the spouse of the agent of that context. The RFP of “Chico” is a function which maps every context *c* to the truth-value Truth if *c* is located in Chicago, and to False otherwise. And so on.

Let's consider another example. Suppose I fix the reference of "Hydro" by saying:

I hereby stipulate that "Hydro" is to refer wrt every possible world to the heaviest elementary constituent of *this* [pointing to a liquid].

Suppose I am demonstrating a sample of water when I perform this act. So "Hydro" will rigidly designate the element oxygen. But of course it's *a posteriori* that Hydro is oxygen. Because that is *a posteriori*, we can imagine it having turned out that Hydro is not oxygen. So it will seem to us possible for Hydro to have turned out not to have been oxygen. But this is an illusion. Hydro is necessarily oxygen. How is the illusion to be explained?

Our procedure tells us to take the reference-fixing transform of "Hydro is not oxygen." That is "The heaviest elementary constituent of *this* is not oxygen." Relative to some contexts, like the actual context, where what's being demonstrated is a sample of water, that sentence expresses a proposition which is (in the world of that context) false. But relative to other contexts, in which a sample of some other liquid is being demonstrated, the sentence will express a proposition which is (in the world of those contexts) true. So our RFP for "Hydro is not oxygen" will yield the value true relative to some contexts and false relative to other contexts. In a situation of this sort, let's say that the RFP is **fickle**. When the RFP yields *the same* extension relative to every context (for which it yields an extension), on the other hand, say that the RFP is **faithful**. The RFP for "Hydro is not oxygen" is fickle. According to the Strengthened Kripke Strategy we're elaborating, that explains why it *seems* possible to us for Hydro to have turned out to have been something other than oxygen. The fickleness of the RFP indicates there is a genuine possibility we are intuiting: the possibility of one's having, at the time of fixing "Hydro"'s reference, been demonstrating a sample whose heaviest elementary constituent is not oxygen.

That is the intuitive way to extend Kripke's Strategy to cases where the reference-fixing description includes indexicals. We'll see that this intuitive fix perfectly suits the formal developments we'll be looking at in Chapter 4.

This Strengthened Kripke Strategy for explaining illusions of contingency is more systematic than anything we find in Kripke. In addition, it helps itself to the assumption

that reference-fixing descriptions will always be readily available—an assumption that *Kripke* would be loath to accept.<sup>16</sup> So although this Strengthened Kripke Strategy is *suggested by* the Strategy Kripke actually employs, we shouldn't expect him to sign on.

Myself, I think this Strengthened Kripke Strategy will be able to explain only a very few illusions of contingency. However, let's imagine the Strengthened Kripke Strategy being advocated by someone who thinks it will be much more comprehensive than that. Suppose the advocate thinks the Strengthened Kripke Strategy is the *only way* to explain illusions of contingency. In fact, let's make it an assumption of the Strengthened Kripke Strategy that this is so. So according to the Strengthened Kripke Strategy, if there is an illusion of contingency, it must *always* be explicable in the way we've described. Since we can suppose that necessary *a posteriori* truths can always give rise to illusions of contingency, it follows that:

(SKS-1) A sentence S expresses a necessary *a posteriori* truth only if S's RFP is fickle.

Now, having gone this far, there's no obvious reason to limit ourselves to necessary truths. For *every a posteriori* truth, whether it's necessary or not, our ability to conceive it being otherwise will give rise to an intuition of contingency. So far, we've only been discussing cases where the truth in question is necessary, and so the intuition of contingency is illusory. But there's no reason why the Strengthened Kripke Strategy shouldn't be viewed as an account of *all* these intuitions of contingency, regardless of whether they are illusory. Hence, we can take the following to be the central claim of the Strengthened Kripke Strategy:

(SKS-2) A sentence S expresses an *a posteriori* truth (that is, a truth which is knowable only *a posteriori*) only if S's RFP is fickle.

It's tempting to conjoin that with the inverse:

(SKS-3) A sentence S expresses an *a priori* truth (that is, a truth which is knowable *a priori*) only if S's RFP is faithfully true.<sup>17</sup>

<sup>16</sup> Consider the necessary truth that so-and-so is Gödel's father (if Gödel exists). This truth generates as much an appearance of contingency as the truth that water is made of H<sub>2</sub>O. But for names like "Gödel," Kripke has argued we have no reference-fixing descriptions. So when I seem to intuit possible worlds in which someone else is Gödel's father, what I'm intuiting *can't be* someone else having the properties that fix the reference of the name "Gödel" for me. There are no such properties.

<sup>17</sup> However, I claimed in fn. 1, above, that even if p is knowable *a priori*, not-p can still be conceivable. (Perhaps not-p could not be conceivable when p is *known a priori*. It may be that *knowing*

In what follows, however, we will be concerned primarily with SKS-2 and its consequences.

It is an immediate consequence of the Strengthened Kripke Strategy that if we fix the reference of a term like “water” with the reference-fixing description “the clear drinkable liquid around here,” the sentence “Water is the clear drinkable liquid around here” will express something which is knowable *a priori*. For the reference-fixing transform of that sentence is “The clear drinkable liquid around here is the clear drinkable liquid around here,” which expresses a truth relative to every context (setting aside for the moment contexts where there is no clear drinkable liquid nearby). Hence the RFP of “Water is the clear drinkable liquid around here” will be faithfully true. It will yield a truth for every context. According to SKS-2, that means that it can not express an *a posteriori* truth. So if what it expresses is knowable at all (as it surely is), it has to be knowable *a priori*.

Many will view this result as unexciting. After all, they’ll say, if we fix the reference of “water” in the way described, isn’t it obvious that it *is* knowable *a priori* that water is the clear drinkable liquid around here?

It is not obvious. In fact, the task of Part II of this essay will be to persuade you that it’s false. Fixing the reference of “water” to be the clear drinkable liquid around here does *not* make it knowable *a priori* that water is the clear drinkable liquid around here. That’s something we can know only *a posteriori*. In general, I will argue, fixing the reference of “N” with the description “the F” will not enable us to know *a priori* that N is the F.

So I will argue that “Water is the clear drinkable liquid around here” expresses something we can know only *a posteriori*. Yet, the RFP of “Water is the clear drinkable

---

something *a priori* requires being in an evidential situation from which no defeating evidence could dislodge one. This is controversial. All I’m claiming here is that being in an evidential situation from which not-p is conceivable is compatible with p’s being knowable *a priori*.) Doesn’t that suggest that *a priori* truths can give rise to illusions of contingency in much the same way as necessary *a posteriori* truths? If so, and we wanted to explain those illusions of contingency with the Strengthened Kripke Strategy, we’d need the relevant sentences to have fickle RFPs. That would be a reason for avoiding SKS-3.

[If the 2D theorists *did* say that knowable *a priori*  $\square$  faithful RFP, this would be akin to the positivist’s claim that *a priori* truths are those truths which are expressed by a special kind of sentence, a sentence such that just understanding it is enough to know that it’s true.]

liquid around here” certainly is faithful (again, setting aside for the moment contexts where there is no clear drinkable liquid nearby). So we have a sentence that expresses an *a posteriori* truth, but its RFP is not fickle. SKS-2 must be mistaken, then. It cannot explain, and requires us to deny, the *a posteriority* of the fact that water is the clear drinkable liquid around here.

Similar examples demonstrate the Strengthened Kripke Strategy’s inability to explain the *a posteriority* of many *necessary* facts. For example, fix the reference of “Trudy” to be the firstborn child of Harry Truman. Arguably, being a child of Harry Truman will be essential to Trudy (perhaps being the *firstborn* child won’t be). Now, just as it’s not knowable *a priori* that water is the clear drinkable liquid around here, neither do I believe it’s knowable *a priori* that Trudy is a child of Harry Truman. So it will be a necessary *a posteriori* truth that Trudy is a child of Harry Truman. But the associated reference-fixing transform is “The firstborn child of Harry Truman is a child of Harry Truman.” This will express a truth relative to every context (setting aside contexts where Harry Truman has no children). So the RFP of “Trudy is a child of Harry Truman” is faithfully true. The sentence expresses a necessary *a posteriori* truth, but its RFP is not fickle. Both SKS-1 and SKS-2 are mistaken, in this case.

In general, the Strengthened Kripke Strategy can succeed only if we can know objects *a priori* to have those properties we use to fix the reference of their names. This is a fundamental assumption of the Strengthened Kripke Strategy. Let’s call it SKS-4:

(SKS-4) If you fix the reference of “N” with the description “the F,” then you will be able to know *a priori* that N is the F.

Judging from contemporary discussions, many philosophers seem prepared to allow the assumption SKS-4. However, I will argue that they are wrong to do so. The assumption is mistaken and rests on a number of confusions about the relations between Twin-Earthable thoughts and epistemology. It must go, and so too must go the Strengthened Kripke Strategy which is built upon it. Perhaps *some* illusions of contingency can be explained in the way we’ve sketched in this chapter. But the Strengthened Kripke Strategy can’t have anything like the comprehensiveness it reaches for, and which is codified in SKS-1 and SKS-2. We’ll need to find a different account of what happens in most of the cases where necessary truths give us the illusion of contingency.

Jackson and Chalmers are strong advocates of the Strengthened Kripke Strategy. In Chapter 4, I will set out will set out and explain their philosophical framework. We will see how it builds upon the Strengthened Kripke Strategy, and how it too takes on board the assumption SKS-4.

Before we do that, though, we should pause and look more closely at the relation between the character of Twin-Earthable terms like “water” and the characters of their reference-fixing transforms.