

*The Fourteenth Annual Harvard-MIT
Graduate Student Philosophy Conference*

Cambridge, Massachusetts
March 4th and 5th, 2006

Peter Godfrey-Smith

KEYNOTE SPEAKER

Contents

1. <i>How to Perceive the Past with your Eyes Shut</i> JOHN BENGSON (TEXAS)	1
2. <i>Tought-Experiment Intuitions and Truth in Fiction</i> JONATHAN ICHIKAWA (BROWN)	15
3. <i>Making Sense of Perceptual Defeasibility</i> CIARA FAIRLEY (UCL)	25
4. <i>Getting Back to Nature</i> LEONARD KAHN (OXFORD)	37
5. <i>Imagery, Evidence, and the Mind-Body Problem</i> PETER LANGLAND-HASSAN (CUNY)	49

*How to Perceive the Past with your Eyes Shut**

John Bengson[†]

UNIVERSITY OF TEXAS, AUSTIN

I can read with my left eye. I can read with my right.
I can read Mississippi with my eyes shut tight!
Mississippi, Indianapolis and Hallelujah, too!
I can read them with my eyes shut!
That is VERY HARD to do!

Dr. Seuss, *I Can Read with My Eyes Shut!*

1 Introduction

THERE IS A TENDENCY to assume that memory and perception are entirely distinct: that is, if x is a memory, then x is not a perception, and if x is a perception, then x is not a memory. My aim in what follows is to pose a (somewhat unsettling) challenge to this tendency.

The distinction between memory and perception is typically drawn in either one of two ways, each of which involves an appeal to a seemingly innocuous assumption about perception. The first is that we perceive only the present.¹ Most likely it is with this seeming platitude in mind that Alva Noë (2003; p. 96) maintains that perception is only of that which is “in the here and now” and David Lewis (1980; p. 79) suggests that one sees only when “before [one’s] eyes various things are present and various things are going on.”² The second is that in order to perceive, one’s sensory organs must be

*Earlier versions of this paper were presented at the University of Wyoming in fall 2004 and conferences in spring 2005 at the City University of New York and the University of Colorado at Boulder. I would like to thank my commentators and audiences at each of these events. Special thanks are owed to Franz-Peter Griesmaier, Marc Moffett, Jennifer Wright, Kendall Walton, Michael Tooley, and Michael Levin for their particularly helpful comments on previous drafts of this work.

[†]jsteele@mail.utexas.edu

¹ Perhaps the most explicit such appeal is Sutton’s (2004) assertion, “We remember experiences and events which are not happening now, so memory seems to differ from perception.”

² Although it may seem plausible to admit of *auditory* exceptions to this general principle, such as hearing thunder from a distant storm, most philosophers (including Noë and Lewis) tend to assume that, at the very least, all *seeing* (often considered the paradigm example of perception) is of the present. Paying heed to this tendency, I will focus my discussion of this principle on the specific case of visual perception. But I see no reason to doubt that the arguments given would apply *mutatis mutandis* to other perceptual modalities as well.

operating at the time of perceiving. In other words, one must *use* one's sensory organs to perceive. This second assumption is quite widespread among both philosophers and cognitive scientists; for instance, Gilbert Ryle (1949; p. 246) claims that "A person can see things, only when his eyes are open, and when his surroundings are illuminated" and Leo Hurvich (1981; p. 26) asserts that "we see objects and colors only when our eyes are open and light enters them."

Apparently, each of these authors considers his statement to be more or less obvious, not a thesis to be stated and argued for. Indeed, not one of them provides an argument directly supporting either of these widely shared assumptions. Presumably, this lack of support goes unnoticed because a denial of these assumptions seems highly counterintuitive: in short, it sounds quite bizarre to say that we can perceive something not currently present, or that we can do so without using our sensory organs.

Despite this orthodoxy, there are good reasons to think that neither of these claims is as bizarre as it may at first sound. I begin, in §2, by arguing that there are cases of perception, such as gazing at the stars, in which we perceive the past. In §3, I introduce a thought experiment that develops Kendall Walton's thesis that to look at a photograph is to genuinely *see* whatever it is a photograph of, namely, a past scene. Then, in §§4-6, I consider reasons to think that we may also perceive past scenes in virtue of *remembering*. If correct, this indicates that, despite widespread assumptions to the contrary, one cannot draw a principled distinction between perception and memory.

2 Interstellar Perception

Let us begin with an argument in favor of perception of the past, one which takes seriously ordinary star-gazing.³ Consider the so-called "North Star", Polaris. Polaris is 650 light-years away, so its light takes roughly 650 years to reach us. So, when we look at Polaris now (in the twenty-first century), we are seeing it as it was in the fourteenth century. And if Polaris had suddenly exploded in 2003, we would not know it until the year 2653. This is because we cannot see Polaris as it is presently; we can only see it as it was, in the past. As a result, this is an instance of perception of the past.⁴

Note that this conclusion is in line with the causal theory of perception. According to this widely held analysis, to perceive something is to have experiences caused by that which is perceived.⁵ In order to rule out deviant causal chains, it has been suggested that these experiences must be counterfactually dependent on that which is perceived.⁶

³Robert Solso has suggested another argument for the claim that we can perceive the past. He writes, "One implication of the finite speed of light is that we always see the past" (1994; p. 8). The idea seems to be this: perception involves a causal transaction; causal transactions take time; so, it follows that there is always a sort of delay in perception. Although this may be right, it is unclear to me whether Solso's desired conclusion—namely, that we always perceive the past—is correct. Though I lack the space to fully address this issue here, it would seem that *if* Solso is correct, this only strengthens the argument given below. However, nothing in what follows presupposes this.

⁴Although the *light* of Polaris is present to the eye, we perceive *Polaris*, and not proximal arrays of light.

⁵The causal theory of perception has been defended by Grice (1961), Strawson (1974), Goldman (1976), and Jackson (1977), to mention only a few. For an argument that despite its apparent flaws, we ought to accept some version or other of the causal theory, see Coates' (2000) response to the non-causalist.

⁶See Lewis (1980). I realize that treating the counterfactual requirement as an amendment to the causal theory of perception may not solve the problem of deviant causal chains, as Davies (1983) and others have

Hence, an agent A perceives a perceptual object o if

- (α) A has an experience as of o , and
- (β) A 's experience is appropriately causally and counterfactually dependent on o .

This makes clear that gazing at Polaris is genuine perception of the past. For in such a case,

- (α_1) A has an experience as of Polaris, and
- (β_1) A 's experience is appropriately causally and counterfactually dependent on Polaris.

Clearly the match that occurs during such star-gazing is not a “lucky accident” but an instance of genuine perception. In such a case, we can legitimately be said to be in perceptual contact with what our present experience is of—namely, something in the past.⁷

3 Snapshot Perception

Let us engage in a bit of science fiction in order to further develop this point. Borrowing from (Grice 1962), let us suppose that Martians finally arrive on earth. Suppose further that they are physiologically and psychologically similar to normal human beings. The only significant difference is that Martians possess a more sophisticated perceptual system: they can “freeze” a given visual image (perhaps by arresting the patterns of light currently on the retina), allowing them to enjoy a given perception while the world around them continues to change.⁸

When they freeze a given image, do the Martians continue to perceive what it is an image of? It would seem so. After all, they are enjoying an experience that, all things being equal, is appropriately causally and counterfactually dependent on what it

argued. However, it would appear that no satisfactory solution has yet been given to the problem of deviant causal chains; Grice (1961), Goldman (1976), Jackson (1977), Peacocke (1979), Searle (1983), Davies (1983), Pendlebury (1994), Coates (2000), among others, have proposed solutions, though it remains unclear whether any of them is adequate. This is not the place to evaluate these proposals, so I will assume that the “right” sort of causal connection is, in addition to being counterfactually supporting, one that is *appropriate*. So, to perceive something is to have experiences that are appropriately causally and counterfactually dependent on what they are of, as stated in the text. Whether fleshing out the appropriateness condition has implications for the central argument of this paper is addressed in §5.

⁷It is worth mentioning that Lewis has challenged the idea that the stars we perceive in the night sky are in the past. He argues that star-gazing consists in perceiving something “not straight-forwardly past; for lightlike connection has as good a claim as simultaneity-in-my-rest-frame to be the legitimate heir to our defunct concept of absolute simultaneity” (1980; p. 83, n.9). Lewis appears to believe that if ϕ is seen now, then because it is within an observer’s light cone, ϕ is simultaneous with what is happening now— ϕ is “not straightforwardly past”, but present. If this is so, then an explosion of Polaris that occurred almost seven centuries ago is present so long as we see it now. But, presumably, such an explosion does not occur in the present. Indeed, it does an injustice to our concept of the present to apply it to that which occurred almost seven centuries ago: *pace* Lewis, we clearly consider such an event to be past.

⁸Of course, the frozen image is not an afterimage, which merely “overlays” a visual scene.

is of. So, if the Martians did in fact perceive the scene initially, because their frozen perceptions continue to satisfy the causal and counterfactual conditions that mark out genuine perception, it would seem to follow that they still perceive the scene seconds or minutes later when it is frozen in their perceptual field. Thus, such “snapshot perception” enables the Martians to perceive (indirectly but genuinely) a given scene long after it has occurred.⁹

One might object that the causal theory of perception can be modified to undercut the possibility of snapshot perception. But such a move appears to have undesirable consequences insofar as snapshot perception might not be wholly science fictional. Imagine creatures in possession of a perceptual system such that they perceive in virtue of rapidly processed frozen images of their surroundings. Such creatures would possess a peculiar sort of snapshot perception. Interestingly, these creatures may be *Homo sapiens*. Robert Solso (1994; p. 26) writes,

When we [human beings] look at an object [...] we do not see it all at once [...] but go through series of scans in which the eye momentarily stops on one feature [...] then darts on to another part [...] and then on to another [...] and so on. [...] Since this scanning/stop maneuver takes place over very short time periods, the subjective experience is that we are seeing [an object] all at once, when, in fact, our visual perception of it is built up from a series of discrete “snapshots.”

In short, homogenous visual experiences result from the rapid processing and subsequent binding of a series of initially discrete images. The fact that “discrete ‘snapshots’” play an important role in human vision suggests that we, like the Martians, enjoy a form of snapshot perception. Of course, ours are not produced voluntarily and occur too quickly to be noticed. But suppose for a moment that the discrete snapshots underlying human vision could be frozen, or processed at a speed at which we could, with practice, come to notice each snapshot individually. Since we would not deny that we perceive even if it came to light that our visual perceptions somehow relied on this type of snapshot perception, it looks as though we ought to consider snapshot perception a form of genuine perception.

That said, let us push our thought experiment one step further. Suppose that the Martians can pull up previously frozen images at will. In so doing, they put themselves in the same perceptual position as when they initially froze the images in their perceptual field. Accordingly, if they perceived what the images are of when they initially froze them, then it would appear to follow that they perceive what they are of when they pull them up at a later time.

This result enjoys support from reflection on similar cases in the actual world. For a start, Kendall Walton (1984, 1986, 1997) has convincingly argued that viewing photographs involves such snapshot perception. In most cases, photographs are appropriately causally and counterfactually dependent on what they are of.¹⁰ Walton observes

⁹Now, the Martians may not be perceiving *simpliciter*, if that is taken to mean direct (unmediated), as opposed to indirect (mediated), perception. But whether or not their snapshot perception is direct perception is beside the point, since indirect perception is still perception (see below).

¹⁰Walton (1986) recognizes that over—and under—developed or artistically altered photographs are a few

that unlike the content of a painting or sketch, which depends on what the artist believes that he or she sees, the content of a photograph “is determined by what is really there before [the photographer], regardless of what he [or she] thinks” (Walton 1984; p. 264). Because they have such “natural dependence” (i.e., counterfactual dependence not mediated by an agent’s intentional states) and thus preserve real similarity relations between objects, photographs, Walton argues, are *transparent*.

To see the force of this transparency thesis, consider relevant similarities between photography and other devices which clearly do not preclude but rather enable (indirect) perception. For instance, we would not deny that someone with eyeglasses perceives what her visual experiences are of. Even though the lenses act as mediating transducers, if appropriate causal and counterfactual dependence is maintained, then she still sees (albeit indirectly). Because mirrors also maintain such dependence, we should not deny that when looking at oneself in a mirror, one is in fact looking at oneself.¹¹ A periscope produces a similar result: so long as appropriate causal and counterfactual dependence is maintained, its mirrors act as mediating transducers which allow one to see (again, albeit indirectly). Since photographs also maintain appropriate causal and counterfactual dependence, they, like eyeglasses, mirrors, and periscopes, appear to act as mediating transducers that allow one to see (once again, albeit indirectly). As with other artificial or prosthetic devices that, because they maintain natural dependence while preserving real similarity relations between objects, enable one to see (indirectly but genuinely) things that one could not see otherwise, photographs, though imperfect, make a “contribution to the enterprise of seeing” (*ibid*, 251).

Interestingly, that looking through photographs is on a par with clear cases of perception is consistent with how we ordinarily speak about viewing photographs. Just as we are likely to assert that we see ourselves when looking in the mirror, although we do so indirectly, we are likely to assert that we see a photographed object, albeit indirectly. When viewing a photograph of the Taj Mahal, for instance, it is only natural to (non-metaphorically) say things like “I see the Taj Mahal.”¹² The same is true of film, television, and other so-called “moving pictures.” Consider watching a game on the television: we frequently and unhesitatingly say that we see the game, the players, the fans, the ball, the last-second-shot, and so on. In fact, we often think we see such things *better* via the television than can those (e.g., the referees) who are actually there, seeing them directly!

Obviously, photographs are of objects or events which occurred at some time in

instances in which this claim (to varying degrees) might not hold. Film and television are instances in which it could.

¹¹This is not to say that one never sees one’s reflection. One sees one’s reflection and *thereby* sees oneself. I should mention that, as far as I know, all parties in this debate in philosophical aesthetics accede that mirrors are transparent. See Walton (1997), Currie (1995), Carroll (1995), Carroll (1996), Cohen and Meskin (2004).

¹²Of course, the contextual nature of seeing-talk complicates things. Consider the question “Did you see the Taj Mahal?” In certain contexts, one can legitimately answer by saying, “Yes, I saw the Taj Mahal”, if, for instance, one was just shown a photograph of the Taj Mahal. But in other contexts, one can legitimately answer by saying, “No, I have never seen the Taj Mahal”, if one has never seen the Taj Mahal in person (regardless of whether one has or has not seen photographs of it). Incidentally, one with a propensity for linguistic analysis might say that on the present view, statements made while looking through photographs, such as “A sees *o*”, are to be analyzed as “A sees *o* in virtue of seeing the photograph.” An adverbial analysis is also available: “A sees *o* photographically.” Either way, A sees *o*. The same considerations apply *mutandis* to other types of perceiving-talk.

the past. So, if this transparency thesis is correct, photographs are a way in which we perceive the past. This makes plausible the somewhat surprising notion that our visual system is, in a way, just as sophisticated as the Martians. Human vision can be *extended* through the technology afforded by cameras, and so looking at a photograph is roughly equivalent to pulling up previously frozen perceptions.¹³ Viewing a photograph is engaging in Martian-like snapshot perception.

4 The Transparency of Episodic Memory

There is reason to think that photography is not the only real-life example of Martian-like snapshot perception. For memory can be very much like photography. Of course, memories are not photographs. But insofar as some memories function like photographs by capturing past scenes, the comparison is apt.

Consider first how we sometimes talk about our remembrances. For instance, we often say things like “I can still see the look on his face.” Such statements reflect a peculiar fact about memory: we often remember past scenes in such a way as to make them “present” to us once again. This is most evident in eidetic memory, as Ian Hunter has noted:

One of the pictures used by Allport [in his 1924 study of memory in children] depicted a street scene and contained, among other details, the German word *Gartenwirtschaft* written above the door of an inn in the background. The word was quite meaningless for the English children and was not usually reported at first [. . .]. But on being pressed to observe [sic] more closely, each of the thirty children whose eidetic imaging was strong saw [sic], often to his surprise, the small letters above the door. Three of these children spelled out the word without error, seven got no more than two of the letters wrong, and only five failed to give at least five letters correctly [. . .]. There was, of course, no question of the word having been memorized (quoted in [Zemach 1969](#); p.15).

Is it not appropriate to say, as Hunter does, that these children are ‘seeing’, or ‘observing’, the words in the picture?

Putting this question to the side for a moment, consider two noteworthy features of episodic memories, the class of memories of which eidetic memories are members.¹⁴ Episodic memories concern neither propositions nor embodied skills, but, rather, one’s experiences of past events and episodes.¹⁵ Tyler [Burge \(2003\)](#) has referred to episodic memory as a type of “experiential memory”, for through it we “re-live” experiences that

¹³ In light of the facts about photography cited in this section, this way of putting the point appears to be licensed by the view which [Clark and Chalmers \(1998\)](#) call *active externalism*.

¹⁴ Incidentally, there may be a third such feature, in addition to the two mentioned in the text. Suppose that, as [Noë \(2002a, 2002b, 2003, 2004\)](#) and [Jonathan Cohen and Aaron Meskin \(2004\)](#) have argued, perception is always egocentrically indexed. As [Burge \(2003\)](#) has pointed out, a given episodic memory has *de se* form, and is therefore egocentrically indexed, in three ways: “It is indexed to my having experienced the act or event, to my having been the agent or subject of it, and to my perspective as agent in the past act or event” (294).

¹⁵ Episodic memory, which involves a phenomenal experience of a particular past episode or event, is

we have previously had (see also [Martin and Deutscher 1966](#), [Brewer 1996](#)). This is the first interesting feature of episodic memories, namely, that they involve being in a state that is qualitatively identical or very similar to a specific past experience. For instance, when remembering one's childhood, one may have an experience as of the house in which one grew up or the yard where one often played. In such cases, past scenes become present to us once again.

The second noteworthy feature of episodic memories concerns their veridicality conditions. As C. B. Martin and Max Deutscher (1966), David Wiggins (1976), and others have argued, in order to count as a genuine memory, there must be an appropriate causal connection between a past scene and one's present experience. Presumably, this causal connection must be counterfactually supporting. This explains why it is that some episodic memories can be compared with photographs. Like photographs, genuine episodic memories involve experiences that are appropriately causally and counterfactually dependent on what they are of. In effect, like photographs, genuine episodic memories maintain natural dependence and preserve real similarity relations between objects.

Given this, some episodic memories appear to have much in common with eyeglasses, mirrors, and periscopes. It would seem that like these prosthetic devices, genuine episodic memories make a contribution to the enterprise of perceiving. In short, there seems to be reason to believe that genuine episodic memories, like photographs, are *transparent*.

Because this conclusion may strike one as surprising, I want to be clear about the sense in which it is reasonable to hold that some episodic memories are continuous with so-called "ordinary" perception. For one, there is the possibility of snapshot perception. Second, there is the way that we sometimes talk about our remembrances. In addition, as we have just seen, episodic memories are in ontologically relevant ways on par with "ordinary" perceptions. That is, in some episodic memories,

- (α_2) A has an experience (*qua* memory) as of *o*, and
- (β_2) A's experience (*qua* memory) is appropriately causally and counterfactually dependent on *o*.

That genuine episodic memories satisfy the causal theory of perception is at least *prima facie* reason to think that, in such cases, we can legitimately be said to be in perceptual contact with what our present memories are of.¹⁶

sometimes called 'direct', 'personal', or 'recollective' memory. It is typically distinguished from memory concerning facts (i.e., remembering *that*), which is often called 'semantic' or 'propositional' memory, and memory concerning embodied skills (i.e., remembering *how*), which is often called 'procedural' or 'habit' memory.

¹⁶In the next section, I express skepticism towards the ability of popular modifications of the causal theory to undermine this conclusion, and in §7 suggest that simply rejecting it on the grounds that it is plainly unacceptable invites unattractive consequences. Interestingly, if this conclusion is correct, it would explain Martin and Deutscher's (1966: p. 165) observation that "reliving something is not merely remembering it." Reliving something is more than just remembering: it is *perceiving*. I should mention that Eddy Zemach (1969) has argued in a very different manner for a more general version of this view; though I do not find his argument convincing, his defense suggests that heterodox views of perception may inhabit a surprisingly attractive region of conceptual space.

This returns us to the above passage on eidetic memory. *Prima facie*, some of the children described there were in fact perceiving. For what they achieved is comparable to the achievement of the Martians described earlier, who could see by calling up previously frozen visual images at will. Insofar as these children’s memories, retrieved voluntarily, were appropriately causally and counterfactually dependent on a past scene, one could say that these children were engaging in a process of Martian-like snapshot perception.¹⁷

5 Episodic Memory vis-à-vis “Ordinary” Perception

One might object that the causal theory of perception can be easily amended to undercut the thesis that remembering is sometimes perceiving. It is tempting to think that specifying precisely what sort of causal connection is ‘appropriate’ will show that memory cannot be genuine perception. But even the most sophisticated accounts of the unruly appropriateness condition in the causal theory of perception, accounts which typically involve an alleged solution to the problem of deviant causal chains, do not seem to pose a problem for this thesis. In fact, it would appear that some episodic memories are able to satisfy each of the conditions—e.g., causally relevant aspect (Searle 1983), differential explanation (Peacocke 1979), teleological function (Davies 1983), reliable mechanism (Pendlebury 1994), and so on—for perception introduced by these proposals, and that whatever sort of causal connection we decide is appropriate for perception will also obtain in some episodic memories.

I suggest, then, that one must look elsewhere than the appropriateness condition in the causal theory of perception in order to ground an objection to the above argument. To this end, one might appeal to either one or both of two intuitively relevant differences between memories and “ordinary” perceptions. These differences were noted above, in the introduction. The first is that the perceptual objects of “ordinary” perception, but not memory, are always in the present. But we saw in §2 that there are clear cases of perception, such as the perception of Polaris, in which we perceive something in the past. Accordingly, we are left only to consider the second difference, which is that “ordinary” perception, but not episodic memory, requires the use of sensory organs.¹⁸

¹⁷Perhaps the proper linguistic analysis of statements concerning episodic remembering, statements of the form “A can still see *o*”, is one of the following: either “A perceives *o* in virtue of remembering”, or “A sees *o* memorially”. Either way, A sees *o*. An important difference between these analyses will be discussed in §6. Note that although the preceding discussion has focused directly on visual perception, the argument given appears to hold *mutatis mutandis* for all other perceptual modalities, as noted above.

¹⁸Another *prima facie* relevant difference between “ordinary” perceptions and memories lies in their seemingly disparate phenomenologies. While experiences of the former type are typically vivid, lively, and forceful, experiences of the latter type are typically dull, languid, and weak. An obvious response is that, first, many episodic memories (such as those discussed in §6 below) are extremely vivid, lively, and forceful. Second, there appear to be many “ordinary” perceptual experiences that are *not* vivid, lively, and forceful (e.g., perceptions one has when in a drunken stupor), from which it follows that this cannot be a generally necessary condition for perception. Third, and finally, it is not clear that we should expect memories and “ordinary” perceptions to have similar phenomenologies, even if they are both genuine perceptions. Different forms of perception possess disparate phenomenologies; so, insofar as the view is that memory and “ordinary” perception are different forms of perception, we should not expect the phenomenologies to be

This difference looks to be merely incidental, not one that is ontologically relevant. Because episodic memories involve being in a state that is qualitatively identical or very similar to a specific past experience, episodic remembering requires the operation of certain perceptual processes at the time of remembering. Accordingly, in episodic memory, the visual system, say, is still working; it is just doing so without the eyes. One might respond that this is not genuine perception because the eyes are not being used. But it seems arbitrary to privilege the eyes in this way, making them more important than the rest of the visual system. What is important to seeing, it seems, is not so much the eyes themselves, but the proper functioning of the visual system. Although the eyes normally play a role in vision, there seems to be no reason to assert *a priori* that the visual system cannot function properly in the eyes' absence. In fact, it seems reasonable to suppose that it is possible to see without the eyes (as in, e.g., artificial and prosthetic vision).¹⁹ In light of this, it is unclear how one could defend a response that appeals to the absolute necessity of the eyes without either unduly disqualifying clear cases of perception or, worse, begging the question.

6 Constructive Memory?

There is one further objection worth considering. Walton (1984; p. 275, n.12) has suggested that one might avoid the heterodox thesis that certain memories are perceptions by noting that memory is *constructive*, suggesting that memories are more like paintings, whose contents depend upon an agent's beliefs, than photographs or perceptions, whose contents do not. If this were so, it would imply that episodic memories are not perceptions because they lack the requisite counterfactual dependence on their objects.

However, although it appears correct to say that beliefs *can* alter how one remembers things, there appears to be no argument that they *must* do so—at the very least, not that they must do so any more, or in a different manner, than in “ordinary” perception. In fact, empirical research shows that some memories contain at least as much detail of the remembered scenes as some “ordinary” perceptions do of the perceived scenes.

Consider first the eidetic memories of the children in Allport's study: even if these memories are constructive in nature, it seems that they contain enough of the detail of the remembered picture to count as veridical. Of course, they are by no means perfect.

similar.

¹⁹Let me attempt to describe such a case. Suppose that an apple is before Sue, and Sue is looking at the apple. Suppose that Joe has a visual experience as of that apple—that is, Joe's visual experience matches the apple. Now, Joe's brain is connected to Sue by a sophisticated mechanism that (let us suppose) preserves causal relations. So, Joe's experience is appropriately counterfactually and causally dependent on the apple. Joe sees the apple. If this is right, it shows that a perceptual object need not be “before one's eyes” in order for one to see it, and thus one's sensory organs need not operate at the time of perceiving in order for one to see.

Now, this case suggests another. Suppose that Joe is not connected to Sue, but someone else. Actually, the person to whom Joe is connected is Joe himself, in the sense that his present brain states are connected to his past brain states (causally, as it were). In such a case, the sophisticated mechanism that preserves causal relations is simply Joe's brain, and his present visual experience is simply an episodic memory of a past apple. If Joe sees the apple when he is connected to another person, then so long as appropriate causal and counterfactual dependence is maintained, there appears to be no reason to deny that Joe sees the apple when he is connected to himself.

But just as perception need not be perfect to count as genuine perception (e.g., one can still see without one's glasses or contact lenses), an episodic memory need not be perfect to count as a genuine memory (e.g., a groom can still remember his wedding scene even if he cannot recall the lacework on the fringe of his bride's dress). There is little room to doubt that for the children in Allport's study, a different picture would have caused correspondingly different memories, perhaps even so as to exhibit a strict isomorphism between the memories and the remembered picture. If so, it appears appropriate to say that their memories possess the requisite counterfactual dependence on their objects.²⁰

Second, empirical research on "flashbulb memory", where significant events are remembered in detail, also suggests that episodic memories, though imperfect, sometimes possess the counterfactual dependence necessary to achieve veridicality (Brown and Kulick 1977). In *Searching for Memory*, Daniel Schacter (1996; p. 196ff) points out that although flashbulb memories are not "etched indelibly in the mind", they can nevertheless be "highly accurate." For instance, many Americans remember with surprising accuracy the details of their experience when they were told in 1963 that President Kennedy had been assassinated. Although their memories of the surrounding events are often hazy, they can remember the scene in question in such a way that it is "present" to them once again.

To be sure, there are many cases in which this is not so. And there are many cases in which it is difficult, if not impossible, to judge whether or not a given memory possesses the appropriate sort of counterfactual dependence. But there are at least a few cases, such as those involving eidetic and flashbulb memory, in which a given memory contains the requisite amount of detail of the remembered scene. These do not appear to be isolated cases, for many have argued that memory is generally reliable, in which case it produces memories that are veridical more often than not.²¹ And given the arguments offered in §§4–5, a fair principle seems to be that, all things being equal, a given episodic memory is an instance of perception just in case it is veridical.

7 Conclusion

Part of the shock of Dr. Seuss' rhyme, in which the Cat in the Hat asserts that he can read with his eyes shut, comes from the simple fact that we typically read—and, more generally, *see*—with our eyes open. Indeed, insofar as most perception is "ordinary" perception, *most* cases of perception are of what is present and involve the operation of one's sensory organs. But we have seen that there are reasons to think that "ordinary" perception is not the only perception there is. Plausibly, some episodic memories are

²⁰ Indeed, as Martin Davies (1983; p. 417) points out, "very little in the way of matching is intuitively required for an experience to be a perception of an object." In addition, note that given the *prima facie* theory-ladenness of observation and the influence of top-down processing theories in cognitive science, many believe that "ordinary" perceptions are themselves constructive insofar as they counterfactually depend, at least to some degree, on the perceiver's beliefs and desires. If so, the objection under consideration suffers a still-birth. See Hanson (1958), Bruner (1957), Gregory (1966/1997), Fodor and Pylyshyn (1981) and Marr (1982) for discussions of the theory-ladenness of observation and/or top-down processing.

²¹ See, e.g., Harman (1973; 1986), Pollock (1974; 1986), Martin (1992), Dummett (1992), Plantinga (1993), Burge (1993; 2003), Chisholm (1997), Hamilton (1998), and Owens (2000).

instances of another form of perception, one that allows us to (visually) perceive the past, even with our eyes shut.

One might consider this conclusion highly counterintuitive, and reject it on the grounds that it is plainly unacceptable. In such a case, one may take the argument of this paper to be a *reductio ad absurdum* of the causal theory of perception,²² one which forces us to reconsider the plausibility of this widely held analysis.

However, it is not apparent that this is a defensible response. For the considerations outlined in the preceding pages—in particular, the possibility of snapshot perception, the way we sometimes talk about our remembrances, and the fact that episodic memories are in ontologically relevant ways on par with “ordinary” perceptions—suggest that an adequate analysis of perception *ought* to include some episodic memories. In light of this, I recommend that we not dismiss the argument of this paper without first considering where it leaves us.

If the argument is accepted, it seems that we can adopt either one of the following two positions. First, we could understand it as a defense of the claim that episodic memory is a form of perception, namely, memorial perception, which is related to but distinct from other forms of perception (such as visual, auditory, and olfactory perception). This position would do some justice to the seeming platitude that episodic memories are distinct from other, more “ordinary”, forms of perception, while at the same time acknowledging their ontologically relevant similarities. Accordingly, we would be in possession of a non-reductive account of episodic memory that places it alongside other forms of perception. However, it is also possible to understand the argument as a defense of the reduction of episodic memory to previously established forms of perception, such as seeing, hearing, and smelling. In other words, episodic memory is nothing but seeing, hearing, or smelling in a particular way, in which case an episodic memory of, say, a visual scene ψ just is a way of seeing ψ , namely, memorially. In addition to giving us a reductive account of episodic memory (presumably, a conclusion desirable to those who privilege ontological parsimony), this position makes the causal theory of perception look very impressive, for it endows this widely held analysis with broad explanatory scope and unique predictive success. Although neither interpretation is forced by the above argument, both leave us without a principled distinction between perception and memory.

References

- William Brewer. What is recollective memory? In D. C. Rubin, editor, *Remembering our Past*. Cambridge University Press, Cambridge, 1996. 7
- Roger Brown and James Kulick. Flashbulb memories. *Cognition*, 5:73–99, 1977. 10
- Jerome Bruner. On perceptual readiness. *Psychological Review*, 64:123–152, 1957. 10
- Tyler Burge. Memory and persons. *Philosophical Review*, 112(3):289–337, 2003. 6, 10

²²Or, at the very least, a *reductio* of many of the causal theory’s most plausible forms.

- Tyler Burge. Content preservation. *Philosophical Review*, 102:457–488, 1993. [10](#)
- Noël Carroll. Towards an Ontology of the Moving Image. In C. A. Freeland and T. E. Wartenberg, editors, *Philosophy and Film*. Routledge, New York, 1995. [5](#)
- Noël Carroll. *Theorizing the Moving Image*. Cambridge University Press, Cambridge, 1996. [5](#)
- Roderick Chisholm. *A Theory of Knowledge*. Prentice Hall, New Jersey, 1997. [10](#)
- Andy Clark and David Chalmers. The Extended Mind. *Analysis*, 58:10–23, 1998. [6](#)
- Paul Coates. Deviant Causal Chains and Veridical Hallucination: A Problem for the Anti-Causalist. *The Philosophical Quarterly*, 50(200):320–331, 2000. [2, 3](#)
- Jonathan Cohen and Aaron Meskin. On the epistemic value of photographs. *Journal of Aesthetics and Art Criticism*, 62(2):197–210, 2004. [5, 6](#)
- Gregory Currie. *Image and Mind: Film, Photography, and Cognitive Science*. Cambridge University Press, Cambridge, 1995. [5](#)
- Martin Davies. Function in perception. *Australasian Journal of Philosophy*, 61:409–426, 1983. [2, 3, 8, 10](#)
- Michael Dummett. Memory and testimony. In *The Seas of Language*. Oxford University Press, Oxford, 1992. [10](#)
- Jerry Fodor and Zenon Pylyshyn. How direct is visual perception? some reflections on gibson's ecological approach. *Cognition*, 9:139–196, 1981. [10](#)
- Alvin Goldman. Discrimination in perception. *Journal of Philosophy*, 73:771–791, 1976. [2, 3](#)
- R. L. Gregory. *The Intelligent Eye: The Psychology of Seeing*. Princeton University Press, Princeton, New Jersey, 1966/1997. [10](#)
- H. P. Grice. The causal theory of perception. *Aristotelian Society*, Supplementary Volume, XXXV:121–152, 1961. [2, 3](#)
- H. P. Grice. Some remarks about the senses. In R. J. Butler, editor, *Analytic Philosophy*, first series. Blackwell, Oxford, 1962. [3](#)
- Andy Hamilton. False Memory Syndrome and the Authority of Personal Memory-Claims: A Philosophical Perspective. *Philosophy, Psychiatry, and Psychology*, 5(4): 283–297, 1998. [10](#)
- Norwood Russell Hanson. *Patterns of Discovery*. Cambridge University Press, Cambridge, 1958. [10](#)
- Gilbert Harman. *Thought*. Princeton University Press, Princeton, 1973. [10](#)
- Gilbert Harman. *Change in View*. MIT Press, Cambridge, MA, 1986. [10](#)

- Leo M. Hurvich. *Color Vision*. Sinauer Associates, Inc., Cambridge, MA, 1981. [2](#)
- Frank Jackson. *Perception: A Representative Theory*. Cambridge University Press, Cambridge, 1977. [2](#), [3](#)
- David Lewis. Veridical Hallucination and Prosthetic Vision. *Australasian Journal of Philosophy*, 58:239–49, 1980. Reprinted in [Lewis \(1986\)](#). [1](#), [2](#), [3](#)
- David Lewis. *Philosophical Papers, II*. Oxford University Press, Oxford, 1986. [13](#)
- David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, San Francisco, 1982. [10](#)
- C. B. Martin and Max Deutscher. Remembering. *Philosophical Review*, 75:161–196, 1966. [7](#)
- M. G. F. Martin. Perception, concepts, and memory. *Philosophical Review*, 101(4): 745–763, 1992. [10](#)
- Alva Noë. On what we see. *Pacific Philosophical Quarterly*, 83:57–80, 2002a. [6](#)
- Alva Noë. Is the visual world a grand illusion? *Journal of Consciousness Studies*, 9 (5/6):1–12, 2002b. [6](#)
- Alva Noë. Causation and perception: The puzzle unravelled. *Analysis*, 63(278):93–100, 2003. [1](#), [6](#)
- Alva Noë. *Action in Perception*. MIT Press, Cambridge, MA, 2004. [6](#)
- David Owens. The authority of memory. *European Journal of Philosophy*, 7(3):312–329, 2000. [10](#)
- Christopher Peacocke. *Holistic Explanation: Action, Space, Interpretation*. Clarendon Press, Oxford, 1979. [3](#), [8](#)
- Michael Pendlebury. Content and causation in perception. *Philosophy and Phenomenological Research*, 4(54):767–786, 1994. [3](#), [8](#)
- Alvin Plantinga. *Warrant and Proper Function*. Oxford University Press, Oxford, 1993. [10](#)
- John Pollock. *Knowledge and Justification*. Princeton University Press, Princeton, 1974. [10](#)
- John Pollock. *Contemporary Theories of Knowledge*. Hutchinson, London, 1986. [10](#)
- Gilbert Ryle. *The Concept of Mind*. Barnes and Noble, New York, 1949. [2](#)
- Daniel L. Schacter. *Searching for Memory: The Brain, the Mind, and the Past*. Basic Books, New York, 1996. [10](#)
- John Searle. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, Cambridge, 1983. [3](#), [8](#)

- Robert L. Solso. *Cognition and the Visual Arts*. MIT Press, Cambridge, MA, 1994. 2, 4
- Peter F. Strawson. Causation in perception. In *Freedom and Resentment and Other Essays*. Methuen, London, 1974. 2
- John Sutton. Memory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. 2004. URL <http://plato.stanford.edu/archives/sum2004/entries/memory/>. 1
- Kendall L. Walton. Transparent pictures: On the nature of photographic realism. *Critical Inquiry*, 11:246–277, 1984. 4, 5, 9
- Kendall L. Walton. Looking again through photographs: A response to edward martin. *Critical Inquiry*, 12:801–808, 1986. 4
- Kendall L. Walton. On pictures and photographs: Objections answered. In R. Allen and M. Smith, editors, *Film Theory and Philosophy*. Clarendon Press, Oxford, 1997. 4, 5
- David Wiggins. Locke, Butler, and the Stream of Consciousness and Men as Natural Kind. In A. Oksenberg-Rorty, editor, *The Identities of Persons*. University of California Press, Berkeley and Los Angeles, 1976. 7
- E. M. Zemach. Seeing, ‘Seeing’, and Feeling. *Review of Metaphysics*, 23:3–24, 1969. 6, 7

Thought-Experiment Intuitions and Truth in Fiction

Jonathan Ichikawa*
BROWN UNIVERSITY

1 Introduction

WE START (where else?) with a story.

Professor McStory was teaching an epistemology course. “It’s surprisingly difficult,” he said, “to provide an analysis of knowledge.” A student in the back row raised his hand. “Yes, Brian?”

“I don’t see why we should think it’s so hard to provide an analysis of knowledge,” Brian said. “I think I know what knowledge is – it’s justified true belief.” Professor McStory marveled privately at how conveniently the dialectic was progressing. “Let me tell you a story.” The class leaned forward attentively. “Listen to this, and see if you think that knowledge is justified true belief.

Joe had left his watch at home, and he wanted to know what time it was. Luckily for him, there was a clock on the wall. As a proficient reader of clocks, Joe had no difficulty in determining that the clock read 10:15. ‘Good,’ thought Joe, ‘I still have fifteen minutes until Mr. Pumbleton will be expecting me.’ Joe had arranged an important meeting at 10:30.

However, things were not as they appeared. Poor Joe had formed his belief on the basis of an inaccurate clock! The clock was fifteen minutes fast, and it was already 10:30! But fate smiled on Joe that day – for due to a careless error on the part of Mr. Pumbleton’s secretary,

*ichikawa@gmail.com

Mr. Pumbleton thought Joe's appointment was at 10:45. So Joe's belief about how much time he had until Mr. Pumbleton would consider him late was true after all.

Think about Joe's belief about how much time he had," Professor McStory suggested. "It was justified, and it was true, but we can see that it was not knowledge. So knowledge isn't justified true belief." Brian and the rest of the class thought about the story and conceded that Professor McStory was right. Justified true belief wasn't the same as knowledge.

Philosophers often construct thought experiments to test theories; Professor McStory's presentation is a fairly typical example. Central in the project is the invocation of intuition. "Intuition" didn't appear in my story, but it could have. Instead of saying "we can see that it was not knowledge", Professor McStory might've said "intuitively, it was not knowledge." This is standard philosophical methodology. But there is apparent reason to be skeptical about the invocation of intuition to advance philosophical claims; there is now an extensive literature challenging the role of intuitions in philosophical inquiry. Skeptical arguments threaten the whole project of conceptual analysis.¹ If the critics are right, then Professor McStory is presupposing a problematic philosophical methodology.

More recently,² Timothy Williamson has offered a sort of a defense of Professor McStory's strategy. On Williamson's view, there is no special faculty of intuition that we must invoke in order to judge that Joe doesn't know. Instead, the 'intuition' we form is a counterfactual judgment, no different from the naturalistically innocuous counterfactual judgments we make every day. Although Williamson is aptly read as defending McStory's invocation of a thought experiment, it is not right to understand Williamson's project as a defense of traditional philosophical methodology, for the judgments with which he identifies 'intuitions' are scarcely recognizable as the things traditional philosophers had in mind. In particular, on Williamson's view, the contents of the intuitions in question are contingent. This is problematic, from the standpoint of traditional philosophical methodology, for two reasons. First, tradition has it that intuitions like the Gettier intuition have modal content; there aren't supposed to be possible worlds where Joe's belief counts as knowledge. Second, relatedly, the standard view has it that intuitions like the Gettier intuition can be known *a priori*. Williamson's contingent counterfactuals could be known at best empirically.³ Williamson is like the postman in the classic comic strip who finally brings Calvin his package after 6–8 anxious weeks. And Williamson's 'intuitions' are like the disappointing beanie hat that has finally arrived, but doesn't let Calvin fly.

In what follows, I will attempt to deliver an account of thought-experiment intuitions that can fly. I hope to establish that philosophical intuitions such as the Gettier intuition do have modal content, and thus to leave open the question of whether we can

¹ See, for example, the various papers in (DePaul and Ramsey 1998).

² Williamson (2005)

³ The problem isn't that contingent things cannot be known *a priori*; we know that some can. It's that the counterfactuals with which Williamson identifies thought-experiment intuitions are clearly not the kinds of things that could be contingent *a priori*.

know them *a priori*. I suggest that what is missing from Williamson's presentation is a proper understanding of the role of truth in fiction. Once this is supplied, we will have no compelling reason to understand intuitions like the Gettier intuition⁴ as having merely contingent content. I will begin with a restatement of Williamson's argument, then attempt to establish a way around it.

2 Williamson's Argument

How shall we formalize the Gettier argument? Here is a first stab:

$K(x, p)$: x knows that p .
 $JTB(x, p)$: x has a justified true belief that p .
 $GC(x, p)$: x stands to p as in the given Gettier story.

(1) $\diamond \exists x \exists p GC(x, p)$
 (2_n) $\Box \forall x \forall p [GC(x, p) \rightarrow (JTB(x, p) \& \sim K(x, p))]$
 \therefore (3) $\diamond \exists x \exists p (JTB(x, p) \& \sim K(x, p))$.

In English:

(1) It's possible for some x to stand to some p as in the Gettier story.
 (2_n) Necessarily, if x stands to p as in the Gettier story, then x has a justified true belief that isn't knowledge.
 \therefore (3) It's possible for someone to have a justified true belief that isn't knowledge.

In this formulation, (2_n) represents the Gettier intuition. And this Gettier intuition is a necessity claim, just as traditional philosophers think it is. But Williamson argues that this cannot be the correct formulation of the Gettier intuition. For, (2_n), says he, is false. The key to seeing this is that Gettier stories, such as the one that Professor McStory tells in my introduction, are inevitably underspecified. There are details that are left out of the text, and some of them are important. It is possible, consistent with the text of the story, to insert pieces that will prevent the story from presenting a case of justified true belief that isn't knowledge. Call a story in which the given Gettier text is true, but there is no case of non-knowledge justified true belief (hereafter NKJTB), **bad** Gettier cases. Gettier cases with NKJTB are **good** ones.

Following are two ways we could understand Professor McStory's story as a bad Gettier case by adding consistent bits to it:

(a) Joe had left his watch at home, and he wanted to know what time it was.
 Luckily for him, there was a clock on the wall. As a proficient reader of clocks,

⁴I follow Williamson in using the Gettier intuition as a paradigm, but the point will generalize to all thought-experiment-based intuitions about particular fictional cases. So what I have to say here should bear also on the modal status of intuitions about whether or not twater is water, Mary learns something new, it is wrong to blow up a fat man to save several lives, etc.

Joe had no difficulty in determining that the clock read 10:15. **Joe decided to believe that the clock was accurate, even though it was only one of dozens of clocks decorating the wall, and each clock showed a different time.**

- (b) As a proficient reader of clocks, Joe had no difficulty in determining that the clock read 10:15. “Good,” thought Joe, “I still have fifteen minutes until Mr. Pumbleton will be expecting me.” **Joe knew that Mr. Pumbleton was watching him via closed-circuit television and would be able to accurately predict that it would take him fifteen minutes to reach his office from his current location.**

Both (a) and (b) are bad Gettier cases—in neither case does Joe have NKJTB. In (a), Joe’s belief fails to be justified; in (b) his belief qualifies as knowledge. Williamson argues that given the possibility of bad Gettier cases, (2_n) , the necessity claim, cannot be right. For sometimes, when x stands to p as in the Gettier story, x is standing to p as in a *bad* Gettier story. So (2) , the Gettier intuition, must be weakened. Here is Williamson’s version of the argument:

- (1) $\diamond \exists x \exists p \text{GC}(x, p)$.
 (2_{cf}) $\exists x \exists p \text{GC}(x, p) \square \rightarrow \forall x \forall p [\text{GC}(x, p) \square \rightarrow (\text{JTB}(x, p) \& \sim \text{K}(x, p))]$.
 \therefore (3) $\diamond \exists x \exists p (\text{JTB}(x, p) \& \sim \text{K}(x, p))$.

In English:

- (1) It’s possible for some x to stand to some p as in the Gettier story.
 (2_{cf}) Counterfactual: If some x were to stand to some p as in the Gettier story, then anyone who stood to any proposition in the same way would have NKJTB.
 So, (3) It’s possible to have NKJTB.

This formulation of the Gettier argument, like the first, is valid. It is also sound, for the weaker (2_{cf}) is true, where the original (2_n) was false. It’s possible to stand to a proposition in a way matching the text’s description of Joe’s relation to *I have fifteen minutes* without NKJTB – but the relevant counterfactual does not include the bad cases. *If* someone *were* to stand to a proposition in the way matching the text, it *wouldn’t* be bad, even though it’s logically possible to be in such a position in a bad way. *If* someone *were* to stand to a proposition in a way matching the story, it would be good.

But now we’re just a short step from the conclusion I want to resist: (2_{cf}) , which is meant to correspond to the Gettier intuition, is true, but it is only contingently true. This is so because the truth of counterfactuals depends on characteristics of the *nearest* possible worlds where the antecedent is true.⁵ And which worlds are the nearest worlds depends on what the actual world happens to be like. To know this counterfactual requires knowing many contingent facts about the actual world. If the Gettier intuition is like this, then it is not the sort of thing that traditional philosophy takes it to be.

⁵ For simplicity of language, I assume a possible-worlds semantics for counterfactuals. This assumption is dispensable for my project.

Williamson’s account leaves intuitions as mere judgments about contingent matters of fact. As such they cannot be known *a priori*.

3 Responses

Williamson considers and rejects various alternate formulations of the Gettier intuition which attempt to preserve its modal content. It is worth rehearsing his arguments against two of these responses, as they will provide useful points of comparison with my own solution. Here is the first alternate formulation Williamson considers:

$$(2_{cf\exists}) \quad \exists x \exists p \text{GC}(x, p) \Box \rightarrow \exists x \exists p [\text{GC}(x, p) \Box \rightarrow (\text{JTB}(x, p) \& \sim \text{K}(x, p))].$$

(If some x were to stand to some p as in the Gettier story,
then there would be some example of NKJTB.)

This weakening of (2_{cf}) still provides a valid Gettier argument, but Williamson shows that this move cannot save the traditional understanding of intuitions, for two reasons. First, it is implausible that this is what we mean by the Gettier intuition; the intuition is about a particular situation. We don’t read about Joe and Mr. Pumbleton and judge that *somebody* has NKJTB – we judge that *Joe* does. So the logical structure of $(2_{cf\exists})$ doesn’t match that of the Gettier intuition. Second, even $(2_{cf\exists})$ is contingent; in worlds with lots of bad Gettier cases, including some similar to our Gettier story, and no good ones, the counterfactual is still false.

Another rejected solution move I’d like to consider is to concatenate an implicit ‘things are otherwise as we expect’ clause to the Gettier story. Again, Williamson offers two reasons to reject the move. The first is that what *we* expect is still a contingent matter. One way to see this is to recognize that we’re introducing another counterfactual: ‘what we expect’ will have to be evaluated as something like ‘what the relevant members of the community *would* expect *if they were given* the story’. So this move won’t help the modal intuitions team. The second objection Williamson raises is that invocation of a clause like this begins to threaten the status of premise (1). It’s obvious that it’s possible for the text Professor McStory gave to have been true; it’s much less obvious that if the story had concluded, “and things were otherwise as you’d expect”, it could have been true. Who knows (*a priori*) what crazy things some people might expect?

4 Truth in Fiction

Williamson gives us the predicate $\text{GC}(x, p)$, x stands to p as in the Gettier story. The way he treats it, $\text{GC}(x, p)$ holds any time that x stands to p in a way that makes each sentence of the body of text true. But there are more resources available here. I suggest that what is missing from Williamson’s analysis, and from the two attempted responses to it, is the recognition of the difference between a text and a *story*.⁶ There is more to a story than the literal claims of the sentences used to tell it. Insights from the philosophy of fiction can help us to understand the subclass of fictions that are thought experiments.

⁶For a compelling argument that the story is not identical to the text, see Currie (1991))

In particular, we can invoke the notion of *truth in fiction*.⁷ Some, but not all, fictional truths are explicitly stated in the text. It is true in my fiction that Professor McStory was teaching an epistemology course; this is so because my text included the sentence, “Professor McStory was teaching an epistemology course.” Some fictional truths are not so directly identified. It’s true in my fiction that Brian had exactly two eyes, even though I didn’t say so. It’s even true in my fiction that Professor McStory was male, even though it’s consistent with the text I wrote that the protagonist was a woman. The challenge of a theory of truth in fiction is to explain all of this.⁸

Using truth in fiction as a placeholder for now, we can make use of it in a reformulation of the Gettier argument. Recall the ‘first stab’ formulation from §2:

- $$\begin{aligned} (1) & \quad \diamond \exists x \exists p \text{GC}(x, p) \\ (2_n) & \quad \Box \forall x \forall p [\text{GC}(x, p) \rightarrow (\text{JTB}(x, p) \& \sim \text{K}(x, p))] \\ \therefore (3) & \quad \diamond \exists x \exists p (\text{JTB}(x, p) \& \sim \text{K}(x, p)) \end{aligned}$$

The problem with this formulation was that it is unsound; (2_n) is false, because there are bad stories that can be described by the Gettier text. However, now that we are armed with truth in fiction, we may try to reformulate the argument in a way that doesn’t move to Williamson’s merely contingent counterfactual (2_{cf}). We replace GC, which merely demands that the claims of the Gettier text be met, with GC^{tf}:

- $$\text{GC}^{\text{tf}}(x, p): \quad x \text{ stands to } p \text{ in the way that it is true in the fiction that Joe stands to } I \text{ have half an hour.}$$

The worry was that there are bad stories for the particular text; GC^{tf} specifies that we’re talking about this particular *story*, instead of this particular text. So we may say:

- $$\begin{aligned} (1_{\text{tf}}) & \quad \diamond \exists x \exists p \text{GC}^{\text{tf}}(x, p) \\ (2_{\text{tf}}) & \quad \Box \forall x \forall p [\text{GC}^{\text{tf}}(x, p) \rightarrow (\text{JTB}(x, p) \& \sim \text{K}(x, p))] \\ \therefore (3) & \quad \diamond \exists x \exists p (\text{JTB}(x, p) \& \sim \text{K}(x, p)). \end{aligned}$$

That is:

- $$\begin{aligned} (1_{\text{tf}}) & \quad \text{It’s possible to stand to a proposition the way that it’s true in the Gettier fiction that Joe stands to his belief.} \\ (2_{\text{tf}}) & \quad \text{Necessarily, anyone who stands to a proposition in the way that it’s true in the fiction that Joe does to his belief would have NKJTB.} \\ \therefore (3) & \quad \text{It’s possible to have NKJTB.} \end{aligned}$$

There’s no mention of counterfactual here. So have I defeated the argument that the Gettier intuition is contingent? In this formulation, (2_{tf}), the Gettier intuition, invokes

⁷The seminal piece on truth in fiction is Lewis (1978). See also Walton (1990) and Currie (1990). For recent developments in truth in fiction, see Hanley (2004) and the papers cited therein. Walton uses the term ‘fictionality’ where others, including myself, use ‘truth in fiction’.

⁸Also, fictions are not determinate on every proposition. It’s neither true nor false in my fiction that Professor McStory skipped breakfast that morning.

the notion of truth in fiction.⁹ So everything depends on our true theory of truth in fiction; the Gettier intuition can be *a priori* if we can access fictional truths, given texts, *a priori*. We might be worried about the plausibility of this line. On one of Lewis's views, for example, truth in fiction *is* a counterfactual: *p* is true in the fiction iff, were the story true, *p* would be true.¹⁰ This would quickly collapse my formalization into Williamson's, with respect to the modal and epistemic status of the Gettier intuition. But there are good reasons to think that this view is not right, and some of them locate the problem just with the counterfactual.¹¹ The best sort of objection to this view is raised by Lewis himself:

Suppose I write a story about the dragon Sculch, a beautiful princess, a bold knight, and what not. It is a perfectly typical instance of its stylized genre, except that I never say that Sculch breathes fire. Does he nevertheless breathe fire in my story?¹²

It's not true in the actual world that if there were a creature matching Sculch's textual description, it would breathe fire, but it's true in the fiction that Sculch does, so the counterfactual account of truth in fiction is wrong. But is there good reason to think that the correct account will serve us better in this respect? Any theory of truth in fiction which demands a contingent judgment will render the Gettier intuition in my formalization as a judgment about a mere contingent matter of fact. Is there good reason to think that the correct theory of truth in fiction will allow us to come to a judgment of necessity? Is there hope for the traditional view that thought-experiment intuitions are available *a priori*? I think that there is, and in what follows, I shall try to spell out some grounds for optimism.

I will not attempt in these short pages to defend a particular view on the *analysis* of truth in fiction. This is not necessary for my project—all that I need is an *a priori* way to know fictional truths. Such a technique need not provide an analysis; indeed, the one I will propose fails in that regard. But if I can deliver an *a priori* way to know that proposition *p* is true in fiction *F*,¹³ then I'll have saved traditional thought-experiment intuitions from Williamson's argument.¹⁴ The solution I have in mind derives from Kendall Walton. It begins with this major theme, which is now a mainstay in the philosophy of fiction: to engage properly with a fiction is to engage in a certain kind of game of make-believe. Walton gives us this principle:

(TFI) Principles of generation are functions from texts to prescribed imaginings; the prescribed imaginings are the fictional truth.¹⁵

⁹So it is a consequence of my view that to have intuitions about thought experiments, one must have some understanding of truth in fiction. This seems correct.

¹⁰Lewis (1978). This is (paraphrased) Lewis's *Analysis I*, the simplest of several theories of truth in fiction he puts forward.

¹¹For several good arguments against *Analysis I*, see Currie (1990; §2.3).

¹²Lewis (1978), p. 274. Because of the possibility that it is a conceptual truth that dragons breathe fire, Lewis goes on to further stipulate that the word "dragon" isn't used in the text of the story.

¹³This will, of course, entail that there *is* some analysis of truth in fiction that renders it available to *a priori* recognition. It is beyond my project here to provide that analysis. My concern is epistemic, not metaphysical.

¹⁴This may raise a new worry about my premise (1_{tf}). I shall return to this worry below.

Fictional texts, on Walton's view, are props in games of make-believe. The game—the fiction—includes the text and a series of instructions—what Walton calls *principles of generation*. These principles take the fictional text as input and deliver a series of prescriptions to imagine. If we understand the principles of generation, and we have the text available to us, then we can derive the prescribed imaginings with no (further?) need for contingent experience. Once we have the prescribed imaginings, we may make use of Walton's (TFI) principle to identify truths in the fiction.

This is an attractive line, and would serve our purpose well, but there are problems with (TFI). Not all prescribed imaginings are true in the fiction; sometimes, fictions invite us to imagine fictional mere possibilities, or to imagine an obvious but fictionally false scenario to set up the surprise ending. Perhaps these problems can be overcome, but as it stands, it would be unwise to rest our theory of thought-experiment intuitions on (TFI). Happily, we can work with an uncontroversial principle at a higher level of abstraction. We needn't consider principles of generation for prescriptions to imagine; we may work directly with principles of generation for fictional truths. We abstract from Walton's principle, thus: fictional texts combine with principles of generation to generate fictional truths (which play a certain role in games of make-believe).

(TFTF) Principles of generation are functions from fictional texts to fictional truths.

So once we know the relevant principles of generation and the fictional text, we are in a position to derive the fictional truths. Since, on my suggestion, thought-experiment intuitions are intuitions about fictional truths, the relevant question comes to this: does knowledge of the principles of generation require judgments of contingent fact? If not, then it is plausible that we could know the modal (2_{tf}) *a priori*.

I suggest that there is good reason to think that such knowledge does *not* require such judgments of contingent fact. After all, we're the ones playing the game. Surely I can take a text and give myself some rules of generation about what to imagine, thus making up a game. None of this seems like the kind of thing that would require judgments of contingent fact. Once I do this, I need only reflect on the rules of generation I've given myself. Once I know the rules, I can know that necessarily, these rules (identified rigidly) dictate that such-and-such fictional text generate such-and-such fictional truths. To go back to Joe and the clock, Brian can know the rules of the fiction-game, and he can know that necessarily, given the text of the fiction, the rules of generation make it true in the fiction that Joe has NKJTB. So, necessarily it's true in the fiction (*this* fiction) that Joe has NKJTB. So, necessarily, anyone who is in the same position as the one that it is true in the fiction that Joe is in will have NKJTB. This is exactly what I suggested the correct formulation of the Gettier intuition to be.

$$(2_{tf}) \quad \Box \forall x \forall p [GC^{tf}(x, p) \rightarrow (JTB(x, p) \& \sim K(x, p))].$$

¹⁵Walton (1990), p. 39. (My paraphrase.) Walton offers this account as an attempt to *analyze* truth in fiction. It fails as an analysis, partially because it gets the order of explanation between imagining and truth in fiction backwards. But this is no failure to provide an epistemically useful sufficient condition for truth in fiction. See above.

If the story I've told is right, then Brian has acquired the Gettier intuition, with modal content, without relying on any contingent judgments. But there is an obvious objection to my story: I've claimed that Brian can access the rules of generation without relying on any contingent judgments. This may be problematic; surely, one might object, facts about the language, the social conventions, the author's intentions, and many other contingent facts influence what the correct rules of generation are, and likewise, what is true in the fiction.

There is obviously much that is right in this point, but I will show in the next section that it is consistent with my argument.

5 Differentiating Fictions

We may put the objection in these terms: perhaps we can set up and identify our own rules of generation without contingent judgments. But in order to know they're the *right* rules, we must know many contingent things. My response is simply that the latter requirement needn't concern us here; it's ok that we don't know they're the right rules of generation. I concede that we won't know *a priori* that we're imagining the things that the author intended us to, or the things that other people would imagine; but all of this is beside the point. On my view, the Gettier intuition is formalized in terms of truth in fiction, but it is not formalized in terms of truth in *the fiction that the author intended us to engage with*, or truth in *the fiction that most people would engage with, given the text*. One text, when coupled with two different sets of principles of generation, can generate two separate fictions. If the text I wrote at the start of this paper were in a culture in which all professors are women, then they would probably engage with a fiction in which Professor McStory is female. If it were read in a society of Cyclopes, it would generate a fiction in which Brian has only one eye. If it were read by a society of psychopathic philosophers, it might generate a fiction in which Professor McStory has murdered several students. So the availability of alternate rules of generation does not preclude *a priori* knowledge of truth in *this* fiction – the one I'm engaging with. The rules of generation, taken with the text, fix the fictional truths. This is all I need for the Gettier argument to go through.

- (1) It's possible for someone to be in a position like Joe's in this fiction I'm actually engaging with.
- (2_{tf}) Necessarily, anyone in a position like Joe's in this fiction I'm actually engaging with would have NKJTB.
- ∴ (3) It's possible to have NKJTB.

6 Final Points

Recall what went wrong with the two proposals I discussed in §3. The first proposal involved the weakening of (2_{cf}) to $(2_{cf\exists})$, a counterfactual existential claim. Williamson's first objection there was that the Gettier intuition is not a mere existential – it is about the particular person and situation we're considering. This is preserved in

my account; the intuition is about Joe, the character I'm engaging with. Williamson's second objection was that as a counterfactual, $(2_{cf\exists})$ was still merely contingent—it is false in worlds where there are lots of bad Gettier cases and no good ones. Not so for my (2_{tf}) : regardless of what the actual world is like, my rules of generation establish that it is true in the fiction that Joe has NKJTB.

The second proposed response, the addition of a 'things are otherwise as we expect' clause to the text of the story, is in many ways similar to mine. But Williamson's first objection demonstrates an advantage of my view; *things are otherwise as we expect* looks again like a contingent counterfactual, but my truth-in-fiction account does not. But what of Williamson's second objection here? Williamson worried that such a clause would threaten the obvious truth of premise (1), the claim that the text is possibly true. In this case, there *does* seem to be a parallel worry on my view. How can we be sure that the fiction we're engaging with describes a possible situation?

My response to this point is simply to concede it. There is room for a skeptical worry that the situation we're considering isn't possible. But we already knew that – when thought experiments describe impossible situations, the intuitions they generate are not to be trusted. It has been my project to establish that these intuitions have necessary propositions as contents, and in so doing to undercut the argument that we cannot know their contents *a priori*. A positive epistemology of thought-experiment intuitions would have to include at least a minimal epistemology of possibility. But I am optimistic that we can know *a priori* that the Gettier story (the one we're actually engaging with) is possible.

References

- Currie, G.: 1990, *The Nature of Fiction*, Cambridge University Press. 20, 21
- Currie, G.: 1991, Work and Text. Reprinted in John and Lopes (2004). 19
- DePaul, M. and Ramsey, W.: 1998, *Rethinking Intuition*, Rowman & Littlefield, Lanham. 16
- Hanley, R.: 2004, As Good as it Gets: Lewis on Truth in Fiction, *Australasian Journal of Philosophy* 82, 112–128. 20
- John, E. and Lopes, D. (eds): 2004, *The Philosophy of Literature: Contemporary and Classic Readings*, Blackwell, Oxford. 24
- Lewis, D.: 1978, *American Philosophical Quarterly* 15. Reprinted in Lewis (1983). 20, 21
- Lewis, D.: 1983, *Philosophical Papers, I*, Oxford University Press, Oxford. 24
- Walton, K.: 1990, *Mimesis as Make-Believe*, Harvard University Press, Cambridge, MA. 20, 22
- Williamson, T.: 2005, Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking, *Proceedings of the Aristotelian Society* 105, 1–23. 16

Making Sense of Perceptual Defeasibility

Ciara N. L. C. Fairley*

UNIVERSITY COLLEGE, LONDON

Davidson once claimed that the only thing that could justify a belief was another *belief*.¹ That seems very implausible. If I see a piece a burnt toast in front of me and believe, on that basis, that the toast is burnt, part of what seems to be justifying my belief is my perceptual experience. So, unless experiences *are* beliefs, it can't be true that beliefs can *only* be justified by other beliefs. They can also be justified, at least in part, by one's perceptual experiences. In this paper I'm not going to do any more to persuade you that what Davidson says is false. I think it's obvious that it's false. What's much less obvious, and what I want to look at, is whether perceptual experiences can justify perceptual beliefs *all by themselves*, or whether it's only in conjunction with other *beliefs* that they can do so. Let's call the first view - the view that perceptual experiences *can* justify perceptual beliefs *without* relying on other beliefs for their *justificatory* force - the Purely Experiential View.² And let's call the second view - the view that perceptual experiences can *only* justify perceptual beliefs in conjunction with other beliefs - the Mixed View. My question in this paper is whether the Purely Experiential View is defensible or whether the only defensible view is the Mixed View. I'll argue that the Purely Experiential View had better be defensible, if we want to avoid being overly sceptical about perceptual knowledge. But notice that *both* views are incompatible with Davidson's view - a view we might call the *Purely Doxastic View*.

Why should anyone think the Mixed View *is* the only defensible view? In other words, why should anyone think that when it comes to one's perceptual beliefs, other beliefs must *always* be playing a *justificatory* role? In the rest of this paper I'm going to discuss one particular argument in favour of that view. I think it's the best argument in favour of that view, so it's significant if it doesn't work. I'm going to call the argument the *Argument from Defeasibility*, since the argument claims that it follows from the fact that perceptual justification is *defeasible*, that it's always partly belief-based. And what I'm going to argue is that that doesn't follow; I think that perceptual justification *is* defeasible, but I'll show why that *doesn't* mean beliefs are always playing a justificatory

*ciarafairley2@hotmail.com

¹Davidson (1990).

²That is a view about where *justification* comes from; specifically, about whether it always comes from beliefs. That view is not undermined by the fact that you may need to have beliefs in order to *have* certain experiences; or you may need them to have the belief the experience is then claimed to justify. It's a view about whether perceptual beliefs necessarily draw part of their justification from other beliefs.

role. So I'll show why the Argument from Defeasibility doesn't give us any reason to think the Mixed View *is* the only defensible view. As far as that argument goes, we're free to endorse the Purely Experiential View and I'll give two reasons why that, in fact, *is* the view that we ought to endorse.

Since the Argument assumes that perceptual justification is defeasible, we need to know what that means, and we need to know how it's supposed to follow from this that such justification must always be partly belief-based. To say it's *defeasible* is to say it can be *defeated*. In the literature there are two sorts of cases in which people like to claim that a perceptual justification actually gets defeated. One is Goldman's famous barn example. In this case you're looking at the only real barn in an environment full of fakes that you can't discriminate from the real thing.³ The other kind of case is one I've taken from Mike Martin. Martin writes:

Suppose you know that I have a system capable of causing perfect hallucinations of oranges in subjects, and that I regularly run tests where I alternate the actual viewing of an orange with a perfect hallucination of one. You subject yourself to my machine. Unknown to you the machine has developed a serious fault and is incapable of causing hallucinations: if it looks to you as if there is an orange there, then that could only have been because you are seeing one. Nonetheless, you have information, which seems sufficient to make rational a doubt on your part as to whether there really is an orange before you when it looks to you as if that is what is there. . . you have reason sufficient to undermine the warrant that experience provides for judgment.⁴

Clearly, these are very different sorts of cases. What's in common is just that in both cases the subject *lacks* knowledge we think she might otherwise have had: you would know it was a barn were it not for the presence of the fakes, and you would know it was an orange if you could only ignore the misleading evidence. Those factors *prevent* you from acquiring knowledge and in that sense they 'defeat' your justification. Insofar as there is some single idea that people are trying to track when they talk about a subject's justification being defeated, that's it.⁵ The differences, and of course the interest, lie in the different explanation we give of why that's so - those explanations, in turn, give us different *senses* in which a subject's justification can be defeated.

Take the barn case. Here it doesn't look like we can explain why the subject fails to know in terms of the idea that it's *unreasonable* for her to believe what she does. There need be no sense in which she *ought* to be aware of the fakes. It's just bad luck they happen to be around. Here, the problem seems to lie with the way the subject's

³Goldman (1976).

⁴Martin (2001).

⁵The fact the cases are so different can lead one to think otherwise. Nonetheless, both get labelled as cases of 'defeat' and my focus on knowledge is an attempt to make sense of that common labelling. Though different authors clearly have different cases in mind. Lehrer and Paxson's original discussion of defeasibility clearly has Goldman-esque cases in mind where defeat does not deprive one of 'justification', but merely explains why one's justified true belief doesn't count as a case of knowledge. More recent discussion (like that one finds in Pryor and Peacocke) has tended to focus on the sort of case Martin has in mind. Here there is a much stronger temptation to think that defeat actually takes away one's justification.

externally hooked up to the world - in relying on *that* connection, she could easily have gone wrong. And that's why she's deprived of knowledge. I'm going to call cases like that, cases of External Defeat – to highlight the fact that it's the subject's external connection that's at fault. In the orange case, by contrast, things seem to be the other way round. Here, it's not that there's anything intrinsically defective about the way the subject's hooked up to the world - it need be no *accident* if what she believes turns out to be true. In Martin's case, the machine is broken: as he emphasises, if it looks to her as if there's an orange there that could *only* have been because she's seeing one. So the problem isn't that the subject's connection isn't basically a good one in some sense; it's just that she can't reasonably rely upon it, given the evidence available to her. She can't reasonably exploit the advantage it would normally afford her and persist in her usual perceptual beliefs. To do so would be unreasonable. And it's in that sense in which her justification is defeated. Let's call these kinds of cases, cases of Internal Defeat.

Some people deny that perceptual justification is capable of being defeated in these ways.⁶ I'm going to assume that's a mistake. So I'm not going to dispute that part of the Argument from Defeasibility. The important question for my purposes is whether that shows that the Mixed View is correct. Why should anyone think that? In other words, why think it follows from the fact one's justification can be defeated in these ways, that it's always partly belief-based? How, indeed, do beliefs even *seem* to get into the picture? To see how, let's go back to our examples. In our examples the presence of the fake barns and the evidence you're subject to the hallucination machine both defeat your justification; they're both circumstances whose actual obtaining deprives you of knowledge you might otherwise have had. I'm going to call circumstances like that, *defeating circumstances*. The central idea underlying the Argument from Defeasibility is that where we've got a belief that's *capable* of being defeated, it's always the case that at least part of what *justifies* you in that belief is the belief that you're not in *defeating circumstances*. So that's how beliefs get introduced – it's the belief that you're not in defeating circumstances –that, for example, it's *not* the case that you're in an environment littered with fake barns, or it's *not* the case that there's evidence you're subject to a hallucination machine - that you've got to have according to the Argument from Defeasibility. Now of course, for any given perceptual belief there's presumably an infinite number of ways in which one's justification might be defeated and we might not want to require that one actually believe with respect to each and every one of these ways that things aren't like *that*. So, perhaps we'll require instead that one just have the *general* belief that defeating circumstances don't obtain. But either way, there's always some belief that you have to have according to that Argument.

The point of the Argument, of course, is to suggest that those beliefs (however exactly we cash them out) are an essential part of what makes it the case that your ordinary perceptual beliefs are justified – they are part of the *source* of justification for one's perceptual beliefs. That's why the Argument from Defeasibility is meant to be argument in favour of the Mixed View. So how does the Argument support that claim? I can think of two reasons that might tempt us to think that. The first is that your perceptual beliefs – your beliefs about barns and the like – would presumably *not*

⁶Michael Ayers sometimes gives the impression of thinking this. See Ayers (1991; Ch 19, esp. pp. 170–171).

be justified if you believed that you *were* in a case in which defeating circumstances obtained. If you believed that you *were* in a landscape littered with barn facades or that you really *were* subject to a machine capable of producing perfect hallucinations, then the only reasonable thing to do would be to refrain from relying on your experiences in forming your perceptual beliefs. If you don't, your perceptual beliefs aren't justified. That can make it natural to suppose that part of what makes those beliefs justified *in the ordinary case* is just that you *believe* you're not in a case in which those conditions *do* obtain.

That's the first reason. The second reason looks at what kind of *explanation* we can give of what's going on in cases of defeat. The thought here is that in the case in which defeating conditions *do* obtain – in the case in which there are fake barns, and there is evidence you're hallucinating – the belief that those conditions don't obtain is false. So, if we assume a subject can't acquire knowledge where what she believes rests essentially upon a *false belief*, then we can explain *why* you don't acquire knowledge in cases of defeat if we assume that you have to *believe* those conditions don't obtain. We can explain it in terms of the no false lemmas requirement and that's a requirement we've got *independent* reason to accept. So it's a neat explanatory story.⁷

We have at least two reasons then to think the that belief defeating conditions don't obtain is playing some kind of justificatory role, just as the Mixed View claims it is; the first adverts to the *unreasonableness* of believing defeating conditions do obtain; and the second draws attention to the *explanatory* advantages of that suggestion. Given that's so, why should we want to resist that thought? What's *wrong* with thinking that part of what justifies your perceptual beliefs about oranges, and barns, *is* the belief that you're not in defeating circumstances? Here are two major problems with that suggestion. I think they're decisive – so I'll argue we ought to reject it. The first and most obvious problem is that in order to *believe* defeating conditions don't obtain you have to possess the *concepts* that figure in that belief. But is everyone really that conceptually sophisticated? Those who aren't, lack perceptual knowledge according to the Mixed View. And that looks very implausible. Call that the Hyper-Intellectualisation Problem.

The second problem concerns the epistemic status of the belief that you're not in defeating circumstances. That belief is clearly playing an epistemic role on the envisaged account: it's meant to be part of what *justifies* you in any given perceptual belief. But we might wonder how that can be so unless that belief is itself justified. That's certainly what the so-called Epistemic Regress Argument would have us think: according to that argument, beliefs can *only* play a justificatory role where they are themselves justified. And there's something very intuitive about that thought. But if this belief does need to be justified, then we face the problem of saying what justifies it, and a lack of any obvious candidates. No doubt one *can* acquire a justification for believing such conditions don't obtain, but it seems equally obvious that one has normally done no such thing. Call that the Epistemic Regress Problem.

These worries are independent of one another, but if either is well founded the Mixed View threatens to give rise to an unpalatable scepticism about perceptual knowledge. That, I take it, gives us *prima facie* reason to reject it. So what's the alternative?

⁷That is a line of thought Gilbert Harman is very keen to push. See [Harman \(1973\)](#).

The simplest way to avoid these problems is to avoid endorsing that view in the first place – hence, to deny the belief about defeating conditions *is* always playing a justificatory role. That means undermining the two considerations, which earlier seemed to make that look plausible. How might that be done? Let's start with the appeal to explanation. The idea here, remember, is we can explain why you don't know in cases of defeat if we assume that you have to believe defeating conditions don't obtain. We just appeal to the no false lemma's requirement – and we ought to accept that requirement anyway. I don't want to dispute any of that. So for me the important question is whether we can offer an alternative explanation – one that doesn't appeal to beliefs. That's hardest in cases of internal defeat. In cases of external defeat, after all, we can presumably just appeal to the fact that the subject isn't reliably hooked up to world in order to explain why she lacks knowledge – Goldman's barn spotter could easily have gone wrong in believing what she does, given the presence of the fakes – and knowledge, in general, excludes that kind of easy possibility. So there's no obvious temptation to introduce *beliefs* to explain what's gone wrong.

But that explanation needn't be available in cases of internal defeat. In Martin's orange case, there need be *no* sense in which, by ignoring the misleading evidence and simply persisting in one's ordinary perceptual beliefs, you could easily have gone wrong; we can perfectly well imagine subjects who are highly reliable in doing just that. Nonetheless, we think that doing so isn't a way to acquire knowledge. Indeed, unlike in cases of external defeat we think that doing so isn't even a way to acquire *justified* perceptual beliefs. Given the evidence available, you ought not to believe that there's an orange before you when that's how things look, *whether or not* you could easily have gone wrong in so doing. Here, I take it, there really is a temptation to introduce beliefs to explain why that's so. For here it can seem like the most straightforward explanation of why it wouldn't be reasonable to ignore that evidence – of why doing that isn't a way to acquire justified perceptual beliefs – is just that: in order for those beliefs to be justified you have to be justified in believing that those conditions don't obtain. That's why the evidence undermines your justification, because it renders that belief unjustified and that belief is part of what makes your perceptual beliefs justified.

I think that's the real force of the explanatory challenge; that's what's really intuitive about thinking that that belief is playing a *justificatory* role. That's why I think cases of internal defeat are the more fundamental in some ways – because there does seem to be a connection between whether or not your perceptual beliefs are justified and whether or not you're justified in believing that defeating circumstances don't obtain.⁸ Nonetheless, I think we should resist the further thought expressed above. I think we should resist the temptation to explain that connection – to explain why evidence those conditions do obtain can undermine your justification – in terms of the idea that part of what makes your perceptual beliefs justified to begin with is your being justified

⁸In cases of external defeat one might be justified in believing that defeating conditions did not obtain and *still* lack knowledge (assuming one can be justified in a false belief). So whether or not you know doesn't even seem to depend on whether or not you're justified in that belief. In cases of internal defeat, by contrast, it's tempting to think that your failure to be justified in the belief about the absence of defeaters is all that stands in your way and so all that prevents you from acquiring knowledge: after all, if you were justified in believing the evidence was misleading, you presumably *would* know – given that you could then reasonably ignore that evidence. This is the heart of Kripke's unpublished paradox in this context. See Harman for a discussion. Harman, G. *ibid.*

in believing those conditions do not obtain. That's the move I'm resisting here. And I've already conceded that I don't think we can explain why that's so in terms of the fact that you need be being *unreliable* in ignoring such evidence.⁹ What I think, then, is that we have to just accept it as a basic fact about perceptual justification that that's so. I think there may be no *deeper* explanation to be had of why you can't reasonably ignore such evidence - it's just a fact about the kind of knowledge you can acquire by means of your senses that you can't. I'm going to return to that thought again at the end, but for now that's all I'm going to say about the appeal to explanation.

So what about the first consideration in favour of the Mixed View— the one that points out that your perceptual beliefs wouldn't be justified if you believed that defeating circumstances *did* obtain. And so hypothesises that part of what makes them justified in the ordinary case is your believing such circumstances *do not* obtain. Can we avoid appealing to beliefs in order to make sense of what's going on here? Here, I take it, there really is an obvious alternative. Instead of requiring that one *believe* defeating circumstances *do not* obtain, why not just claim that one has to *lack the belief* that they *do*? If what's meant to be odd is just the *combination* of the two attitudes – one's believing *both* that 'that's a barn' *and* that 'one's surrounded by fakes one couldn't discriminate from the real thing' – then we can surely get round *that* by simply requiring that one just *not have* the latter belief.

Clearly, if the aim is avoid introducing *beliefs* into the justificatory story, this move will only work if we can distinguish between [*believing* it's *not* the case those conditions obtain] and [*not believing* that they *do* obtain]. That is, between believing that not-p and failing to believe that p. But, that's a distinction we obviously can and do make. As Evans points out 'I believe that p' admits of both internal and external negation.¹⁰ If that's right, then why isn't it enough to claim that in order for your perceptual beliefs to be justified you have to *lack the belief* that defeating conditions do obtain, without requiring you *believe* they do not? That needn't be to deny that conceptually sophisticated subjects sometimes do *believe* certain defeaters aren't present. But that doesn't show that what makes their perceptual beliefs justified *is* the fact that they have this belief. As far as what's justifies them goes, it's enough that they merely lack a certain belief. And if that's right, then their justification needn't be partly belief-based. Maybe that shows that a subject's perceptual experiences don't justify her perceptual beliefs *all* by themselves; but she doesn't need anything more in the way of *beliefs*. And that's the claim the Mixed View needs.

By appealing to the absence of belief in this way we avoid both of the original objections to the Mixed View. The first worry, recall, was about the conceptually unsophisticated, and the sceptical implications of requiring them to have concepts it seems implausible to suppose they have. The present proposal gets round that problem because while it's true that you need certain concepts in order to *believe* that defeating circumstances don't obtain; it's not true that you need those concepts in order to *lack the belief* that they do. Indeed, you can lack the belief precisely *because* you lack the concepts you'd need in order to possess it. So the present proposal doesn't have

⁹That is the sort of move one finds in Tyler Burge: the basic idea is that one couldn't be 'reliable' in ignoring 'misleading evidence' (on some suitable reading of reliable). I'm happy if that move can be made to work: it certainly makes things easier. I'm just sceptical that it can be made to work. See Burge (1993).

¹⁰Evans (1982).

the sceptical consequences that the Mixed View seems to have for the conceptually unsophisticated.

The second worry concerned the regressive consequences of the Mixed View. That View claims that part of what justifies you in any given perceptual belief is another belief. But it's natural to think that where what justifies you in a given belief includes another *belief*, we can always ask and expect an answer to the question of what in turns justifies one in that belief. If so, then the Mixed View threatens to give rise to a certain kind of epistemic regress: a regress lots of people think would be vicious. It's therefore an advantage of the present proposal that it also avoids that consequence. It avoids it because one doesn't need to *believe* that defeating circumstances don't obtain. One merely needs to lack a belief to the contrary. It's the *absence* of a certain belief that is doing the epistemic work. And while we clearly can ask of a given *belief* - what justifies it? - the same doesn't seem to be true of the *absence* of a belief.

This proposal is along the right lines but it's not quite right as it stands. To see why not let's go back to the case of the subject who has excellent reason to believe she's subject to a machine capable of causing perfect hallucinations of oranges. In those circumstances, most subjects will no longer believe that there's an orange before them when that's how things looks - they'll refraining from forming such beliefs in the first place. But let's suppose that's not so; let's imagine a subject who, when presented with the evidence, just refuses to believe that anything is amiss and simply persists in her unqualified perceptual beliefs. And suppose this subject doesn't have any more reason to discount the misleading evidence than the rest of us; she's just more stubborn. Now this subject certainly *fails to believe* that defeating circumstances obtain; she *lacks the belief* the present proposal claims she has to lack. But her perceptual beliefs are surely not reasonable. They're not reasonable, you might think, precisely because it's not *reasonable* for her to fail to believe that those conditions do obtain; she fails to have a belief which a normal subject can reasonably be expected to have. And she, no less than the normal subject, is thereby deprived of perceptual knowledge.

Examples like this suggest that it's not *enough* that a subject *merely lack* the belief that defeating circumstances obtain; the fact she lacks that belief must *itself* be reasonable or justified. The *absence* of beliefs, it now seems, no less than their presence, is something for which we can reasonably be held accountable. That clearly poses a problem for the present line of thought. The original idea, remember, was that, while we *could* ask what made a *belief* reasonable, we *couldn't* ask the same thing of the *absence of a belief*. That's how appealing to the absence of beliefs was meant to give us a way of terminating the regress. But if there's no asymmetry in this respect between beliefs and their absence, then we haven't addressed that problem: for we can simply rephrase our original question and ask about what makes it reasonable for her to *lack* the belief that defeating circumstances obtain? The worry is then twofold. First, we want an answer to this question. But second, the most obvious answer bring us right back to where we started. If it's other *beliefs* which must makes your lack of belief reasonable, then the detour through the absence of belief will have been just that: a detour. We'll be back with a partially belief-based story - and that was just what we wanted to avoid.

What should we say in response to this challenge, given that we want to avoid appealing to beliefs? The simplest, and I think the best response to this worry is surely

just to deny that what justifies a subject's *failure to believe* something need be her other beliefs. On reflection, it's clear that's an assumption we've got no reason to accept. For instance, I don't presently believe that 'there's a pink elephant standing outside the door'. And it's surely reasonable for me *not* to believe this. But what *makes it* reasonable for me not to believe this *needn't* be some other belief of mine, or my having *positive grounds* for believing that things *aren't* so. Intuitively, what makes it reasonable for me *not* to believe that there's a pink elephant outside is the fact that I have *no grounds* or *no reason* to believe that there is (which is not to say I have reason to believe that it's *not* the case that there is). The *absence of certain beliefs*, in other words, can be reasonable in virtue of one's *not having any reasons for believing*.

How does that help? Well, similarly, why not say that what justifies one's failure to believe that defeating circumstances obtain is the fact one *lacks any reason* to believe that they do? That, in the ordinary case, what makes one's lack of belief reasonable is just one's lack of grounds? That offers a neat solution to our earlier problem: the problem, that is, of the person who subjects herself to what she has every reason to believe *is* a machine capable of producing perfect hallucinations but who stubbornly refuses to believe that anything is amiss. For while she certainly lacks the *belief* that defeating conditions obtain, she clearly doesn't lack *grounds* for believing that they do. In that respect, though, her situation is precisely *atypical*: *she* doesn't lack grounds for believing that defeating conditions obtain, and that's why *she* fails to acquire perceptual knowledge. *We*, on the other hand, *do* lack reasons for believing they obtain – or, at least, we do most of the time. And that's why *we*, unlike her, count as acquiring perceptual knowledge.

To briefly summarise where we've got to so far. The Mixed View claims that part of what justifies one's perceptual beliefs is one's belief that defeating circumstances don't obtain. That's what I've denied. I suggested instead that: in order for those beliefs to be justified it must be reasonable for you to fail to believe that defeating conditions obtain. So there is something on my account -over and above one's experience - which is playing an epistemic role, and is helping to justify one's perceptual beliefs. But, unlike on the Mixed View, it's not other *beliefs* which are playing that role. It's the *fact* that one lacks reason to believe those conditions obtain, because it's that fact which, in the ordinary case at least, justifies one's failure to believe.¹¹

This revised proposal clearly offers us a solution to both of the two problems I raised earlier. Unlike the Mixed View it doesn't exclude the conceptually unsophisticated from having perceptual knowledge; and, unlike the Mixed View, it *does* offer us a way of terminating the alleged epistemic regress. For what, in the ordinary case, helps justify one's perceptual beliefs is just the *fact* that one lacks reasons for believing that defeating circumstances do obtain. And unlike beliefs, about which we can always ask: and what is the justification for *them?* - we can't ask what one's justification is for *the fact* one has no reason to believe. That doesn't make sense. Of course, we *can* ask what justification one has for *believing* that one has no reason to believe those conditions obtain. That certainly makes sense. But it's a higher order question: it's a question about one's justification for *believing* that one is justified. And we needn't

¹¹Indeed, perhaps it's not even essential that one does fail to believe such conditions obtain, provided that one lacks *reason* for believing they do. That would be one way to make sense of why the sceptic (who claims to believe that such conditions *do* obtain) nonetheless counts as acquiring ordinary perceptual knowledge.

accept that in order to be justified one must have justification for believing that one is justified. Kept on one level, the regress ends; there need be no grounds for the fact there are no grounds for believing something. It can simply be the case that there aren't any. FULL STOP. Of course, in a way the picture offered here is quite unlike the picture of grounds that underlies the Regress Argument; on this picture, things can be reasonable not only in virtue of the *presence* of certain grounds, but also in virtue of their absence. So perhaps it's not an entirely straight solution to the problem that argument raises. But not all solutions are straight solutions.

That is the picture I want to defend and why I think it's worth defending. I'm going to end by outlining what I see as some of the major challenges it faces and how I think they might be met. But before I do that I want to just say something about this way of arguing for the view. My strategy has been to appeal to the *absence* of grounds as a source of justification – because I claim that's what justifies one failure to believe defeating conditions obtain. But that may raise the following worry, namely: why suppose this response is *only* available where the *absence* of belief is concerned? Why not think that certain beliefs, (including the belief that defeating circumstances don't obtain), are reasonable or justified *just* in virtue of the fact that one lacks reasons to the contrary?¹² That they are, as certain philosophers like to put it, default reasonable? Appealing to the absence of beliefs as I do only looks like a well-motivated alternative if we assume there's a significant asymmetry in this respect between beliefs and their absence. But isn't that assumption just another vestige of the bad way of thinking that underpins the Regress Argument? Isn't it precisely asymmetries like that which I earlier criticised and claimed we ought to reject? So, once we have jettisoned the conception of justification which underpins that Argument, why should we continue to make this assumption? Why not just be more radical in our rejection of that whole line of thought?

That, at least, is the worry. The short answer, I think, is that this is something the Regress gets right. I think it *is* a distinctive feature of the *absence* of beliefs that they can be reasonable or justified *simply in virtue of the fact* that one lacks reason to believe. And I think the same can't, in general, plausibly be said for beliefs; ordinarily we just *don't* tend to think that beliefs are 'justified' simply because one lacks reasons for believing otherwise. Consider, for example, my failure to believe that it's raining outside as I sit indoors with no evidence either way. Intuitively, the fact that I *lack reason* to believe it's raining does make my *lack of belief* reasonable: I should refrain from believing it is if I've got no evidence to that effect. But if that's all my reasons (or lack of reasons) point to then I should surely remain agnostic. I shouldn't believe that it's *not* the case that it's raining: my lack of evidence to the contrary does not support taking that stand either. All it supports is my taking no stand (at least until further evidence comes in).

Of course, it's often the case that when we *lack* reason to believe *p*, we also *have* reason to believe that not-*p*. I don't just lack reason to believe there's a pink elephant outside; I also have reason to believe there's not one out there—pink elephants are pretty uncommon and any elephant would struggle to get up the stairs. That can have

¹²One finds that suggestion in the works of Crispin Wright. He claims to have got it from Wittgenstein. Wright (1991)

a confusing effect here. But the fact one may *often* have reasons of both sorts and so may reasonably lack the belief that *p* and reasonably believe that not-*p*, shouldn't be allowed to obscure the underlying distinction.¹³ Intuitively, where one *only* lacks any reason to believe *p*, that doesn't *thereby* make it reasonable for one to believe that not-*p*. It merely makes it reasonable for one to lack the belief that *p*. If that's right then *merely* lacking reasons to believe that defeating circumstances *do* obtain doesn't *by itself* give one reason to believe that they *don't*. It merely makes it reasonable for one to lack the belief that they do.

I think that contrast is intuitive and I think it reflects an important difference in our attitude to beliefs and their absence. When one believes something one's taking a stand on how things are: one positively commits oneself to things being some way. When one merely lacks a certain belief, by contrast, one makes no such commitment: one's intellectually silent as it were. Insofar as beliefs do involve taking such a stand, it's natural to think that they do as such require justification - that they aren't justified until and unless one has reasons against them, but that they require *something* in the way of positive support. Indeed, that's precisely what can seem so intuitively compelling about the picture of justification that underlies the Regress: we tend to think that beliefs *should* be adopted and maintained on the basis of evidence or reasons. Where that argument goes wrong, if it does, is in supposing the same need be true of the absence of beliefs.

So I think we shouldn't be worried about the fact that the present line of thought commits us to an important asymmetry in this respect between beliefs and their absence. However, that line of thought does raise two further questions, which are more problematic and I want to end by briefly touching on them both. The first concerns an asymmetry between what makes a *belief* reasonable and what makes one's employment of a *method* reasonable. The second concerns what it is for one to *lack* any reason to believe defeating circumstances obtain.

Starting with the first question: according to the present line of thought, beliefs aren't reasonable or justified just in virtue of the fact that one lacks reason to the contrary. I said that was intuitive and in some sense reflected the fact that in believing something one's not intellectually silent: one's taking a stand on how things are. But now consider one's reliance on certain methods or procedures for belief-fixation such as sense perception. According to the present line of thought, it's not reasonable to *believe* you can rely on those methods just in virtue of the fact that one lacks reasons for thinking one can't. But it presumably is reasonable to *rely* on them in the absence of reasons for thinking that one can't; one can reasonably rely on them -that is, one can acquire justified beliefs by means of such reliance -provided only that one lacks specific reasons for thinking otherwise. So, in this sense, the present line of thought is committed to thinking that there is a fundamental asymmetry between *holding a belief* and relying upon a certain method or *procedure for forming beliefs*.

Can that asymmetry be explained, and if so, how? Earlier I suggested that one might reasonably lack a given belief in virtue of the fact that one lacks reason to be-

¹³I think it may be equally true in the perceptual case that one often has positive reason for believing defeating conditions do not obtain: their obtaining is pretty unusual after all. What I'm sceptical of is whether those reasons are independent of one's justification for one's ordinary perceptual beliefs. If not, then it won't be of any help to a proponent of the Mixed View.

lieve, and I linked with this the idea that in lacking a belief (unlike in believing) one is ‘intellectually silent’. But it’s hard to see how the same could be said for one’s reliance upon a given method or procedure. In what sense is one intellectually ‘silent’, after all, when one positively relies upon a method?

That’s one challenge facing the present line of thought. I think it can be met. And I think meeting it involves explaining the different *roles* that beliefs and methods play in one’s epistemic life. I can’t avoid the suspicion that the following is somehow relevant, though spelling it out would be more difficult. One *can* reasonably be expected to refrain from holding beliefs in the absence of reasons to the contrary. In the case of beliefs, it *is* reasonable to demand more – to demand that one actually have reason for believing they’re true, not just that one lack reasons for thinking they’re false. Not so in the case of methods—at least methods like ‘the use of one’s senses’.¹⁴ One’s senses may, of course, lead one astray. Nonetheless, avoiding relying on them until and unless one has reason for thinking that one can is not a feasible epistemic policy. Indeed, if for *every* method one relies upon, we require that one first have reason for believing that one can, and assume that acquiring such reasons itself depends upon relying on a method, then it is not any kind of epistemic policy at all. Epistemic inquiry must, after all, begin somewhere. This line of thought needs developing, but I think it’s along the right lines.

The second challenge is to spell out what it is for one to have or to lack reason to believe that circumstances are such that one can’t rely upon one’s senses. One has perceptual knowledge *only* where one lacks such reason on my account. But a sceptic may insist that one *never* lacks such reason. The mere fact one senses can and sometimes do lead one astray is just such a reason. That reason is present in every case. So one’s justification is always defeated; in the absence of further reasons, one is always deprived of perceptual knowledge. Clearly, the right response here is to question the sceptic’s opening assumption that one always has *reason* for believing one can’t rely upon one’s senses.¹⁵ In particular, we should question the assumption that the mere fact one senses *can* deceive one, is *always* a reason for thinking that they can’t *now* reasonably be relied upon. We don’t, in general, think that the mere fact that we have reason for supposing that something is conceivable means we have any reason for thinking that it is actually the case. If not, then, for all that’s been said, we are free to think that perceptual knowledge may co-exist with one’s open acknowledgement that it’s conceivable circumstances could have been such that one would have lacked such knowledge.

So to conclude: I’ve argued that we ought to allow that a subject may lack perceptual knowledge, which her senses would otherwise make available to her, where certain

¹⁴Does what I have said apply to all methods or only some suitably privileged subset (such as those which are metaphysically necessary to any possible being, as Burge would have it?). I mean to stay neutral on that question at this stage. But I’m tempted to think it applies to them all. Hence, that if methods like ‘clairvoyance’ don’t count as ways of acquiring knowledge it’s not because one doesn’t have positive reasons for believing that one can rely upon them. It’s for some other reason e.g. perhaps they aren’t reliable, or perhaps they don’t furnish one with ‘evidence’ in the old fashioned sense.

¹⁵Perhaps we can’t spell out what it is to have or to lack reason in a general and systemic way but only by reference to examples - our willingness to count them as cases in which one is or isn’t justified thereby fixing what it is for something to count as a ‘defeater’. See ? for a defence of such epistemological modesty and (somewhat differently) Harman, G. *ibid.*

kinds of defeating circumstances prevail. I claimed that we could make sense of that without supposing that beliefs are always playing a justificatory role. It's true that one's perceptual experiences don't justify one's perceptual beliefs all by themselves - it also has to be reasonable for one to lack the belief that defeating circumstances obtain. But nothing further need be required in the way of beliefs. One's lack of belief can be reasonable because one lacks any reason to believe - and the fact one lacks such a reason is not itself a belief. It's a fact. Indeed, on closer reflection, this should have been the moral we initially drew from reflection on cases of defeat. For there what prevents one from having knowledge isn't fundamentally that one *believes* that things are amiss, but that things are amiss or one, at least, has *reason* to believe they are. If so, then what's important in the ordinary case is surely just that things aren't so, and that one lacks reason to believe they are. If that's right, then the Argument from Defeasibility doesn't give us any reason to abandon idea that perceptual justification may be wholly non-belief-based. And that's an idea which I hope to have shown isn't without its own appeal.

References

- Ayers, M. R.: 1991, *Locke: vol. 1: Epistemology*, Routledge, London. 27
- Burge, T.: 1993, Content preservation, *Philosophical Review* **102**, 457–488. 30
- Davidson, D.: 1990, A coherence theory of truth and knowledge, in A. R. Malachowski (ed.), *Reading Rorty. Critical Responses to Philosophy and the Mirror of nature (and Beyond)*, Blackwell, Oxford, pp. 120–138. 25
- Evans, G.: 1982, *The Varieties of Reference*, Oxford University Press, Oxford. 30
- Goldman, A.: 1976, Discrimination in perception, *Journal of Philosophy* **73**, 771–791. 26
- Harman, G.: 1973, *Thought*, Princeton University Press, Princeton. 28
- Martin, M. G. F.: 2001, Epistemic openness and perceptual defeasibility, *Philosophy and Phenomenological Research* **LXIII**. 26
- Wright, C.: 1991, Scepticism and Dreaming: Imploding the Demon, *Mind* **100**. 33

Getting Back to Nature

Leonard Kahn*

OXFORD UNIVERSITY

EVALUATIVE PROPERTIES ARE often thought to be problematic. E.g., they are said to be “unverifiable,” to “play no role in observation,” to be “queer,” to suffer from a “location problem,” and indeed “not [to be] in this world at all.”¹ These charges differ in detail, and there are a variety of ways of responding to them.² However, one general way of vindicating evaluative properties is to show that they are either natural properties or suitably explainable in terms of other natural properties. In this paper, I argue that such a naturalistic vindication of evaluative properties can go farther than many have thought. In section 1, I lay the groundwork for my discussion by distinguishing between thick and thin evaluative properties and by clarifying the concept of natural properties. In sections 2, 3, and 4, I defend the claim that thick evaluative properties are natural properties, while I leave a discussion of thin evaluative properties for another time.

1 Natural Properties through Thick and Thin

Let me begin with Bernard Williams’ distinction between thick and thin evaluative properties. According to Williams, the thick is “world-guided”; it is sensitive to features of the world, in a way that the thin is not. Consider Williams’ examples of thick evaluative properties: treachery, brutality, and courage.³ We might say that an act that is courageous if it has some determinate relationship with the agent’s feelings of fear. To take one case, we might think that Alice acts courageously solely in virtue of the fact that she acts while feeling fear, though neither too much nor too little.⁴ In contrast, if we think that Alice acts courageously solely in virtue of the fact that she acts while

*leonardkahn@gmail.com

¹See Ayer (1936/1946; Chapter 6), Harman (1977; Chapter 1), Mackie (1977; Chapters 1 and 5), Jackson (1998; Chapter 1) and Blackburn (1988) respectively.

²Useful overviews include Darwall et al. (1992) and Smith (1994; especially Chapter 2).

³Williams (1985; p. 129). Also see Quinn (1987), Scheffler (1987), and McNaughton and Rawling (2000).

⁴E.g., Aristotle in his *Nicomachean Ethics*, 1115a-1118a. Of course, the distinction between the thick and the thin does not depend on the particular details of Aristotle’s account of courage, though they are useful by way of illustration.

sipping 20-year-old scotch or while feeling a deep yearning for tuba lessons, then we have not understood what courage is. The thick maps on to certain characteristic features of the world and not on to others; in particular, it maps onto characters and acts undertaken in light of a certain amount of fear. In contrast, the thin does not appear to be world guided in quite this way. A strawberry is good in virtue of being red and juicy; a boil is bad in virtue of the very same facts; a violin concerto can be good or bad even though it is not even capable of being red or juicy.⁵

Now let me turn to the notion of a natural property. Michael Smith gives voice to a common point of view when he writes that natural properties are those “which are the subject matter of a natural or social science,” adding that “the social sciences include psychology, and will therefore encompass facts about human wants, needs, and well-being quite generally.”⁶ Though this characterization is helpful as a rough approximation, it threatens to do little more than push the question back to what counts as a natural or social science and to why this distinction matters.⁷

So let me borrow an expression from Geoff Sayre-McCord and say that a natural property is one that has “explanatory potency” while a non-natural property is one that does not. That is to say,

A property P is natural just in case there is some observational fact F such that the best explanation of F invokes P but would not be the best explanation of F otherwise.⁸

By adopting this way of distinguishing natural from non-natural properties, we avoid the further questions raised by the standard way of distinguishing them. To begin with, we avoid difficult questions about what counts as a natural or social science. While distinguishing between sciences and non-sciences is a worthwhile task, all we need be concerned with here are best explanations of observational regularities. When trying to diffuse skepticism about evaluative properties, we need not be concerned with, e.g., whether or not history counts as a *real* science, etc.⁹

⁵Two minor qualifications are in order. First, Williams did not use the term “thin” in opposition to “thick,” though it is hard to imagine what else would serve as a contrast class. At any rate, the term “thin” has stuck in discussions of Williams’ ideas, and I shall use it here. Second, Williams speaks of thick *concepts* rather than thick *properties*. Obviously, the precise nature of the relationship between concepts and properties is beyond the scope of this paper. I shall simply assume that the relationship is close enough to allow talk of thick properties as well as thick concepts.

⁶Smith (1994; p. 17). Since I follow Armstrong (1980) on the relationship between properties and states of affairs, Smith’s distinction between the two in the rest of this passage is of no importance here. See also, e.g., Moore (2003), Brink (1989; p. 22), Rachels (2000; pp. 74–91), Smith (2000b; p. 23), and Copp (2003).

⁷For further criticisms of this way of distinguishing between the natural and the non-natural, see Lemos (1994; pp. 120–124), Griffin (1996; pp. 37–51) and the exchange between Smith (2000a) and Griffin (2000).

⁸Sayre-McCord himself does not discuss explanatory potency as a test for whether a property is natural, though it seems a fine fit to me. See Sayre-McCord (1988a; pp. 261–263). Ridge (2003) appears to accept this understanding of natural and non-natural properties. It is also close to the empirical characterization of natural properties suggested by Copp (2003) and adopted by Stratton-Lake and Hooker *Forthcoming* among others. One qualification is in order: The foregoing test applies to first-order properties only. Some higher-order properties which do not themselves have explanatory potency can be vindicated by being explained in terms of other first-order properties that themselves have explanatory potency, though that is the subject of another paper.

⁹It is reasonable to suppose that properties that have explanatory potency are also *among* the properties

Given this (admittedly somewhat sketchy) background, we have a credible story about why the distinction between natural and non-natural properties matters. Non-natural properties can be seen to be metaphysically suspect because one does not need to posit their existence in order to explain anything we see, hear, taste, etc. In principle, it would be possible to give a complete account of every event that ever has occurred, is occurring, or will occur without once invoking non-natural properties. So positing these properties would be, at best, superfluous and, at worst, mystifying. Moreover, non-natural properties can be seen to be epistemologically suspect because we lack any plausible account of how we come to know when these properties are instantiated. Since these properties do not have a role in our best explanations of observational regularities, we have no reliable guarantee that we can track them. Sceptics seem justified in being suspicious of non-natural properties, and to the extent that we can show evaluative properties are not non-natural, we significantly undermine the grounds for skepticism about them.¹⁰

Yet once we characterize natural properties in this way it is easy to think of thick evaluative properties which *appear* to be natural. Consider some examples. First, the property of being oppressive: The opposition to the institution of slavery in the United States grew during the time between the Revolutionary War and the Civil War. Some argue that this fact is best explained by the fact that slavery became more oppressive during that time. That is to say, oppressiveness has explanatory potency. Again, consider the properties of being cruel. It is plausible to explain the actions of a tyrant by reference to his cruelty and the actions of an excellent ruler in terms of his being just. Yet again, consider the property of being honest and kindness. The former engenders trust, and the latter encourages friendship.¹¹ Indeed, there appears to be a strong *prima facie* case for saying that thick evaluative properties play a role in a wide variety of explanations of observational regularities. Nevertheless, this *prima facie* case has come under attack recently. In the remainder of this paper, I shall defend it from the most powerful three versions of this attack.

2 On Not Being Regular

Some maintain that there are no regularities for thick evaluative properties to explain. Take honesty; a moment ago I repeated the (apparent) platitude that honesty engenders trust. But, e.g., Brian Leiter claims that this is not so. Rather, Leiter tells us that “honesty just as often engenders not trust, but annoyance, bitterness, or alienation; people, as is well-known, do not want those around them to be *too* honest.”¹² Call this “the no-regularity criticism.”

that are part of the ontology of the natural and social sciences, so it is not necessary to reject the standard characterization of natural properties as at least a *partial* characterization. More importantly, is it necessary to rely on it in what follows.

¹⁰For further discussion of this point, see the authors discussed in footnotes 1 through 5 as well as Russell (1910), Strawson (1949), Sayre-McCord (1986), Railton (1992), Smith (1994), Hampton (1996), Miller (2003), Dancy (2006), and Sturgeon (2006).

¹¹See, of course, Sturgeon (1984; p. 245), Railton (1986; p. 10), Sayre-McCord (1988a; p. 276), and Brink (1989; p. 187)

¹²Leiter (2001; p. 95).

However, there are two reasons that the no-regularity criticism fails. To begin with, it assumes that the following form of reasoning is valid:

A certain quantity of X causes event E , and a greater quantity of X causes, not E , but F which is deeply at odds with E . It follows that there is no regularity with respect to E to be explained by X .

But this form of reasoning is *not* valid, as many everyday examples make clear. A certain amount of exercise will cause me to be healthy, but a greater amount of it will cause me to be unhealthy (as a result of excessive strain on my body). Exposure to a certain amount of a virus will make me resistant to it, but a greater quantity will cause me to succumb to it (as a result of overwhelming my immune system). Once again, a certain amount of solar radiation will nurture life on some of the planets that orbit it, but a greater quantity will destroy any such life that might be there (as a result of changing the average temperature of the planet). In short, the mere fact that different quantities of a property call forth different responses does not provide us with reason to believe that the property does not explain any regularities. On the contrary, many unproblematic natural properties do this very thing.

Perhaps advocates of the no-regularity criticism would claim that what I have said involves an uncharitable interpretation of their position; perhaps they would maintain no more than this: Honesty does not regularly engender trust because it can be combined with, say, cruelty in such a way that it engenders, say, distrust instead.

Nevertheless, this revised interpretation does not rescue the no-regularity criticism. For the regularities between *many* properties can be suspended by the presence or absence of other unproblematic properties. There is a certainly a regularity between striking a match and its lighting – even though this regularity will not manifest itself if the match is wet, if oxygen is not present, etc.. Nevertheless, even if one wishes to concede ground here, one can do so without any risk. E.g., one might claim that the thick evaluative property that plays a role in explaining observational regularities with respect to trust is not honesty but rather non-malicious honesty. This property will work just as well, if not better.¹³

There is a second reason to reject the no-regularity criticism: It often relies on a questionable interpretation of the relevant thick evaluative properties. E.g., Leiter confuses honesty with several other distinct properties – e.g., tactlessness, cruelty, and impertinence. Crudely put, one is honest to the extent that one does not lie and, more generally, to the extent that one does not deceive.¹⁴ Offering opinions indiscriminately, stupidly, maliciously, or hatefully is another matter altogether and can be done without being honest at all. The problem with a person who tells me I'm ugly and moronic is *not* that she is honest. Rather, the problem is that she is being heartless. Equally, the problem with a person who tells others that I wet the bed is not that she is honest. The problem is that she is abusing my confidence.

¹³This reply raises some interesting questions about the connections between this attempt to vindicate thick evaluative properties and the position known as moral particularism. While these questions are orthogonal to this paper, see [Dancy \(2000a\)](#) and [Dancy \(2004\)](#). Given the inclusion of evaluative “hedge phrases,” I assume that the list of hedges is not infinite. For worries along these lines that I cannot address here, see [Hooker \(2000\)](#) and [Little \(2000\)](#).

¹⁴Once again, Aristotle is useful for illustrating this point. See his *Nicomachean Ethics*, 1127a – 1128a.

3 The Impotence of Being Earnest

Let me turn to a second criticism. According to this criticism, we do not *need* to invoke properties such as honesty, cruelty, and earnestness in order to explain the relevant observational regularities. Rather we need do no more than invoke *beliefs* about these properties.¹⁵ The properties themselves are superfluous. Call this “the explanatory-impotence criticism.”

Consider a plausible candidate for explanatory impotence: There are behavioral regularities to be explained in communities where the idea of being a witch is taken seriously. But we would be strongly inclined to favor an explanation of such regularities in terms of *beliefs* about witches, rather than in terms of witches themselves. E.g., the creepy sensation that Samuel Sewell felt when a supposed witch walked by is better explained by the fact that he *believed* that this is a witch, rather than by the fact that someone really was a witch. By the same token, Leiter writes, “what people believe or perceive to be ‘just’ probably does engender allegiance, whereas the regularity collapses when we talk about real justice, which is often a threat to privileged groups.”¹⁶

Several comments are in order. First, a defender of thick evaluative properties need not deny that *some* regularities are best explained solely in terms of beliefs about a property rather than the property itself. One loses nothing by conceding that we are better off explaining behavior in terms of beliefs about witches, rather than witches themselves.

Second, a defender of this strategy need not deny that in the case of at least some evaluative properties, beliefs about these properties play some kind of mediating role in the best explanations of the relevant phenomena. Think for a moment about a phenomenon such as relative prices. E.g., imagine that the price of Battlestar Galactica action figures has gone up relative to the price of Firefly action figures. According to orthodox microeconomic theory, it is this change in relative prices that best explains why I buy more Firefly action figures than I would have if the relative prices remained the same.¹⁷ Nevertheless, the relative prices effect my behavior only to the extent that I have beliefs about them. That is, if I had different (i.e., false) beliefs about the relative prices, then I would act differently. But does it follow that the best explanation of my behavior would invoke *only beliefs* about relative prices, rather than relative prices themselves? Does it follow that economists need to reform their ontologies in light of these facts and stop talking about prices, just as we have stopped talking about witches? Clearly, not. Indeed, if we accept belief-desire psychology or something close to it, *all* explanations of human actions will involve mediating beliefs about certain properties.¹⁸ So such an invocation does not, in and of itself, have worrying implications for thick evaluative properties.

Let me consider a brief rejoinder to what I have just said. One might insist that there

¹⁵Though this criticism is usually phrased in terms of beliefs, it could just as easily include other propositional attitudes. E.g., that I hold my breath while passing a cemetery can be explained in terms of my *desire* to avoid being cursed by the dead. However, for the sake of brevity, I shall ignore this point here.

¹⁶Leiter (2001; p. 96). See also Harman (1977; Chapters 1 and 2) and Audi (1997; Chapter 5).

¹⁷E.g., Heyne (1980; p. 13ff).

¹⁸On belief-desire psychology, see Smith (1987), Kim (1996), and Dancy (2000b).

is the following important dis-analogy between, say, properties like relative prices and properties like cruelty. On the one hand, wide-spread consensus about relative prices is possible – indeed, in most cases, it is actual. If I need to have information about the relative prices of Battlestar Galactica action figures and Firefly action figures, I only have to consult an efficient market for these commodities. On the other hand, there is not a similar wide-spread consensus about cruelty, and there is no way to achieve it. So goes the rejoinder.

While there is something to this, as stated it is too strong to be plausible, and weaker statements of it are not worrisome. To begin with, the rejoinder is too strong in as much as it assumes that there is no way to achieve a wide-spread consensus about thick evaluative properties. To be sure, there appear to be two kinds of obstacles to achieving such a consensus, yet neither is insurmountable. First, there are some distorting influences that can stand in the way of consensus about such things as cruelty. E.g., if I believe that cats are automata who cannot feel pain, then it is plausible that I will think that lighting a cat on fire is not cruel. But the problem here is my false belief that cats are automata. This is a false belief that can be corrected, and when it is, it seems clear that lighting the cat on fire *is* cruel.¹⁹ Similar false beliefs about cruelty can be removed in similar ways.

Second, there are at least some deeper questions about the nature of cruelty that resist consensus. E.g., does cruelty always involve an attempt to cause pain, or does it sometimes involve not pain but rather an attempt to show disrespect for another? Questions of this sort might not be resolved solely by removing distorting influences such as those mentioned a moment ago. Nevertheless, difficult questions of this sort are hardly unique to ethical theory. I see no reason to believe that we are much less likely to reach reflective equilibrium about them than about other interesting concepts that are not evaluative.²⁰

Furthermore, the rejoinder is also too strong in as much as it assumes that there is not actual wide-spread consensus about many thick evaluative properties. Surely, almost everyone will agree that inflicting pain on another for sport is cruel. Likewise, almost everyone will agree that standing up for what one believes in the face of stern resistance is courageous. True, we do not agree about every case, and it is par for the course for us to focus on the hard ones. To take an especially vivid example, there was a great deal of public disagreement over the question of whether the 9/11 terrorists were brave. But we should not allow the hard cases to distract us from the wide-spread consensus about many thick evaluative properties. This is especially true in light of the fact that non-theological ethical theory is a relatively underdeveloped area of inquiry and has been pursued rigorously for a fairly short period of time. It is no surprise that hard cases remain.²¹

A final point remains to be made about this rejoinder. Up until now, I have said

¹⁹Putting aside worries raised in Putnam (1962), of course!

²⁰See Brink (1989; p. 197–209). One need only think about debate within the sciences over, e.g., the correct monetary models, the nature of the Cambrian explosion, rational-actor theory, and, until recently, Fermat's Last Theorem. Life is a good deal messier in the sciences than one might think. See M. Friedman (1992), Stiglitz (2003); Dawkins (1990), Gould (1990); Friedman (1996), Green and Shapiro (1996); and Singh (1998).

²¹Parfit (1984; pp. 453–454).

nothing about why, say, honesty and cruelty figure in the best explanations of some regularities while, say, witches do not, except in as much as they are represented in beliefs. A large part of the reason for this is the simple fact that the existence of properties such as honesty and cruelty is consistent with our other beliefs about the natural world while our beliefs about witches are not. It is not just that we do not need to invoke witches in order to explain the actions of the residents of 17th century Salem, Massachusetts; it is that if we did invoke such things, we would have to amend substantially our body of beliefs that form the natural and social sciences – beliefs that are at least as well supported as anything else in our cognitive economy. The situation is different with respect to thick evaluative properties. The existence of at least a great deal of thick evaluative properties is not at all at odds with physics, chemistry, economics, and the like. Indeed, return for a moment to honesty. This property is not just consistent with the social sciences, these sciences even offer explanations of why some people are honest in certain situations, how they gain such dispositions, etc. We would weaken the explanatory value of the social sciences if we banished honesty and the like from the realm of unproblematic properties. So the motivation to explain witches in terms of beliefs simply does not extend to thick evaluative properties.²²

4 Simply the Best?

Let me turn to what I shall call “the best-explanation criticism.” In order to motivate this criticism, I need to examine the claim that explanations involving thick evaluative properties are, in at least some cases, the *best* explanation of certain regularities. Earlier, I simply repeated the claim that these explanations are, in certain circumstances, best. But why should one believe this to be true?

The answer I shall defend is that these properties have pragmatic tenacity. Let us say that a property has pragmatic tenacity just in case it is invoked in a best explanation solely because such an explanation is substantially easier to use than the alternatives. Gilbert Harman offers one example of pragmatic tenacity thus understood: secondary qualities. E.g., it is sometimes useful to explain the fact that an object looks green in virtue of the fact that the object is yellow but is being viewed in blue light. Oddly, Harman denies that the same can be said for thick evaluative properties and facts.²³ Sayre-McCord, correctly in my opinion, points out that this is a mistake. If such properties as colors have pragmatic tenacity, then so too do thick evaluative properties.

Even if [...] explanations could eventually be replaced by others that appeal only to psychological, social, and physical factors, without mentioning [evaluative] facts, they would still be useful in just the way talk of

²²Of course, this line of argument does not show that *all* thick evaluative properties are consistent with well-established natural and social sciences. E.g., desert plausibly presupposes a libertarian conception of free will that fits poorly with a causal picture of the universe. Nevertheless, there are several options here. First, one can argue that desert does not really presuppose a libertarian conception of free will. Second, one can argue that it does presuppose such a conception, but that this does not come into conflict with the natural and social sciences. Third, one can deny that desert is a thick property. Fourth, one can concede all of these points and simply do without a conception of desert.

²³Harman (1977; pp. 22-23).

colors remains useful even in light of theories of light.²⁴

To return to an example I discussed earlier, it is sometimes useful to explain the fact that two people have become friends in virtue of the fact that both are honest. So it seems that thick evaluative properties are invoked in the best explanations of certain regularities and, therefore, have explanatory potency precisely because of their pragmatic tenacity.

Recently, Alexander Miller has argued that it is a mistake to allow pragmatic tenacity to play this role. For pragmatic tenacity is merely an artifact of our own epistemic limitations. But epistemic limitations should be set aside when taking stock of the furniture of the universe. Miller concludes that “explanations which invoke moral facts and properties do not count as best *for God*; thus, they do not count as best in the right sort of way for their availability to earn ontological rights for the higher-order properties in which they trade.”²⁵

However, Miller asks too much from the notion of the best explanation. It is not at all obvious what would count as a *satisfactory* explanation for an omniscient and infinite mind, much less the *best* explanation. Yet it is probable that it would have little relevance for creatures like us. E.g., suppose that an omniscient and infinite mind could understand every event in the universe in terms of nothing but physics. Would only explanations of this type count as the best explanation? Would we be forced to deny explanatory potency to the properties of chemistry, biology, geology, astronomy, economics, psychology, and the like? Of course, thick evaluative properties would also lack explanatory potency under this scheme. But who cares? If there is no more reason to be skeptical about the existence of fairness and courage than there is to be skeptical about mitosis and continental drift, then thick evaluative properties appear to be in good company. Indeed, it seems likely that what we now think of as the basement facts of physics as currently practiced are themselves the sort of thing that would not figure in the best explanation of the universe for an omniscient and infinite mind. After all, if some version of super string theory is correct, then even protons, atoms, and molecules would not figure in the best explanation of the universe.²⁶ It would seem that almost every property with which we are familiar is metaphysically suspect!

Requiring that the best explanation to be the best for an omniscient and infinite mind is a mistake. The restrictions that it imposes on our ontology amount to a kind of *reductio ad absurdum* for this proposal. It seems wiser to think of the best explanation in terms of imperfect and finite minds like our own. Given these limitations, considerations of pragmatic tenacity seem justified.

Indeed, there is something of an irony here. Those sympathetic to naturalism are often atheists or agnostics. It would be amusing if they had to postulate God’s existence

²⁴Sayre-McCord (1988a; p. 275). For my present purposes, it not necessary to distinguish between moral properties and thick evaluative properties, though see the end of this section.

²⁵Miller (2003; p. 174). My italics. Presumably, Miller means for us to imagine the closest world in which God actually exists. If not, then his proposal faces an obvious difficult: The proposal is that, for all x , if x plays a role in God’s best explanation, then it is a genuine natural property. But in worlds were God does not exist, this conditional is vacuously true since the antecedent is always false, and there are no natural properties at all! Since there is excellent reason to believe that we live in just such a world, Miller’s proposal must be viewed with caution.

²⁶Greene (2000).

in order to make out the natural/non-natural distinction—especially since God Himself plausibly falls on the non-natural side of things.²⁷ At any rate, the essential role played here by spooky beings suggests that this is not the right way to understand naturalism.

Nevertheless, there is room for concern. It is understandable to think that it is a necessary fact that those properties that actually have explanatory potency are features of reality, that those properties would continue to be natural in the relevant sense even if we were constituted somewhat differently. But imagine that in the near future our minds are enhanced as a result of biotechnological advances. So even though our minds are not infinite, they are now capable of easily grasping explanations of certain relevant regularities without the need to invoke, say, colors. For us, let us imagine, color no longer has pragmatic tenacity. As a result, it no longer figures as part of the best explanation of any regularity. Have colors suddenly become metaphysically suspect non-natural properties? It is not the case that we no longer have the distinct experience of the red of a freshly-cut rose or of the azure of the Aegean Sea. These colors are as vivid as ever. Why should the advances in our cognitive abilities make us doubt the robustness of these properties? While these questions raise interesting problems, they do not raise any *special* problems for thick evaluative properties. For the very same thought experiments can be done with the properties that are essential parts of the social, and at least some of the, natural sciences.

For the purposes of this paper only, I have accepted the usual assumption that what is relevant is the *best* explanation of empirical regularities. Perhaps this is a mistake. Perhaps we ought to think of explanatory potency as a feature of all of the necessary properties of any *good* explanation. Plausibly, requiring that these features are part of the *best* explanation goes too far and leaves us with unnecessary difficulties, as we have seen. Sadly, I do not have room to argue this point any further here.

In this paper, I have defended the claim that thick evaluative properties are natural properties from some of the better known and more plausible objections. I have not assumed that naturalism is the only way to vindicate evaluative properties from skeptical worries. I have merely assumed that to the extent that they can be so vindicated, they do not *require* any special attention. It remains to be seen whether a similar (or dissimilar!) defense of thin evaluative properties can succeed, but that is a question for another time.

References

- Armstrong, D.: 1980, *Nominalism and Realism*, Cambridge University Press, Cambridge. 38
- Audi, R.: 1997, *Moral Knowledge and Ethical Character*, Oxford University Press, Oxford. 41
- Ayer, A.: 1936/1946, *Language, Truth, and Logic*. Second Edition, Dover, New York. 37

²⁷However, Sturgeon (2003) offers an interesting argument that supernatural beings ought to be considered natural—if they exist.

- Blackburn, S.: 1988, How to be an ethical antirealist, *Midwest Studies in Philosophy* **12**, 361–375. [37](#)
- Brink, D.: 1989, *Moral Realism and the Foundations of Ethics*, Cambridge University Press, Cambridge. [38](#), [39](#), [42](#)
- Copp, D.: 2003, Why naturalism?, *Ethical Theory and Moral Practice* **6**, 179–199. [38](#)
- Dancy, J.: 2000a, A Particularist's Progress, in B. Hooker and M. Little (eds), *Moral Particularism*, Oxford University Press, Oxford. [40](#)
- Dancy, J.: 2000b, *Practical Reality*, Oxford University Press, Oxford. [41](#)
- Dancy, J.: 2004, *Ethics without Principles*, Oxford University Press, Oxford. [40](#)
- Dancy, J.: 2006, Nonnaturalism, in D. Copp (ed.), *The Oxford Handbook of Ethical Theory*, Oxford University Press, Oxford. [39](#)
- Darwall, S., Gibbard, A. and Railton, P.: 1992, Toward *Fin de siècle* Ethics: Some Trends, *Philosophical Review* **101**, 115–189. [37](#)
- Dawkins, R.: 1990, Review of *Wonderful Life* by Stephen Jay Gould, *Sunday Telegraph* . February 25. [42](#)
- Friedman, J.: 1996, *The Rational Choice Controversy: Economic Models of Politics Reconsidered*, Yale University Press, New Haven. [42](#)
- Friedman, M.: 1992, *Money Mischief: Episodes in Monetary History*, Harcourt, Orlando, Florida. [42](#)
- Gould, S. J.: 1990, *Wonderful Life: The Burgess Shale and the Nature of History*, Norton, New York. [42](#)
- Green, D. and Shapiro, I.: 1996, *The Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*, Yale University Press, New Haven. [42](#)
- Greene, B.: 2000, *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*, Norton, New York. [44](#)
- Griffin, J.: 1996, *Value Judgment*, Oxford University Press, Oxford. [38](#)
- Griffin, J.: 2000, Replies, in R. Crisp and B. Hooker (eds), *Well-Being and Morality: Essays in Honour of James Griffin*, Oxford University Press, Oxford, pp. 281–314. [38](#)
- Hampton, J.: 1996, *The Authority of Reason*, Cambridge University Press, Cambridge. [39](#)
- Harman, G.: 1977, *The Nature of Morality*, Oxford University Press, Oxford. [37](#), [41](#), [43](#)
- Heyne, P.: 1980, *The Economic Way of Thinking*, third edition edn, SRA, Chicago. [41](#)

- Hooker, B.: 2000, Moral particularism: Bad and wrong, in B. Hooker and M. Little (eds), *Moral Particularism*, Oxford University Press, Oxford, pp. 1–22. 40
- Hooker, B., Kahn, L. and Stratton-Lake, P.: Forthcoming, *Reasons and Value*, Oxford University Press?, Oxford. 38
- Jackson, F.: 1998, *From Metaphysics to Ethics*, Oxford University Press, Oxford. 37
- Kim, J.: 1996, *Philosophy of Mind*, Westview Press, Boulder, Colorado. 41
- Lleiter, B.: 2001, Moral Facts and Best Explanations, *Social Philosophy and Policy* **18**, 79–101. 39, 41
- Lemos, N.: 1994, *Intrinsic Value*, Cambridge University Press, Cambridge. 38
- Little, M.: 2000, Moral generalities revisited, in B. Hooker and M. Little (eds), *Moral Particularism*, Oxford University Press, Oxford, pp. 276–304. 40
- Mackie, J.: 1977, *Ethics: Inventing Right and Wrong*, Penguin, London. 37
- McNaughton, D. and Rawling, P.: 2000, Unprincipled ethics, in B. Hooker and M. Little (eds), *Moral Particularism*, Oxford University Press, Oxford, pp. 256–275. 37
- Miller, A.: 2003, *An Introduction to Contemporary Meta-Ethics*, Polity Press, New York. 39, 44
- Moore, G.: 1903, *Principia Ethica*, Cambridge University Press, Cambridge. Revised Edition, Thomas Baldwin (ed.) 1993. 38
- Parfit, D.: 1984, *Reasons and Persons*, Oxford University Press, Oxford. 42
- Putnam, H.: 1962, The analytic and the synthetic, in H. Feigl and G. Maxwell (eds), *Scientific Explanation, Space, and Time. Minnesota Studies in the Philosophy of Science*, Vol. 3, University of Minnesota Press, Minneapolis, pp. 358–397. 42
- Quinn, W.: 1987, Reflection and The Laws of Moral Knowledge: Williams on Objectivity, *Philosophy and Public Affairs* **16**(16), 195–209. 37
- Rachels, J.: 2000, Naturalism, in H. LaFollette (ed.), *The Blackwell Guide to Ethical Theory*, Blackwell, Oxford, pp. 74–91. 38
- Railton, P.: 1986, Moral realism, *Philosophical Review* **95**, 162–207. Reprinted in Railton (2003), pp. 3–42. 39
- Railton, P.: 1992, *Philosophy and Phenomenological Research* **52**, 961–968. 39
- Railton, P.: 2003, *Facts, Values, and Norms*, Cambridge University Press, Cambridge. 47
- Ridge, M.: 2003, Moral non-naturalism, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.
URL: <http://plato.stanford.edu/archives/spr2003/entries/moral-non-naturalism/> 38

- Russell, B.: 1910, *The Elements of Ethics*, *Philosophical Essays*, Longmans, Green, London. 39
- Sayre-McCord, G.: 1986, The many moral realisms, *Southern Journal of Philosophy* **24**, Supplement, 1–22. Reprinted in Sayre-McCord (1988b), pp. 1–23. 39
- Sayre-McCord, G.: 1988a, Moral theory and explanatory impotence, *Midwest Studies in Philosophy* **12**, 433–457. Reprinted in Sayre-McCord (1988b), pp. 256–281. 38, 39, 44
- Sayre-McCord, G. (ed.): 1988b, *Essays on Moral Realism*, Ithaca, New York. 48
- Scheffler, S.: 1987, Morality through Thick and Thin, *Philosophical Review* **96**. 37
- Singh, S.: 1998, *Fermat's Enigma*, Anchor, New York. 42
- Smith, M.: 1987, The Humean theory of motivation, *Mind* **97**, 36–91. 41
- Smith, M.: 1994, *The Moral Problem*, Blackwell, Oxford. 37, 38, 39
- Smith, M.: 2000a, Does the Evaluative Supervene on the Natural?, in R. Crisp and B. Hooker (eds), *Well-Being and Morality: Essays in Honour of James Griffin*, Oxford University Press, Oxford, pp. 91–114. 38
- Smith, M.: 2000b, Moral realism, in H. LaFollette (ed.), *The Blackwell Guide to Ethical Theory*, Blackwell, Oxford, pp. 15–37. 38
- Stiglitz, J.: 2003, *Toward a New Paradigm in Monetary Economics*, Cambridge University Press, Cambridge. 42
- Strawson, P.: 1949, Ethical intuitionism, *Philosophy* **24**. 39
- Sturgeon, N.: 1984, Moral explanations, in D. Copp and D. Zimmerman (eds), *Morality, Reason, and Truth*, Rowman and Allanheld, Totowa, New Jersey, pp. 49–78. Reprinted in Sayre-McCord (1988b), pp. 229–255. 39
- Sturgeon, N.: 2003, Moore on Ethical Naturalism, *Ethics* **113**, 528–556. 45
- Sturgeon, N.: 2006, Ethical Naturalism, in D. Copp (ed.), *The Oxford Handbook of Ethical Theory*, Oxford University Press, Oxford. 39
- Williams, B.: 1985, *Ethics and the Limits of Philosophy*, Cambridge University Press, Cambridge. 37

Imagery, Evidence, and the Mind-Body Problem

Peter Langland-Hassan*

CUNY GRADUATE CENTER

1

KRIPKE (1980) FAMOUSLY SHOWED that certain identity statements that appear contingent (e.g. ‘water is H₂O’) may in fact be necessarily true. Much of the persuasive power of his argument derives from his ability to explain why such identities initially seem contingent. While it is impossible that water could have turned out not to be H₂O (given it is H₂O), it *is* possible, suggests Kripke, for one to be “qualitatively in the same epistemic situation” that one normally is in when one encounters water, yet it not be H₂O one is encountering (1980; p. 142). But to imagine this is not to imagine a situation where water is not H₂O, but rather is to imagine a non-H₂O substance that has some of water’s contingent properties (e.g., its looks, taste, and location). When explaining away the misconception that a certain wood table could have been made of ice, Kripke puts this point in terms of the possibility of one’s having identical *sensory evidence* in differing scenarios:

What, then, does the intuition that the table might have turned out to have been made of ice or of anything else... amount to? I think that it means simply that ... I could have the same sensory evidence that I in fact have, about *a table* which was made of ice (1980; p. 142).

The problem Kripke notoriously attributes to mind-body identity theories is that this paradigm for explaining away apparent contingency does not work in the case of mind-body identities. There is no analogous situation where one could, for instance, have the normal sensory evidence one has that one is in pain (i.e. feeling as though one is in pain), while in fact not being in pain. Thus, when it seems *prima facie* clear to one that a pain could possibly occur in the absence of its corresponding brain state, there is no story to be told about akin to the case of water and H₂O that would explain how one was not imagining what one thought one was.

*peterlhmail@yahoo.com

I challenge that conclusion in this paper, arguing that once we gain a finer understanding of Kripke's own paradigm, we see that it too can be extended to explain away the apparent contingency of mind-body identities. My dialectical route will be somewhat indirect. I begin by critiquing a promising alternate strategy for explaining away the apparent contingency of mind-body identities, first suggested by Nagel (1974) and revived by Hill (1997) and Hill and McLaughlin (1999). I then argue that, properly modified, this approach becomes a viable way of understanding Kripke's own paradigm, with the crucial difference that it allows the apparent contingency of mind-body identities to be explained away in the same manner as those pertaining to Kripke's other *a posteriori* necessities.¹

2

In a footnote to his famous "bat" paper, Nagel (1974) makes an interesting suggestion for how we might account for the apparent contingency of mind-body identities. The sense of a contingent relationship, suggests Nagel, results from the opposition of images in two fundamentally different forms of imagination, the "perceptual" and the "sympathetic". "To imagine something perceptually," Nagel tells us, "we put ourselves in a conscious state resembling the state we would be in if we perceived that thing. To imagine something sympathetically, we put ourselves in a conscious state resembling the thing itself. (This method can only be used to imagine mental events and states—our own or another's)" (1974; fn. 11).

According to Nagel, when an experience of pain is imagined occurring in the absence of the brain state to which it is (necessarily) identical, we are splicing together disparate kinds of images. We are using the sympathetic imagination to put ourselves in a state resembling pain—imagining what it's like to be *in* pain—while using the perceptual imagination to imagine perceiving a lack of the relevant brain activity (i.e., imagining seeing that certain neurons are missing or inactive). Something analogous is happening when we imagine the relevant brain state (call it "C-fiber firing") in the absence of any pain experience: we perceptually imagine or "see" the C-fibers firing while simultaneously using the sympathetic imagination to imagine being in a mental state that it does not feel like pain to be in. "Because of the independence of the disparate types of imagination" suggests Nagel, "the relation between [a mental feature and its realizing brain state] will appear contingent even if it is necessary," (1974; fn. 11). The central idea here is that we experience no cognitive discomfort in imagining a pain in the absence of its realizing brain state precisely because we imagine pains and their realizers in such different ways.

Taking Nagel's hypothesis to lend *prima facie* support for distrusting certain appearances of contingency, Hill (1997) attempts to extend the account so that it not only

¹Two alternate popular responses to Kripke's anti-physicalist argument are: 1) pains (and mental states in general) can occur non-consciously, so the phenomenal quality of a pain is not an essential property of pain; thus, imagining feeling pain is not necessarily imagining a pain, and 2) imaginability is *never* a reliable guide to possibility, so we need not concern ourselves with the imaginability of disembodied pains. For reasons I cannot go into here, I believe that both strategies lead down blind alleys. For the purposes of this paper, I will assume, with Kripke, that the property in virtue of which we fix the reference of 'pain' is an essential property of pain.

explains the appearances but explains them *away*. To explain away, one must “provide evidence which calls the reliability of the relevant mechanisms into question” (1997; p. 70). Hill’s concern to explain away is well-founded, for one might grant that imagining mind/brain identities involves disparate forms of imagination, yet wonder why the resultant appearance of contingency should be considered especially unreliable. The answer cannot be that we are antecedently committed to mind-brain identities, since the unreliability of perceptual/sympathetic image-splicing is supposed to provide *independent* warrant for accepting such identities, thereby rendering them less mysterious.²

Hill’s strategy for showing the splicing process to be unreliable is to *assimilate* the cognitive mechanisms it involves to the same class as those that lead us astray in the less controversial case of natural kind/theoretical kind identities. In both kinds of cases, Hill claims, our intuitions of separability are unreliable because they result from our imagining certain *commonsense* phenomena in the absence of the *theoretical* phenomena to which they are identical (1997; p. 71). Just as ‘pain’ picks out a commonsense phenomenon and ‘C-fiber firing’ a theoretical one, so does ‘water’ name a common sense kind and ‘H₂O’ a theoretical one. Our epistemic relations to commonsense and theoretical phenomena are so different that identities involving them will inevitably have an air of contingency.

Though well-motivated, this assimilative approach fails on several counts. First, it seems clear that the imaginability of pain without C-fiber firing does not arise out of the mere fact that pain is paired with a theoretical phenomenon. The same sense of contingency, fueling dualisms from Descartes forward, is present even if we substitute a commonsense (physical) phenomenon for C-fiber firing.³ Suppose we had reason to believe that pain is necessarily identical to the entire brain (or to a beach ball, or any other commonsense physical object). We could just as easily sympathetically imagine the pain occurring in the complete absence of that (perceptually imagined) physical object.

Moreover, if a theoretical kind/non-theoretical kind contrast could do the explanatory work Hill wishes, his appeal to the Nagelian ideas concerning disparate forms of imagination would not have been necessary in the first place. He could simply have put his point in terms of a contrast between things named by theoretical and commonsense terms.

Also, if Hill is to apply his own paradigm to the case of natural kind/theoretical kind identities, he will have to give some *other* account of why necessary *a posteriori* identity statements *not* involving theoretical terms (e.g. “Hesperus is Phosphorous”, “Superman is Clark Kent”) can also appear contingent. A unified account of the mechanisms responsible for the apparent contingency of all necessary *a posteriori* identity statements would, if possible, be preferable.

²Put another way, the identity theorist must not beg the question against those who allow for mind-brain causal interaction yet, trusting their Kripkean conceivability intuitions, deny mind-brain identities.

³Given the little that most philosophers know of neurophysiology, it is even plausible to think that in philosophical contexts ‘C-fiber firing’ plays the role of a more or less commonsense physical kind, standing for, roughly, “any given gray and squishy brain event.” This is not to ridicule the philosopher’s use of ‘C-fiber-firing’, but is just to suggest that his epistemic relationship to the term may not be a deeply theoretical one.

3

Consider again Kripke’s method for explaining away apparent contingency: the possibility of someone’s being in a qualitatively identical epistemic or evidential situation—having the same sensory evidence—when encountering both water and some non-water substance is mistaken for the possibility that water could have turned out not to be H₂O.

My suggestion will be that we can understand the envisaging of qualitatively identical epistemic and evidential situations in various cases where the external world differs as essentially involving an opposing of images in the sympathetic and perceptual imaginations. If such a view is correct, it will allow the desired assimilation, showing the apparent contingencies of natural kind/theoretical kind identities and mind/brain identities to share the same roots.

I will use ‘imagery’ to name that class of mental states that are typically featured in a remembering (or anticipating, or supposing) of what various perceptual and sensory experiences were (or will, or would be) like. Images account for the difference between merely thinking that, for instance, you went to the store, and imagining yourself having done so. Images need not be picture-like, but they will bear a close relation to perception, and to the qualitative aspects of experience, being invariably keyed to some sense modality or other. Thus, to simply think that there is bread baking in the kitchen without drawing on the resources of some sense modality is not to *imagine* that bread is baking in the kitchen. That said, if one defines “thinking” broadly, one can consider imagery as a *kind* of thinking—here the relevant distinction would fall between imagistic thinking and non-imagistic thinking. I intend for this characterization of imagery to be neutral as regards the debate between pictorialists and descriptionalists, which concerns the underlying nature of the representations featured in imagistic thinking, not the very existence of imagery, or its intuitive features. So long as one grants that there is a difference between thinking and imagining, images—be they “pictorial” or “descriptive” in nature—are what account for this difference.

On reflection, we can see that the difference between the perceptual and sympathetic imaginations is somewhat subtle. One and the same mental image can, it seems, serve double duty, being incorporated in both the perceptual and sympathetic imaginations (going forward, the “P-imagination” and “S-imagination”). To see this, we can start by imagining brain state X —a state that realizes an experience as of staring at a red wall—in the two different ways, noting the images the two imaginings make use of (the ‘ \gg ’ signs below signify that the thing on the left *essentially involves* or *makes use of* the thing on the right)⁴:

- (1) S-imagination of $X \gg$ a visual image of a red wall.
- (2) P-imagination of $X \gg$ a visual image of activity in a certain set of neurons, R .
[or, if one is skeptical that we can visually perceive a certain brain state, an

⁴Keep in mind that we are simply *assuming* here that brain state X is identical with a certain visual experience, for the purposes of seeing how the sympathetic/perceptual imagination distinction would account for the apparent contingency of such a relation, were it to hold. So far as independent reasons go for accepting the identity, I think the best (current ones) will be causal: postulating identities between mental states and brain states is a very good (but by no means the only) way of explaining why brain events are closely correlated with mental events. See [Papineau \(1998\)](#) for what I take to be cogent developments of the causal argument for mind-brain identity.

image of what it's like to see certain test equipment revealing the relevant activity].

It is the “disparity” in these two types of imagination to which Nagel attributes the apparent contingency of mental/physical identities. Now imagine a red wall W and, in particular, what it would be like to see W .

(3) P-imagination of $W \ggg$ a visual image of a red wall.

Lastly, sympathetically imagine a certain brain state Y that is the brain state of someone who is looking at brain state X above (that is, they are seeing the brain of the person—a neurological patient, let's suppose—who is in X).

(4) S-imagination of $Y \ggg$ a visual image of activity in a certain set of neurons, R .

Notice that (1) & (3) and (2) & (4) incorporate the same mental image, even though each pair couples an instance of the perceptual imagination with an instance of the sympathetic imagination. The conclusion we should draw from this is simple: the difference between the sympathetic and perceptual imaginations does not reside in the intrinsic nature of the images they incorporate. Rather, it arises out the different ways in which a particular image may be used in one's mental economy.

Before I go on to explain these different uses, I should stress the following: the content (or functional role) of an act of imagination (perceptual, sympathetic, or otherwise) need not (perhaps *must* not) be fixed entirely by the nature of the image featured in the act of imagination. Two acts of imagination can easily incorporate the same image without themselves being the same (as a rough analogy, one and the same string of English letters can have different meanings). This is one of the upshots of the reasoning above, but is evident from consideration of more mundane cases as well—for one can use the same mental image to imagine one's mother *and* to imagine someone who looks exactly like one's mother, without these two acts of imagination amounting to the same thing.⁵

Given that there is *more* to an act of imagination than the image it incorporates, let me now try to say what it is that distinguishes acts of the sympathetic imagination from acts of the perceptual imagination (aside from the fact that only “mental” states can be sympathetically imagined, itself a not particularly illuminating stipulation). When we use an image to represent our *last line of evidence* (or, as I'll also call it, “transparent evidence”) for an external state of affairs, we are conceptualizing that image sympathetically. When we use the image to represent properties inherent in an external state of affairs itself (or to represent “non-transparent evidence”), we are conceptualizing the image perceptually.

The idea of a “last line of evidence” or “transparent evidence” bears a close relation to the view that perception is in some sense “transparent”. When we can use one and the

⁵Thus this account does not run afoul of Wittgenstein's (1953) famous worry: “I see a picture; it represents an old man walking up a steep path leaning on a stick.—How? Might it not have looked just the same if he had been sliding downhill in that position?” (*ff.* 137, fn. 2(b)) Does allowing that there is more to an act of the imagination than what is given in its attendant imagery conflict with pictorialism? I do not think so; the pictorialist need only allow that the (pictoral) image featured in a given act of the imagination does not itself exhaust the content of that act of imagination.

same image both to represent our perceptual evidence for a thing and to represent the thing we would normally take the imagined evidence to be evidence *for*, the evidence we are imagining is our last line of evidence, for there is no more basic evidence we can call on to justify the perceptually formed belief. A moment ago I was looking at a pen. When I imagine that pen now, I form a certain visual image. And when I imagine what my evidence was that there was a pen of that sort in front of me, I conjure the same image, but use it to represent my *evidence* for the pen and not the pen itself. In so doing, I have reached what I am calling my “last line of evidence” for the pen, and, by extension, have engaged in an act of the sympathetic imagination.

One can, of course, perceptually imagine an everyday object as a way of imagining evidence for some state of affairs. For instance, I might perceptually imagine an unmade bed, and rightly consider what I am imagining to be a representation of evidence that someone was sleeping in a certain bed not long ago. But here we must use a *different* image (of a person sleeping in the bed) to represent the thing we would take the unmade bed to be evidence for.⁶

So, whereas Nagel simply stipulates that the sympathetic imagination pertains only to *mental* states, we now have an independent, epistemic characterization of it: sympathetically imagined states are *evidential* states, the imagination of which involves the same image as the imagination of the states of affairs such evidence is taken to be evidence *for*.

4

We can now cast Kripke’s paradigm for explaining away the apparent contingency of natural kind/theoretical kind identities in terms of the sympathetic and perceptual imagination distinction. When we misconceive the possibility that water could have turned out not to be H₂O, what we are really doing is sympathetically imagining a typical last line of evidence for water, and then “splicing” this (sympathetic) image with a perceptual image of some non-H₂O substance that we are interacting with while having this perceptual evidence.⁷ This is simply a more fine-grained way of understanding Kripke’s suggestion that one “could, *qualitatively* speaking, be in the same epistemic situation” as someone whose experiences are being caused by a substance of a different nature (1980; p. 150). It is our somewhat covert tendency to engage in this kind of image-splicing that makes ‘water is H₂O’ appear contingent.⁸

⁶ I am advancing a necessary, not a sufficient, condition for something being an act of the sympathetic imagination. Thus, I have not *reduced* the notion of a sympathetically imagined state to a “last line of evidence”, but have merely provided a partial, epistemically grounded, *characterization* of it.

⁷ Again, it is not crucial that the image of XYZ used in the perceptual imagination be any different, *qua* image, than that used to perceptually imagine H₂O—the important difference here too will be in how that image is *used* in one’s mental economy. It is enough that H₂O and XYZ are conceptually “tagged” as different substances. Yet it bears noting that the “disparity” that Nagel originally noted between the sympathetic and perceptual imaginations has shifted from being a pictorial one to being a largely functional one.

⁸ Why doesn’t a similar splicing process make “water = water” seem contingent? Well, it *would* make “water = water” seem contingent if one took one’s justification for accepting the identity to be grounded in empirical evidence. If, on the other hand, one feels justified in accepting the identity claim independent of any empirical evidence, then the splicing process will not be brought to bear on the statement to begin with.

Turning to the mind-body case, a sensation of heat is evidence for the presence of heat itself. And the image we use in an imagination of our typical sensory evidence for heat is the same as we might use to imagine heat itself. Thus imagining one's (typical) sensory evidence for heat qualifies as a case of the sympathetic imagination. Of course, a sensation of heat can also make one aware that one is in a certain brain state, couched in a neuroscientific vocabulary, if one knows some brain science (or even if one doesn't, but just thinks, "there must be something happening in my brain expressible in a neuroscientific vocabulary that accounts for this sensation. . . I name this neuroscientifically-specified state of affairs '*N*'"). In this situation, a heat sensation can serve as an intermediary that makes us aware of *itself* under another description. That is, imagining a heat sensation can be a way of imagining evidence that brain activity of a certain scientifically-specified kind is occurring.

This is where the important similarity with the heat/molecular motion and water/H₂O cases arises. In all these cases, the illusion of contingency arises from the fact that we believe we could have the same type of evidence we in fact have for a perceptually imagined object or kind—where "evidence" is always a sympathetically imagined state, the sort that can be a "last line of evidence"—yet it be some other perceptually imagined object or kind that is causing or realizing that evidence.⁹ In short, as epistemic agents with restricted forms of access to features of our environment, we are committed to the idea that states of affairs in that environment are underdetermined by the evidence we have for them; we could have the same last lines of evidence, yet things not be as we take them to be.

Of course, if a type identity theory is true and pain is type-identical with a neurophysiological state of some kind (e.g., "C-fiber stimulation"), it is *not* possible to have the same type of evidence we normally have for that state yet have it determined by some other state. However, even in this case, it remains *rational to suppose* that our evidence does not rule out the falsity of the theory that equates pains with C-fiber firings (or, alternately, it remains reasonable to think that our evidence does not logically entail the truth of the theory that equates pain with C-fiber firing). So, even if pain *is* type-identical to C-fiber stimulation, it will feel unnatural to call it a necessary identity just because it is unnatural for us to think that it is not even *possible* to have the same evidence for the occurrence of a certain kind of state while a different theory of what it is evidence for is true.¹⁰ Thus we see how what is normally a reliable methodological principle—the idea that theories are underdetermined by our evidence for them—would, in the mind/body case, lead one to feel that a necessary identity is

Thus, we can think of it as a two-step process through which the appearance of contingency arises. First, one (rather automatically) considers whether one's justification for accepting an identity statement is grounded in empirical evidence of some kind. If the answer is No, then the splicing process does not begin (since one has essentially *stipulated* that by one token of "water" or "Bob" one is referring to the same thing as with the other token of "water" or "Bob"). If the answer is Yes, then the splicing mechanisms kick in, giving the identity an air of contingency. Thanks to Jonathan Ichikawa for pressing me on this point.

⁹Clearly, causation and realization are *not* the same relation, yet they bear a similarity that warrants the present assimilation: both causes and realizers *explain*, insofar as they help *determine*, the nature of one's evidence.

¹⁰A similar story can be told for why the truth value of Goldbach's conjecture (that every even number larger than 2 is the sum of two primes) may seem contingent, even if it is necessary (that is, why it seems as though a presumably necessary truth could come out either way).

in fact contingent. Ironically, what fuels the sense of contingency may be, at bottom, an aversion to idealism. We are not comfortable concluding that, given our current perceptual evidence, some aspect of the external world (i.e. some brain state) could not possibly be any other way than the (neuroscientific) way we take it to be.

It is sometimes suggested that Kripke's appeal to identical epistemic situations succeeds in explaining away apparent contingency just because it shows we were imagining something else that *really is possible*.¹¹ But this reading is unwarranted, since we cannot know *a priori* that it really *is* possible to be in identical epistemic situations while encountering different substances. Kripke himself makes this point, when discussing the presumably contingent (but *perhaps* necessary) relation between sensations of heat and heat itself:

Of course, it might be part of the very nature of human beings that they have a neural structure which is sensitive to heat. Therefore, this too [i.e., certain kinds of sensations being *of* heat] could turn out to be necessary if enough investigation showed it (1980; p. 133)

So, what matters in explaining away apparent contingencies is not that we *really could* have the same evidence that we normally do for the presence of a substance (e.g., water) in our environment while it is some other substance (e.g. XYZ) we are encountering, but that, given our evidence, we cannot rule the possibility out.¹² Note that the constraints on what we cannot rule out *a priori*, are much looser than the constraints on what we may rationally suppose *is* the case. The former matches up with a notion of "epistemic possibility" akin to the one Kripke allows, according to which it is correct to say gold *might* turn out not to be an element (1980; fn. 72). The latter is constrained by considerations of theoretical strength and simplicity—given our evidence, we cannot rule out the possibility that Tom Cruise is a robot, but this does not make it rational to suppose that he actually is one.

If I am right, we now have what Hill was originally searching for: an assimilation of the factors responsible for the apparent contingency of mind-body identities to the class of those responsible for the apparent contingency of natural kind/theoretical kind identities. Further, it is an assimilation that fits with Kripke's own explanation of why necessary natural kind/theoretical kind identities appear contingent. Thus we need not take the apparent contingency of mind-brain identities to weigh against their necessity.

References

Chalmers, D.: 1996, *The Conscious Mind*, Oxford University Press, Oxford. 56

¹¹ This is the leading idea of David Chalmers's (1996, 2002, 2004) deployment of the two-dimensional semantic framework for understanding Kripke's arguments.

¹² Here is the argument spelled out: P1) For all we know *a priori*, it is an essential property of humans that we have neural structures that realize sensations of heat only in presence of heat. P2) So we cannot know *a priori* that we humans really could be in the same epistemic situation we are normally in around heat but be encountering something that is not heat. P3) Regardless of how the empirical investigation of our own essential properties plays out, Kripke's appeal to identical epistemic situations still explains away apparent contingencies. C) So what matters for his paradigm of explaining away to succeed is *not* that the scenarios imagined really are possible, but only that they are not ruled out *a priori* by our evidence.

- Chalmers, D.: 2002, Does conceivability entail possibility?, in T. Gendler and J. Hawthorne (eds), *Conceivability and Possibility*, Oxford University Press, Oxford, pp. 145–200. [56](#)
- Chalmers, D.: 2004, The foundations of two-dimensional semantics, in M. García-Caprintero and J. Macià (eds), *Two-Dimensional Semantics: Foundations and Applications*, Oxford University Press, Oxford. [56](#)
- Hill, C.: 1997, Imaginability, conceivability, possibility and the mind-body problem, *Philosophical Studies* **87**, 61–85. [50](#), [51](#)
- Hill, C. and McLaughlin, B. P.: 1999, There are fewer things in reality than are dreamt of in chalmers's philosoph, *Philosophy and Phenomenological Research* **59**(2), 445–454. [50](#)
- Kripke, S.: 1980, *Naming and Necessity*, Harvard University Press, Cambridge, MA. [49](#), [54](#), [56](#)
- Nagel, T.: 1974, What is it like to be a bat?, *Philosophical Review* **4**, 435–450. [50](#)
- Papineau, D.: 1998, Mind the gap, *Noûs* **32**, 373–388. Supplement: Language, Mind and Ontology. [52](#)
- Wittgenstein, L.: 1953, *Philosophical Investigations*, Blackwell, Oxford. Edited by G.E.M. Anscombe and R. Rhees, English translation by G.E.M. Anscombe. [53](#)