Cartesian Intuitions

by

Jeff McConnell
A.B., Harvard University
(1977)

Submitted to the Department of Linguistics
and Philosophy in Partial
Fulfillment of the Requirements for the
Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology
May 1994

Signature of Author . .
Department of Linguistics and Philosophy
March 9, 1994

Certified by . . . . . .
Professor Ned Block
Thesis Supervisor

Accepted by . . . . . .
Professor George Boolos
Chair, Committee on Graduate Students (Philosophy)

Cartesian Intuitions

by

Jeff McConnell

Submitted to the Department of Linguistics
and Philosophy on March 9, 1993, in Partial Fulfillment
of the Requirements for the Degree of Doctor of Philosophy

ABSTRACT

At the core of the essay that follows is a set of intuitions that distinguish the mental and subjective from the public and objective. I call these intuitions Cartesian intuitions even though Descartes himself ignored some of them. I argue that some of them survive the best efforts of critics to explain them away. This, I contend, is the basis of the mind-body problem, which should be seen as a paradox, in which both materialist and dualist lines of argument seem conclusive. My aim is thus to clarify and to bolster what I call the neo-Cartesian half of the mind-body paradox, to show that there really is a paradox, one that remains with us, undissolved.

I use two forms of arguments to accomplish this. One is the Knowledge Argument, according to which there are certain things which you can know everything physical about but not know everything about and which, because of that, make physicalism false. The Knowledge Argument says that your having this extra knowledge depends on the existence of special properties called qualia, which are supposed to be separate from any physical or functional properties. I argue that the Knowledge Argument is what survives from Descartes's original critique of materialism. I defend the Knowledge Argument, or at least I defend the claim that it has not so far been refuted.

The other form of argument is the Absent Qualia Argument. I use it further to support the case against functionalism, the approach to the mind according to which mental states are definable by their causal relations to behaviors, perceptions, and each other. Against this position, I argue that there might be so-called absent qualia states which filled the causal roles of some of our genuine mental states but did not look or feel like anything. The functionalist response to this position is that since absent qualia states cause the same beliefs as genuine states do we could never know whether we were having real states or absent qualia replicas. I argue that there is no such skeptical problem.

At the heart of both arguments is an account of our direct reference to our own phenomenal states. Critics of the Knowledge Argument have argued that the Cartesian intuitions the argument

exploits can be explained away as a result of two distinct forms of reference: roughly, the direct reference distinctive to the mental, the descriptive reference distinctive to our neurophysiological talk. But I argue that we could not refer directly to our phenomenal states unless we did so by way of properties distinct from any neurophysiological properties. Critics of the Absent Qualia Argument can argue that our direct reference to phenomenal states requires mental processing with phenomenal aspects, processing that necessarily runs outside any causal role that might be shared with absent qualia replicas. These side-effects mean that absent qualia are impossible in us. I grant that but argue that this would not rule out the possibility of absent qualia in nonsentient creatures. They might be similar enough to have states with identical causal roles to our phenomenal states; but since they would be free of experience, there would be no chance for the side-effects that make absent-qualia states impossible to realize in us.

Thesis Supervisor: Ned Block

Title: Professor of Philosophy

Biographical Note

Jeff McConnell received his A.B. degree from Harvard College in 1977. His essay "In Defense of the Knowledge Argument" is forthcoming in *Philosophical Topics*. He has taught at Tufts University since 1988. Before that, he was a contributing writer on national security matters for *The Boston Globe* from 1985 to 1988. His work is mainly in metaphysics, moral theory and the application of ethics to public policy.

Acknowledgements

In a sense, this essay began in September 1975 when I attended a seminar in Ithaca by Sydney Shoemaker on his just-published "Functionalism and Qualia." At the time, these ideas were very foreign to me. I felt that something was wrong in what he said but I also felt that there was something very important in it. The issues were too hard for me, however, and on the several times I encountered the paper in the next ten years I continued to think that they were too hard. Nevertheless, in the summer of 1985, after abandoning an earlier effort to write a dissertation on reference, I resolved to devote myself to the Shoemaker paper and the issues it raised until I could say what I thought was wrong. What is before you is an interim report on my progress, nine years later.

Two people deserve particular thanks in helping me complete this essay. My advisor, Ned Block, has read, discussed and given written comments on every inch of not only the dissertation itself but also many earlier drafts of it. His own writings on the mind-body problem and his clear and insightful thinking about what I have written on it here have been indispensible. Steve White probably rescued this thesis, while serving as a substitute adviser at an earlier crucial stage when Ned was on sabbatical. Although disagreeing with them, he took very seriously the arguments appearing in Chapter Seven and Chapter Eight, at a time when I did not know whether I had anything at all to say on these issues. Now a colleague, he has been always willing to discuss problems I have had. I have come to believe that his own formulation of the Property Dualism Argument is fundamental to understanding the mind-body problem and it informs much of what I have written here.

Other people also deserve thanks. It is due to Noam Chomsky that I came to M.I.T. and it is due to his writings that I and many others take Descartes as seriously as we do, and I thank him for both things. Brian Crabb spent many useful hours discussing Chapter Four and Chapter Five with me. I have benefitted greatly from the continual resistance to my views shown by Dan Dennett; it is hard for me anymore to write anything on these topics without hearing a voice in my head saying, "What would Dan say?" My other committee members, Bob Stalnaker and Sylvain Bromberger, read earlier drafts of the thesis and spent many hours with me. Both gave me encouragement at times of discouragement. I particularly wish to thank Sylvain for serving as my adviser while I was working on the earlier thesis topic, and it has been a matter of much disappointment that we have not had the opportunity to work as closely together since then.

The same can be said of a number of friends who watched this work evolve over the years: Jon Church, John P. Kelly, Lisa Schur and Harvey Simon.

To Janet Chumley, I am grateful for advice and patience. She put up with a lot as a finished product emerged only in fits and starts. To Reed, I am both grateful and sorry for time apart as I have done the last work on this thesis. Daddy can now play with you in the evenings again.

Table of Contents

List of Illustrations

CHAPTER ONE SOME CARTESIAN INTUITIONS

Descartes taught us that there is something to be learned of the world just from the ways we imagine it to be. The lesson we get from him is not quite the one he set out to give but it starts there.

I. Two Cartesian Intuitions, an Explanatory Gap and the Humean View about Them

From a high position, you peer out upon a body of water that meets the sky in a long, unobstructed horizon. A magnificent sunset is slowly unfolding. A warm breeze rustles the leaves. What you see before you, what you hear behind you--these are paradigms of what we call "physical." The light of a tremendous fireball millions of miles away is refracted and dispersed by the atmosphere through the many wavelengths of the color spectrum. Energetic air molecules, moving through the trees and colliding with anything in their way, create ripples of sound waves in all directions.

This is what goes on around you. But what about what goes on <u>inside</u> you--what about <u>how</u> this looks and sounds and feels to you? How could <u>that</u> be anything physical? Looks and feels may seem not to be the right sorts of things to be called "physical." After all, can we not imagine them being pried free of physical things? We seem easily able to imagine the idea, for example, of leaving our bodies.

According to a Greek legend, Hermotimus's soul would leave his body from time to time to visit distant lands, returning to report things only an observer could have known, until one day it came back to find that his enemies had burned his body.[1] Many primitive peoples have reportedly used the idea of leaving the body to explain what happens during dreams. It may even seem to us, in certain flights of fancy, that we could trade in the physical world completely and still keep the looks and feels. This we imagine doing when we entertain the thought that our entire lives are mere dreams or when, on some accounts, we imagine ourselves going to Heaven.[2]

---

1. See Pliny the Elder, Natural History, Bk. VII, sec. 174; in ed. by H. Rackham, tr. (Cambridge, Mass.: Harvard University Press, 1942), at vol. 2, p. 623; see also Lucian's work, The Fly (or Muscae Laudatio) in A. M. Harmon, tr., Lucian, vol. 1 (Cambridge, Mass.: Harvard University Press, 1913), at p. 89.

2. For the classic account of primitive explanations of dreams as disembodied experiences, see E. B. Tylor, Primitive Culture, 4th ed. (London: John Murray, 1903), vol. 1, pp. 440-445. The philosophical notion that one's entire life could be a dream is, of course, an old one; see, for instance, Plato's Republic, Bk. V at 476. One must be careful in interpreting conceptions of disembodiment as intuitions of a nonphysical realm. The pre-Socratics, while seeing the soul as distinct from the human body, each thought it to be comprised by one of the substances, with the exception of earth, believed by the Greeks to be elements (air, water, fire); see Aristotle, De Anima, Bk. I, ch. 2. Even these vague conceptions of a "tenuous" soul, "like a wind or flame or ether, which permeated [the] more solid parts" of the body, to use Descartes's words, are materialistic enough to clash with his argument; see the Second Meditation at AT VII 26, in John Cottingham, Robert Stoothoff and Dugald Murdoch, The Philosophical Writings of Descartes (Cambridge: Cambridge University Press, 1984 and 1985), 2 vols., at vol. 2, p. 17. Throughout, I shall use the "Cottingham translation."

Let me call these intuitions of differences between the mental and the physical Cartesian intuitions. For it was Descartes who first drew attention to such intuitions and argued that merely by virtue of our having them we know that something about them is right. We have all at one time or another had intuitions about disembodiment like those he described. But I shall also call Cartesian other intuitions which Descartes paid less or no attention to but which are equally familiar.[3]

Horror films and the folklore of voodoo are filled with talk of zombies. Some probably think of them primarily as people robbed of free will, but a more extreme idea is that they are entirely missing the inner light of consciousness. Descartes imagined machines that could without consciousness imitate certain kinds of human behavior. He thought that all animals but humans were such machines and that, with none of the looks or feels of experience, they acted entirely on the basis of programming and stimulus.

On the other hand, it was also Descartes's view that internal programming and external stimuli were limited in what they could produce. Neither beasts nor machines, he thought--much less zombies--could pass what has become known as the Turing test, producing behavior indistinguishable

---

3. I take the phrase "Cartesian intuitions" and the distinguishing of these two categories of them from Saul Kripke, Naming and Necessity (Cambridge, Mass.: Harvard University Press, 1980), p. 148.

from that of normal humans. He argued that beasts and machines lack the capacity for making creative, unprogrammed responses peculiar to a res cogitans--a rational mind like each of us. If this were right, their psychologies and presumably their physiologies would also differ from those of normal people. So a still more extreme idea is that there could be beings who were zombie-like in lacking an inner light but who were similar to us in every other way--in behaviors that could pass the Turing test, in their psychologies, even in their physiologies.

Although I am labeling this notion as a second kind of Cartesian intuition, it is not a proposal found in Descartes's writings. He argued that nothing could produce the behavior characteristic of normal humans without the inner light of a conscious soul. But this was a conclusion from philosophical reasoning. Even Descartes would have allowed that, pre-reflectively, we are at least able to wonder if other people are zombies or robots or mere Cartesian beasts. "If I happen to look out my window and see men walking in the street," he muses in the Second Meditation, "... what do I really see, except hats and coats that could be covering robots?"[4] To wonder this is to entertain it as a possibility that they are. And this is to see the world as if it consists of more than just its physical aspects, as if it has an inner realm of

---

4. At AT VII 32.

consciousness beyond its physical aspects that we can at least imagine to be missing even when all its physical aspects are present.

According to both these sorts of Cartesian intuitions--those of disembodiment and those of zombiehood--we have direct access and make direct reference to a realm apart from the realm of the physical, to which we only have indirect access and make indirect reference. It is not from indirect behavioral evidence that res cogitantes exhibit that we can deduce this further realm, Descartes argued, but only from the direct evidence of the cogito--the "I think."[5]

This leaves us with an explanatory gap between our knowledge of the physical and our knowledge of the mental, one that cannot be bridged by naive common sense alone. This much is uncontroversial. I may have physical evidence about some people from their physiologies or from their behaviors and still may wonder of them what their experience is like, or whether they have any experience at all. I may have direct evidence in my own case of having experience and of what it is like to have but may still wonder what its physical basis is. How do we get from the one to the other? Our intuitions of the physical are very different from our intuitions of the mental. In fact, they seem

---

5. If it turns out that all res cogitantes are in fact conscious minds, capable of the cogito, as Descartes maintained, then this would require a further argument.

incommensurable.

But can we from such intuitions alone conclude anything about the world--about whatever these intuitions are themselves about? For some time, the prevalent position has been that we cannot. Let me call this the Humean view. Hume argued that none of our factual knowledge--none of our knowledge about the real world--is a priori. We cannot know anything about the way the world is, according to Hume, merely from the ways we think about it in an armchair, independently of our experience of it. He argued, for example, that only through experience can we know how, or if, such things as thought and physiological activity are connected. Experience provides knowledge a posteriori, and there is no other knowledge possible.[6]

Those who after Hume continue to take the Humean view

---

6. In one relevant passage, Hume writes: "For tho' there appear no manner of connexion betwist motion or thought, the case is the same with all other causes and effects.... If you pretend, therefore, to prove a priori, that such a position of bodies can never cause thought; because turn it which way you will, 'tis nothing but a position of bodies; you must by the same course of reasoning conclude, that it can never produce motion; since there is no more apparent connexion in the one case than in the other. But as this latter conclusion is contrary to evident experience, and as 'tis possible we may have a like experience in the operations of the mind, and may perceive a constant conjunction of thought and motion; you reason too hastily, when from the mere consideration of the ideas you conclude that 'tis impossible motion can ever produce thought, or a different position of parts give rise to a different passion or reflection." See A Treatise of Human Nature, Book I, pt. IV, sec. V, par. 30. The Cartesian, of course, holds not that thought is caused by motion but that it is a form of motion.

make arguments like his a starting point but supplement them with others. According to one argument for the position known as scientific materialism, for example, the best overall explanation we have for the evidence of the sciences is a materialist one. This line of thought entails that the Cartesian intuitions are illusory about the existence of a further realm beyond the physical but that we only know this a posteriori, in light of synoptically constructing the best overall explanation for the evidence of the sciences. If on some versions there remains some work for philosophers to do from their armchairs, it is not to establish materialism, which they purportedly cannot do. The task instead is, where possible, to reconcile us to our knowledge of materialism by providing ways in which the Cartesian intuitions could be illusory.[7] Philosophy, if this were

---

7. In "The Big Idea," Times Literary Supplement, July 3, 1992, p. 5, Jerry Fodor seems to express something like this view. With a slight exaggeration, he puts it this way: "... we're all materialists for much the reason that Churchill gave for being a democrat: the alternatives seem even worse. The new research project is therefore to reconcile our materialism to the psychological facts; to explain how something that is material through and through could have whatever properties minds actually do have." Philosophy, he goes on, can join hands with psychology in this research project. Such a picture as this, initially a reaction to the anti-synoptic, anti-metaphysical, hands-off approach to our conceptions of the world, scientific and otherwise, appearing in Wittgenstein's writings, found early expression in J. J. C. Smart's Philosophy and Scientific Realism (London: Routledge & Kegan Paul, 1963), ch. 1, and, inter alia, in W. V. O. Quine's books like Word and Object (Cambridge, Mass.: M.I.T. Press, 1960) and Ontological Relativity and Other Essays (New York: Columbia University Press, 1968). The seminal book-length defense of scientific materialism is that of D. M. Armstrong, A Materialist Theory

right, might make it conceivable that phenomenal properties could be reduced to the physical, or shown to be irreducible because of limitations on human powers of theory construction, or eliminated altogether as fictional. But only the best explanation based on the evidence of the sciences could provide any grounds for knowledge about the way the world actually is. Cartesian intuitions could thus never shift the burden of proof to the materialist.

II. The Neo-Cartesian View, the Mind-Body Problem and Descartes's Modal Argument for Dualism

The Cartesian rejects the Humean view. At least in the case of the human mind, the Cartesian claims that we can know something of the world a priori. For we are human minds, and because of this, we know we exist. And we know this a priori. Moreover, according to this alternative view, we have a special relation to ourselves that we do not have to the physical world, one revealed in our Cartesian intuitions about the differences between the physical world and certain aspects of us. On the basis of these intuitions, the Cartesian claims, we can rightly conclude that the world consists of two realms--the physical realm around us and the phenomenal realm within us.

The orthodox Cartesian (such as Descartes) often

---

of the Mind (London: Routledge & Kegan Paul, 1968).

supplements this account of the world and ourselves with further claims: that we have privileged access to our own phenomenal properties, and that our knowledge of them is certain; that the realm of phenomenal properties is not just different from but is separable from the physical realm and could exist without it; that our bodies are lifeless, mindless machines, that phenomenal properties are borne by souls or ghosts, and that each of us is no more or less than a ghost in a machine; that body and soul are distinct substances. You can, however, accept the weaker Cartesian claims--that the phenomenal and the physical are two distinct realms of properties (in a sense I hope to clarify) and that we know this a priori--without all the rest which the orthodox Cartesian may add. I shall call this pair of weaker Cartesian claims the neo-Cartesian view.

The neo-Cartesian does not need a sophisticated account of the a priori, one of the sort the orthodox Cartesian may try to provide. Tactically, it is wise to minimize the risk of being trapped in the spider's web of controversies about whether any of the terms commonly associated with the a priori--like self-evident, certain, incorrigible, and so forth--are true of anything at all. Roughly, we can say that knowledge is a priori if it can be gained by reason alone, without appeal to the particular facts of experience. Beyond that, I want to avoid commitment to any more substantive account of what the category of the a priori is

or includes. It is enough for the neo-Cartesian to point out some form of access to knowledge of the world separate from those public forms paradigmatic of our access to physical aspects of the world.

Part of my purpose in this essay shall be to show that the neo-Cartesian view survives the counterarguments of its critics, both past and contemporary. I shall not be claiming, however, that the neo-Cartesian view is true--that it shall forever survive the counterarguments to it. It is hard to see, given the worldview we get from the natural and the biological sciences, how any form of dualism could be true, since it is hard to see where nonphysical phenomenal properties could come from or how they could interact with the physical world. If we ignore for a moment the direct evidence each of us has which gives the neo-Cartesian view its pull, then it does seem that some form of the scientific materialism must be right and that the best explanations of the world must be materialist ones. This means that there is currently an explanatory gap in our understanding of how to put together the physical and the phenomenal, but it means more than that. Our present epistemological state is one of paradox, with our having apparently conclusive arguments supporting both reducibility and irreducibility between the physical and the phenomenal. A rational, well-informed observer can be in a fairly stable state of

reflective equilibrium[8] about the makeup of the world, yet be split between two independent, fundamentally different accounts, each of which when taken in isolation seems to be demonstrable. And we do not have even a glimmer of insight about how to dissolve the paradox.

One writer has labeled as principled agnosticism such an inability as this to decide between materialism and dualism.[9] But the above view is a special case of that. Here, the inability to decide comes not from inadequate information or incomplete reasoning but from two sets of information and reasoning each of which, when taken in isolation from the other, seems to give a conclusive verdict but one incompatible with the verdict from the other. This special case of paradox-inspired agnosticism appears not be resolvable without moving outside the bounds of common sense.

---

8. I borrow this term from John Rawls, A Theory of Justice (Cambridge, Mass.: Harvard University Press, 1971), p. 20, where it is used in a related way to refer to the state of equilibrium that may follow a process of mutual adjustment of first-order judgments and higher-order principles about justice in political structures.

9. Owen Flanagan, Consciousness Reconsidered (Cambridge, Mass.: M.I.T. Press, 1992), p. 1. Flanagan's example of a principled agnostic about the mind-body connection is Thomas Nagel; see his "What Is It Like to Be a Bat?," Philosophical Review 83 (1974), reprinted in his Mortal Questions (Cambridge: Cambridge University Press, 1979), p. 176, and his The View from Nowhere (New York: Oxford University Press, 1986), p. 47. Nagel, however, displays none of the sense of paradox I recommend here; while he holds that there is currently an "explanatory gap" in getting from the material to the phenomenal (see below), he does not recommend any form of dualist reasoning.

This is the most compelling form of what has traditionally been labeled the mind-body problem: we cannot fit consciousness into our picture of the world, even though this seems to be the only picture possible. I will call this form the mind-body paradox. It is Kierkegaardian in the way it confronts us. For Kierkegaard, to be religious is to embrace an Absolute Paradox: that of something infinite, God, becoming incarnate in something finite, the person of Jesus Christ. For anyone struck by the Cartesian's arguments, the problem is how to include consciousness, something which seems immaterial, within a world which seems wholly material. The mind-body paradox should appear much more daunting than Kierkegaard's, however--while Kierkegaard's God is at best hidden and Kierkegaard's believer is a mystic, the evidence of consciousness is all around.

Ordinarily, the mind-body problem is portrayed more modestly, with none of this air of paradox: What is the relation between consciousness and the brain? Is the mind something different from the brain, though connected with it, or is it the brain?[10] Often it is depicted, particularly by the optimists or ideologues of contemporary cognitive science, as being, if not a matter of normal science, then at least a form of stagesetting for a

---

10. See, for example, Thomas Nagel's short introduction to philosophy, What Does It All Mean? (New York: Oxford University Press, 1987), pp. 27, 28.

scientific revolution and a new paradigm. The solution is seen awaiting new mechanical models, more ingenuious than before. Dualism is considered a form of impatience, the sin of concluding merely from our present inability to figure out a physical basis for our phenomenology that it has none.

This is the criticism of even those who conclude that the explanatory gap is permanent because we will always lack the conceptual resources to bridge it, such as Locke and perhaps Hume. Cartesian intuitions, on this so-called noumenalist view,[11] are the best evidence against materialism but are not good enough because they can be explained on other grounds--that is, by limitations on human

---

11. Hume writes that "as the constant conjunction of objects constitutes the very essence of cause and effect, matter and motion may often be regarded as the causes of thought, as far as we have any notion of that relation" (my italics). See the Treatise, Book I, Part IV, Section V, par. 33; see also the Enquiry, secs. 8 and 12, pt. 1. Noumenalism not only appears to have been Hume's view but is suggested by the views of Locke, who argued that ideas of secondary qualities do not resemble the qualities themselves or any other physical qualities but were arbitrarily connected to them by God (see the Essay, Book II, Ch. VIII, sec. 13, and Book IV, Ch. III, secs. 12 and 13). Recently, a noumenalist view has been defended by Joe Levine, "Materialism and Qualia--the Explanatory Gap," Pacific Philosophical Quarterly 64 (1983), pp. 354-361; and by Colin McGinn, "Can We Solve the Mind-Body Problem?", in his Problems of Consciousness (Cambridge, Mass.: Blackwell, 1991); and it can easily be read into Thomas Nagel, "What Is It Like to Be a Bat?," op. cit., and The View from Nowhere, op. cit. McGinn (op. cit., p. 3) coins the term "cognitive closure" to pick out the conceptual inadequacy he thinks we humans have in psychophysical matters. Flanagan, op. cit., is responsible for this use of the term "noumenalism," a perhaps unfortunate choice, since here, unlike in Kant's use of the term, "noumenal" properties are known--directly, through experience.

conceiving, akin to blindness. Considering the mind-body problem from this perspective, a person responds to the question of whether we can solve it with a yes-and-no answer. We cannot, since the inadequacy of our conceptual resources are seen to create a permanent explanatory gap, but actually we can, since from this fact it would follow that dualists cannot meet their burden of proof and that materialism is true.[12]

The Cartesian rejects these assessments. The Cartesian, orthodox or nouveau, views the explanatory gap as a gap not in us but in the world. And the inference that the world contains nonphysical properties comes, according to the Cartesian, not as a last-ditch resort to mysticism, an expression of impotence at an intractable problem, but as a clearcut solution, a conclusion guaranteed by argument; it comes not by default but by principle. The Cartesian argues that whatever limitations there may be in our understanding of the world from limitations in our conceptual resources, they are irrelevant to understanding the inference to dualism. That inference is sanctioned, it is held, not by an inability to explain but by simple, obvious principles which, even if not themselves easily explainable, are justly regarded as reasonably certain. But even the Cartesian should regard the Cartesian conclusion as mysterious.[13]

---

12. McGinn, op. cit., pp. 17-18.
13. See Flanagan, op. cit., pp. 9-11.

The mystery is to reconcile the Cartesian conclusion with the considerations that motivate scientific materialism, for the clear-headed Cartesian should recognize the force of most of them, too. The Cartesian's would not agree that these considerations defeat Cartesianism, but it is even open to the Cartesian to see this as a paradox.

What are the considerations Descartes felt to mandate the inference to dualism? His writings suggest two kinds of arguments, one more successful than the other.[14] The one he is most identified with, the less successful one, I shall call the Cartesian argument. It is a modal argument, employing a premise linking the conceivability of differences to the genuine possibility of those differences. We can conceive of differences between the mental and the physical because of our intuitions surrounding disembodiment and zombiehood. We can also conceive of having multiple physical realizations (although Descartes himself never wrote of this): having different bodies from those we in fact have or bodies made of different material. From the further premise that the genuine possibility of a difference between two things entails an actual difference, it seems

---

14. It is typical of many of Descartes's critics to attack strawman arguments and to ignore his actual arguments. One of the most flagrant examples of this is Gilbert Ryle's attack on what he calls "Descartes's myth" and "the dogma of the Ghost in the Machine" in his The Concept of Mind (New York: Barnes and Noble, 1949), pp. 11-24, where none of Descartes's actual arguments are even alluded to.

possible to distinguish selves and souls from human bodies.

Clearly, there are naive versions of the first premise, the conceivability-to-possibility principle, which won't work. Consider the most unsophisticated version possible. To claim simply that the mere distinguishability of things into separate apparent sorts of things entails a real distinction between them invites the objection that there are often two ways of picking things out which, even though they may seem to come apart in imagination, are in fact necessarily connected. Although it may have once been conceivable, for example, that water was distinct from hydrogen oxide, all that follows is that there are two different ways of referring to water, linked a posteriori.

This problem might not appear fatal to the modal argument. The Cartesian can take advantage of a more sophisticated premise by setting a higher standard on what is taken to deliver genuine possibilities. Descartes' own standard was that of clarity and distinctness; another, suggested by remarks by Kripke, requires that the case of distinguishability not be explicable as an illusion of contingency. Yet I shall argue below that the more work this higher standard is given, the more implausible it is.

And the modal argument fails for other reasons.[15]

---

15. One essential-properties objection is due to Thomas Nagel, The View from Nowhere, op. cit. pp. 47-48; the style-of-reference objection is due to Brian Loar, "Phenomenal States," in Philosophical Perspectives 4 (1990), pp. 84-85.

Consider the Cartesian's claim of distinguishing phenomenal states from physical states. As in the water case, picking out things in different ways does not mean we have different things, but for different reasons. There are two kinds, which I discuss in more detail in Chapter Three. On the one hand, styles of reference we use in the case of the mental differ from styles paradigmatic of reference to the physical. Phenomenal states are ordinarily open to direct reference; the referring paradigmatic to physical states ordinarily is not direct. Phenomenal and physical concepts (and terms that express them) differ in conceptual role and may be cognitively independent. On the other hand, we have no reason to conclude from the modal argument alone that the properties which seem to distinguish a mental from a physical state are not both essential properties of one and the same thing. The modal argument, therefore, does not support the Cartesian's conclusion. It leaves open the possibility that the Cartesian merely employs different means--different styles of reference, different essential properties--to pick out the same things.

III. Descartes's Epistemological Argument for Dualism, the Functionalist Response and Two More Cartesian Intuitions

The other kind of consideration on behalf of dualism suggested by Descartes's writings was an epistemological argument. The argument requires careful scrutiny, but in this case, unlike that of the modal argument, something substantial survives. Descartes invites us to imagine that the external world is a mere illusion, that an evil deceiver is deceiving us about the world in every possible way. Descartes's thought experiment, exploiting intuitions of disembodiment, suggests the following argument: I can doubt everything physical about the world without doubting everything about the looks and feels in it, since the fact that it looks and feels to me the way it does is beyond doubt; therefore, there are features to the world beyond its physical features.

This argument is akin to a more recent one, derived by replacing doubt with knowledge. This is the so-called Knowledge Argument: that since I can <u>know</u> everything physical about the world without <u>knowing</u> everything about looks and feels there are features to the world beyond its physical features. But despite its similarities to the argument suggested by Descartes's thought experiment, the Knowledge Argument exploits neither kind of Cartesian

intuitions mentioned so far, of disembodiment or of zombiehood. While the doubt mentioned in the argument suggested by Descartes's writings raises the radical possibility of the absence of the physical world, the Knowledge Argument does not entertain this possibility or that of the absence of minds. It is compatible with denying them. It is based on a set of closely related but separate Cartesian intuitions. These I shall identify as intuitions of the duality of information. To know what is like to somebody to look at or feel something is intuitively a separate piece of knowledge from any of the knowledge you have of objects in the external world; the first is direct, the second is indirect, and either can exist without the other.

My main task in the essay that follows is to argue that the Knowledge Argument and arguments like it can be given more convincing defenses than they have been given, even by their current defenders. Even if it turns out our introspectively acquired phenomenal knowledge of looks and feels picks out the same pieces of the world as does our physical knowledge, it could only do so by way of properties of the world different from any of the properties by which our physical knowledge would pick out pieces of the world. This argument relies on the Fregean insight that meanings are separate from referents and provide routes to them in virtue of properties of the referents. Thus, thoughts

identifying the phenomenal pieces of the world with certain physical pieces of the world could do so, according to the Fregean, only in virtue of different routes to the same things, routes that would reach their referents by way of distinct mental and physical properties of the referents. It is this argument that sanctions the inference from our Cartesian intuitions of a duality of information about the world to the dualist conclusion about two distinct realms of properties in the world.

This is the fundamental dualist insight that survives the close scrutiny of Descartes's arguments. And later I offer as a conjecture that this is the dilemma about consciousness that is the core of the mind-body problem and makes it persist. This is what provides a rational, well-informed observer in a fairly stable state of reflective equilibrium about the makeup of the world with a sense of paradox by splitting this person between dualism and the fundamentally different account of scientific materialism.

The overall objective of this essay, then, should be seen as twofold. First, I try to motivate the conjecture that showing the Knowledge Argument unsound would be tantamount to solving the mind-body problem. Second, I defend the neo-Cartesian approach to the problem by way of the Knowledge Argument, or at least the claim that it has not so far been shown unsound and is strongly supported by common-sense principles.

I argue that there are two main sorts of common-sense counterarguments to the Knowledge Argument--a functionalist one, and what we might call a conceptual-role approach--and that neither is successful. The conceptual-role approach is analogous to the counterargument that defeats the modal argument. According to this non-Fregean line, the two forms of knowledge, phenomenal and physical, differ not in the properties by which or to which they refer but in their styles of representing the world. One style depends on direct demonstrative reference; the other depends on descriptive reference. They differ in their conceptual roles and may be cognitively independent. This form of counterargument worked against the modal argument for the following reason. In that case, there was no assurance that there was anything more to the appearances of contingency that seemed to make it possible to pull mentalistic and physicalistic representations apart than the possibility that two different kinds of reference were picking out the same things. In the case of the Knowledge Argument, on the other hand, there is no claim of contingency between pieces of the world, only of nonidentity between properties of it.

Assume, for the sake of argument, that it is not possible to pull the two kinds of representations apart, that they have the very same referents in common. Still, I shall argue later, the very distinction between direct demonstrative reference and descriptive reference in those

cases we human beings are most familiar with--our introspective, phenomenal knowledge and our nonintrospective physical knowledge--entails the existence of distinct sorts of properties. We directly pick out referents in introspection at least partly in virtue of their looks and their feels, and these are different from any of the paradigmatically physical properties in virtue of which we pick out aspects of our brains, whether we do this descriptively or demonstratively.

The other common-sense line of counterargument advanced against the Knowledge Argument has been to argue that although there are nonphysical properties they are functional properties--characterizable entirely by reference to causal relationships among stimuli, behavioral responses and psychological states. Thus, they would not be irreducibly mental. However, the reasoning by which the Knowledge Argument refutes physicalism also seems prima facie to defeat functionalism. Just as someone might know everything physical without knowing everything, so someone might know everything physical or functional without knowing everything. If someone were to know that some functional description picked out the same mental state as some mentalistic description, the person could only do so by way of distinct routes, ones that reached the common referent in virtue of distinct causal properties of the referent separate from its phenomenal properties. The Knowledge

Argument at least requires a functionalist to be an analytic functionalist, but it seems that the analytic functionalist falls prey too. A blind person, for example, might know that some mental state satisfies any functional description you like without knowing everything about it (such as what it's like to have).

IV. Absent Qualia Intuitions Against Functionalism and Problems of Skepticism

There are some who will remain analytic functionalists and resist this conclusion. If dualism is the alternative, they reason, then there must be a way out of the Knowledge Argument, however counterintuitive. But there is a further counterargument backing up the Knowledge Argument. It is rooted in intuitions that parallel the anti-physicalist intuitions of zombiehood. This further argument shows not just that mental properties are distinct from functional ones but that they don't even supervene. According to these new intuitions, it is possible for brain states or states of mind to satisfy a functionalist account of pain (or any other mental-state type) yet lack qualitative character. If these intuitions are right, then since it should by functionalism be like something to have these states but is not, functionalism is false. I will call these <u>absent-qualia intuitions</u>, and because of their resemblance to the others, I will call them Cartesian. The

nonqualitative functional states they suggest the existence of I will call absent-qualia states. We now have six categories of Cartesian intuitions, which I summarize in the following table.

Against Physicalism	Against Functionalism
Intuitions of Disembodiment	-----
Intuitions of Zombiehood	Absent-Qualia Intuitions
Intuitions of Multiple Physical Realizability	-----
Intuitions of Distinct Information about the Mental and the Physical	Intuitions of Distinct Information about the Mental and the Functional

Fig. 1 Six Categories of Cartesian Intuitions

The absent-qualia intuitions proceed from the following ideas. According to a standard account, the functionalist is committed to the existence of functional definitions for each type of phenomenal state. These definitions identify the states of that type with whatever states are characteristic effects of those states' characteristic causes and characteristic causes of their characteristic effects. Thus, pain, for example, is identified with whatever characteristically is caused by pain's causes and causes pain's effects--with whatever, say, is caused by bodily injuries, etc., and causes grimacing, the desire to be rid of pain, and so forth.

According to the most defensible of the absent-qualia

intuitions, whatever functional roles can be filled phenomenally, say with pain, can also be filled nonphenomenally, at least hypothetically. Phenomenal properties, or qualia, by these intuitions, are like the fluid of a hydraulic computer. If we suppose the calculations of such a device to be driven by the movements of fluid, these intuitions suggest that a functionally equivalent device, one that performed the same calculations and did this by the same program, could be driven in some other way, such as by the movements of electrical currents. In that case, a functionalist theory of fluid would be false and an "absent fluid" hypothesis, the hypothesis that a functionally equivalent device might lack fluid altogether, would be true. Yet the fluid in such a hydraulic computer would not be epiphenomenal, since its computations would occur in virtue of the movements of its fluid. By similar considerations, these intuitions go, there might be something lacking pain but functionally equivalent to someone in pain, going through the same pain-related mental processes but doing so in virtue of something other than pain. And the pain might be thought of much like the fluid: crucial to our thoughts and behavior and thus not epiphenomenal, even though its role could be filled by something else in a functionally equivalent system.[16]

---

16. This example is due to Ned Block, "Are Absent Qualia Impossible?", Philosophical Review 89 (1980), pp. 262-263.

It is worth recalling that these intuitions, although I call them Cartesian, are not Descartes's. Functionalism would be false for Descartes not because the functional roles of pains and other states can be filled both phenomenally and nonphenomenally but because pains and other mental states have no set functional roles at all.[17] Descartes's point is made not for pains but for thoughts: that humans aren't beasts or machines, since they can think and act on their thoughts in creative ways that go beyond the programmed responses of beasts or machines. Pain and all other phenomenal qualities are supposed to be a kind of thinking and would receive a similar account.

Rejecting functionalism has seemed to some to entail an extreme form of skepticism. If the causal connection between pain and pain belief could be broken and you were able to believe that you were in pain without being in pain, how can you ever know that you are in pain at all?

This view seems to assume that knowledge of one's phenomenal states requires one to distinguish them from all alternatives. But I shall argue in a later chapter that knowledge does not require that. It is no more sound to

---

17. These Cartesian intuitions against functionalism correspond to the disembodiment intuitions directed against physicalism. They would be placed in the empty upper-left space in Fig. 1. I believe that they can be defended as well, although I do not do so here. Such "madman intuitions" are discussed in the account of "mad pain" in David Lewis, "Mad Pain and Martian Pain," in David Lewis, Philosophical Papers (New York: Oxford University Press, 1983), vol. 1, p. 122.

argue this in cases of introspective knowledge than it is in cases of perceptual knowledge. Since the existence of Cartesian sorts of radical perceptual error does not jeopardize claims of knowledge in more ordinary cases, so long as these other cases are grounded in reliable mechanisms of belief formation, radical introspective error would not either.

Moreover, even if one could distinguish one's states from any alternative, this is not inconsistent with the existence of absent-qualia states, as the anti-skeptical argument requires. I can in my own case, I shall argue, know that I am having the qualitative states I am and not absent-qualia replicas because I can know a priori that in me, absent-qualia replicas are impossible. The psychological mechanisms required for direct references to qualitative states, I will argue, have side-effects upsetting any functional isomorphism between them and absent-qualia counterparts. But this is consistent with the existence of absent-qualia states in nonsentient creatures without direct demonstrative reference of an introspective sort.

V. The Structure of the Argument to Follow

In the seven chapters which follow, my aim is to clarify and to bolster the neo-Cartesian half of the mind-body paradox, to show that there really is a paradox, one that remains with us, undissolved. I do this by showing in more detail why the Cartesian intuitions set out in this first chapter, those underlying the neo-Cartesian view, remain intact, despite the best efforts of their critics. To accomplish this, I will more more fully describe and evaluate the pieces of reasoning set out above in which these intuitions have been put to use: the modal argument, the Knowledge Argument, the considerations behind the anti-skepticism argument, and my arguments against and for the possibility of absent qualia.

In Chapter Two and Chapter Three, on the modal argument, I compare Descartes's version of it--or, perhaps more accurately, one <u>like</u> Descartes's--with the superficially similar argument against materialism due to Saul Kripke, which is not dualist but is closer to agnosticism. The conceivability-to-possibility principle itself can be modified to escape many counterarguments against it; still the objection I mentioned remains. That is the objection that appearances of contingency may be illusory from picking the same things out both directly and descriptively, or through separate properties of them. I

argue that the Kripkean position, while not committed to the conceivability principle, misunderstands that because of the overwhelming case for materialism the opponent of materialism has the burden of proof, and that the Kripkean agnosticism is inadequate for meeting the burden. By contrast, I argue that Descartes understood the burden, and that while he failed to meet it, his effort foreshadows neo-Cartesian arguments which do meet it.

In Chapter Four, on the Knowledge Argument, I clear the way for the main argument of this essay by setting aside some standard objections, focussing on the specific version of the argument advanced by Frank Jackson. I argue that although past criticisms of Jackson's version are unsuccessful, it and its argumentative strategy must ultimately be rejected, since what they assume, that simply your knowing everything physical but not knowing everything is enough to contradict physicalism, is false.

This clears the way for setting out in Chapter Five what is right about the Knowledge Argument and about the Cartesian tradition more generally. I develop the different, stronger version of the Knowledge Argument according to which, by contrast with Jackson's version, the knowledge we have of some of our mental states could not be <u>about</u> those states unless we picked them out in virtue of properties of them distinct from any physical properties. Even though our knowledge of looks and feels is

distinguished from our physical knowledge in virtue of different forms of reference, this can only be so by way of properties different from any by which our physical knowledge refers. As I stated above, the very distinction between direct, demonstrative reference and descriptive reference in cases that are relevant entails the existence of distinct sorts of properties. If this is right, qualia provide routes to our mental states distinct from those provided by any physical properties, contradicting materialism.

In Chapter Six, I will fill out my argument that skeptical considerations are not the obstacles to the Knowledge Argument they may at first seem to be. As I stated above, there is no general epistemological principle backing an anti-skepticism argument for functionalism and against the Knowledge Argument. Even if there are failures of transparent access and incorrigible access to the phenomenal characters of our mental states, they do not threaten our ability in principle to know the characters of our states in <u>normal</u> cases. Any argument against the possibility of absent qualia based merely on the threat of skepticism requires an epistemological principle that makes our knowledge of our own qualia depend upon evidence more comprehensive than is justified. I will argue that our evidence is not as complete as published versions of the argument require; our evidence would be so complete as this

if we had transparent access to the phenomenal properties of our own mental states, but I will argue that we lack it in crucial ways.

In Chapter Seven and Chapter Eight, on the possibility of absent qualia, I fill out the position set forth above in support of that possibility, refuting fuctionalism and bolstering the Knowledge Argument. Even though there is no general epistemological guarantee that we can distinguish our qualitative states from our nonqualitative ones, I still argue in Chapter Seven and Chapter Eight that we can nevertheless distinguish our genuine states from ersatz counterparts on the alternative grounds mentioned above. However, I argue that this fact is consistent with the existence of enough absent-qualia states to undermine functionalism since, as I also said above, absent-qualia states are possible in nonsentient creatures that do not do any distinguishing.

CHAPTER TWO DESCARTES'S MODAL ARGUMENT

Descartes's principal argument in the Meditations for the dualism of mind and body relies on some strong tie between what can be conceived and what is possible. It is possible, and therefore actual, he argues, that his mind is distinct from his body because of what he can conceive about his mind and his body.

Just what this assumed connection is between the conceivable and the possible and how general and necessary the reliance on it is in arguments for dualism is the subject of this chapter and the next. Descartes's argument is a powerful one, and it can be made even more powerful by dropping some of its problematical but unnecessary aspects. This power, however, has gone largely unappreciated. Part of the reason for this is a failure to see what Descartes actually assumed about conceivability and possibility. Another part is a failure to understand how his argument can be and has been improved on.

I. From Conceivability to Possibility

Michael Hooker, for example, states that while Descartes rejected (P"), from the conviction that some theological mysteries were beyond human comprehension, he endorsed and relied on (P'), its converse, in his argument for dualism.

(P′) For all p, if p's truth is conceivable, p's truth is possible.

(P") For all p, if p's truth is possible, p's truth is conceivable.

"He argues that it is conceivable that the mind exists without any bodies existing, and from there concludes that distinctness is actual," Hooker writes.[1] But he contends that there is no notion of conceivability making Descartes's argument sound.[2]

Hooker also finds reliance on (P′), or what I shall call the Simple Conceivability Principle, in the writings of Saul Kripke. In a passage from Naming and Necessity, Kripke writes: "One can imagine ... various things in [Queen Elizabeth's] life would have changed: that she should have become a pauper; that her royal blood should have been unknown, and so on. One is given, let us say, a previous history of the world up to a certain time, and from that time it diverges considerably from the actual course. This seems to be possible. And so it's possible that even though she were born of these parents she never became queen."[3]

---

1. Michael Hooker, "A Mistake Concerning Conception," in Stephen F. Barker and Tom L. Beauchamp, eds., Thomas Reid: Critical Interpretations (Philadelphia: Philosophical Monographs, 1976), pp. 86-87.

2. Michael Hooker, "Descartes's Denial of Mind-Body Identity," in Michael Hooker, ed., Descartes: Critical and Interpretive Essays (Baltimore: Johns Hopkins, 1978).

3. Kripke, op. cit., p. 113.

Hooker takes the argument, through tacit appeal to the Simple Conceivability Principle, to be "that it is possible for someone to exist without the properties we can conceive them lacking."

He finds appeal to this principle, too, in Kripke's attempt at the end of Naming and Necessity to discredit certain arguments for physicalism, calling it a "contemporary version of Descartes's argument." According to Hooker's account of this "contemporary version," Kripke "argues from the conceivability of mind-body distinctness to its possibility, and from there, via the necessity of identity to dualism."[4] If that is right, the reliance on the Simple Conceivability Principle (P') is transparent.

Clearly, the principle has had supporters. Hooker appears to be right that Hume in the Treatise endorsed it. But Hooker is wrong about Descartes. The conceivability principle on which Descartes's argument for dualism depends is much more subtle than Hooker's simple principle. Hooker himself quotes this passage from Comments from a Certain Broadsheet:[5] "We should note that even though the rule,

---

4. Hooker, "A Mistake Concerning Conception," op. cit., p. 87.

5. At AT VIIIB 351-352. Even Hume, in two of the three passages Hooker cites supporting (P'), claims only that possibility can be derived from clear conceivability or from distinct conceivability. Hooker's error is made by others as well. See, for example, Christopher Hill's account of what he calls the "Cartesian argument" in his Sensations (New York: Cambridge University Press, 1991), p. 90, where he ascribes to Descartes the belief in something like (P').

'Whatever we can conceive of can exist,' is my own, it is true only so long as we are dealing with a conception which is clear and distinct..." (my emphasis).

I see no reliance on the Simple Conceivability Principle in Kripke's challenge to physicalism, even if there is tacit appeal to it elsewhere. I shall argue below that Kripke does not, contrary to Hooker, rely on quite the conceivability principle Descartes uses, if he uses one at all. Rather, Kripke argues only that our Cartesian intuition of mind-body contingency cannot be explained away by familiar means and that we should not endorse materialism until it is explained away.

My aim in this chapter is to set out Descartes's argument for dualism and, after dispensing with several side issues, to continue to focus on what is the central problem for the Cartesian: the nature and role of the conceivability-to-possibility principle. I will show, by mapping some of the logical geography in which they reside, how Cartesian views move beyond Hooker's simple principle to be subtler views than has sometimes been appreciated. I will argue ultimately that the Cartesian has a dilemma about what reading and role to give the conceivability principle, and that the choice between two alternatives distinguishes a position something like Kripke's, an agnostic one, from one like Descartes's, which I will call the orthodox Cartesian position.

In section II, I set out Descartes's argument from conceivability, and there and in section III, amend it to yield a conclusion directly contradicting psychophysical event identity. In section IV and section V, I distinguish clear and distinct conceivability from several other notions with which it has been confused, and I argue that several counterarguments to Descartes rest on this confusion. In section V I set out the Cartesian's dilemma about what reading and role to give the conceivability principle. I set out the two alternatives distinguishing the agnostic from the orthodox Cartesian positions, which I go on to examine and criticize in more detail in Chapter Three.

II. Descartes's Argument

In the Sixth Meditation, Descartes makes something like the following argument for believing that the mind is distinct from the body.[6] (I present the argument and the modifications of it which follow by way of schemata, with Greek letters as placeholders for names.)

---

6. At AT VII 78.

Descartes's Argument from Conceivability

(1) If I can conceive clearly and distinctly of α and ß that they are two things existing apart from each other, then α and ß are two distinct things.

(2) I can conceive clearly and distinctly of my mind's existing apart from my body.

(3) I can conceive clearly and distinctly of my body's existing apart from my mind.

Ergo, (4) my mind is distinct from my body.[7]

Nowhere does Descartes explicitly argue that pains, tastes, thoughts and the like--particular mental features of him--are also distinct from his body. What he does write, however, makes it easy to see how such an argument would go. Thought, according to Descartes, is an attribute of my mind, one essential to making it a mind and not something else--in fact, it is the only attribute essential in this way--and my

---

7. To be precise about this, Descartes would have said that they are really distinct. According to Descartes, there are three sorts of distinction: real distinction, modal distinction and conceptual distinction. The first holds only between substances (or entities), and it holds of substances A and B if and only if A can exist apart from B (which, in the case of substances, entails that B can exist apart from A). The second holds between modes or between a mode and a substance; it holds of mode A and mode or substance B if and only if A can exist apart from B (which, in this case, does not entail that B can exist apart from A--Descartes's shape is modally distinct from Descartes, since he can exist apart from that shape, although Descartes's shape, he contends, cannot exist apart from Descartes). The third holds between A and B when neither can exist apart from the other. Descartes gives the example of a substance and its duration: they are "conceptually distinct," although neither can exist without the other. See Principles of Philosophy, pt. I, secs. 60-62, at AT VIIIA 28-30.

being in pain is a mode of that attribute. Since my mind is distinct from my body, and since my being in pain is a mode of what it is that makes my mind a mind, my being in pain could not be a mode of my body's attributes, he would argue, and is thus distinct from my body.[8]

For Descartes, it was essential to proceed this way. It is not possible, he would have said, to argue that thoughts are distinct from my body just because I can clearly and distinctly conceive them existing apart. They cannot be conceived that way without the mediation of a mind, a soul. He thought of his mind, his soul, as a "complete thing," and this allowed him, he believed, to make certain inferences about it merely from his conception of it, without knowing everything about it. Particular mental aspects, on the other hand, he did not view as "complete things," and without the ability to place them within a "complete thing," he argued to Arnauld, there was always the chance that there would be hidden aspects to them outside his conception of them that would undermine such inferences.[9]

I now intend to depart from Descartes's version and

---

8. See Principles, pt. I, secs. 53, 56, at AT VIIIA 25-26. On the kind of mode pain is said to be, see Principles, pt. IV, secs. 190-191, at AT VIIIA 316-318.

9. For a discussion of Descartes's use in the argument for dualism of the notion of a "complete thing" and of the related notion of a substance, see Bernard Williams, Descartes (New York: Penguin, 1978), pp. 113-114 and 124-129. For more discussion, see the end of section IV below.

discuss an argument for dualism which parallels it but is slightly different. After Hume's critique of the notions of self and soul,[10] it would be best to rid the modal argument of any commitment to self or soul. This is possible and can be done without any deep cost to Descartes's insights.

Suppose, then, that in contradiction to Descartes's dualism, it turned out that some neurological theory of pain were true. In that case, pains would turn out to be instances of a particular kind of physical process, presumably a certain kind of stimulation--call it "C-fiber stimulation." Let us say, then, by stipulation, that the best materialist theory of pain, a true one if there is one, identifies pain with C-fiber stimulation, whether or not there may be other non-materialist theories that would better comport with the evidence. Thus, by stipulation, if any physical thing is identical to some pain, it is a C-fiber stimulation. The Cartesian would deny the truth, and even the possibility, of any such neurological theory of pain.[11] One Cartesian way of continuing the argument goes

---

10. In A Treatise of Human Nature, Book One, pt. IV, sec. VI.

11. Descartes comes closest to making such an argument in pt. I, sec. 61, of the Principles at AT VIIIA 30 when he argues that motion is modally distinct from doubt. For this reference and other assistance in understanding Descartes, I am indebted to Paul Hoffman. In her paper "Cartesian Dualism" (in Michael Hooker, ed., Descartes: Critical and Interpretive Essays, Baltimore, 1978), Margaret Wilson suggests caution, on the basis of Descartes's Sixth Meditation argument that the faculty of sensation does not belong to his essence and on the basis of several more obscure passages, in "attributing to him the view that we

as follows. Assume that (A) and (B) are true.

(A) I am having pain at time t.

(B) I am having C-fiber stimulation at t.

With (1′), slightly modified from (1), as one premise and the mental-token counterpart to (2) I label (2′) as another, this argument derives the counterpart to (4) which I label (4′).

Descartes's Argument, First Modification

(1′) If I can conceive clearly and distinctly of α and ß that they are complete things existing apart from each other, then α and ß are two distinct things.

(2′) I can conceive clearly and distinctly of the existing of my pain at t apart from my C-fiber stimulation at t.

Ergo, (4′) my pain at t is distinct from my C-fiber stimulation at t.

---

can clearly and distinctly conceive our sensations apart from any physical state or occurrence" (pp. 207-210). Apart from what I say in the text, however, such caution seems unwarranted; sensations are non-essential to Descartes, but so is any other particular mode of thought. The attribute of thought is essential to Descartes, but no mode of that attribute is. Wilson is right to assert that the argument in Meditation Six "is not intended by Descartes to make any claim that he can clearly and distinctly conceive his sensations ... independently of anything physical" but "is concerned only with the isolation of Descartes's essence as a thinking thing" (p. 208; italics in original). Yet neither assertion contradicts my view that Descartes is committed to the belief that sensations--as modes, or instances of modes, of Descartes's essential attribute--are, notwithstanding their own non-essential character, distinct from his body.

I will provisionally call (4') the Non-Identity Thesis. (I will make a slight alteration in the next section.) As a first approximation, it is fair to think of Descartes's argument for the Non-Identity Thesis as intended, if sound, to undermine any sense we have that the world is a wholly material world. I mean by that a world consisting entirely of things and stuff that occupy space, of fields of force that operate through space, and of the objective properties intrinsic to being something that occupies space or operates through it. In our own cases, in particular, it is intended to contradict whatever sense we have--through science, philosophy or naive common sense--that all explanations of our mental lives and behaviors are about the workings of physical parts of our bodies and of the fields of force passing through them. If the Non-Identity Thesis is true, not just some specific identifiable physical process fails to be the process of pain but all actual physical processes of the brain fail. The term to the right of the identity sign, by stipulation, names the actual physical process--whichever one it happens to be--picked out by the true neurological theory of pain, if there is one. I will use the term materialism loosely to denote this sense of ours that the world is wholly material.

I will use the term token physicalism to denote the view that every individual mental item, or every token of a type of mental state or process, is identical to an

individual physical item, or a token of a type of physical state or process. Descartes's argument is inconsistent with token physicalism. According to the argument, at least one mental item, my pain at t, is not identical to any physical item.

Now, consider two ways that Descartes's argument for the Non-Identity Thesis could be true. These are two ways in which premise (2'), that I could conceive clearly and distinctly of the my pain at t existing apart from my C-fiber stimulation at t, could turn out to be true. The two ways reflect two distinct Cartesian intuitions I discussed in Chapter One.

If disembodiment can be conceived clearly and distinctly and the conceivability principle is true, the Cartesian reasons, then my pain at t could exist apart from any body--and thus could exist without my body--and thus cannot be identical to any physical feature of it. On the other hand, if multiple realizability can be conceived clearly and distinctly, then if the conceivability principle is true my pain at t could once again exist apart from my body. It could exist without the very body I now have or, even if I might have this very body, without the physico-chemical constitution this very body now has, and thus cannot be identical to any actual physical feature of it. This, then, is the outline of Descartes's argument.

III. Caveats About Mental Objects and Mental Events

Before I begin to assess the core intuitions underlying Descartes's argument in the next section, I will modify my representation of it one last time to sidestep a difficulty with an assumption underlying this version of it. The assumption is problematical, but it can be dispensed with without jeopardy to the argument. Moreover, despite what some philosophers have claimed, the offending assumption is not even made by Descartes.

The assumption is that there are "mental objects"--painful pains, itchy itches, red after-images, and so forth--with some of the same observed properties as physical objects. Sometimes this complaint is made against "sense data" or "mental particulars" or "phenomenal individuals," but it can be made against mental <u>parcels</u> or <u>quantities</u> as well--anything which is essentially mental and instantiates or realizes observed properties. The complaint against mental objects is that it is easily conceivable that nothing mental has the properties they seem to. If my seeing red requires there to be a red mental object for me to see, there really must be something red in the universe. But the fact that I see red does not seem to entail that there is anything red in the universe, whether outside my body or inside my nervous system. Brains, after all, are gray throughout.

The materialist might then just concede this much of the Non-Identity Thesis--that *if there is anything* identical to my pain at *t* then it is distinct from my C-fiber stimulation at *t*. And the materialist might then deny, on the basis of an argument parallel to the one against red sense-data, that there is anything satisfying the referring expression "my pain at *t*." The materialist could then go on to assert that materialism requires something different, event identities rather than object identities. On this story, it requires only that my having of pain at *t* = my having of C-fiber stimulation at *t*. Here, the materialist argues that the previous situation does not arise: the event of my seeing red is not itself red and thus does not have a property, redness, which my brain may lack, but the event of my having of pain still has all the properties, such as hurting, that the event of my having C-fiber stimulation has.

But the materialist's strategy is of limited utility. The dualist can modify the argument to escape it. From what I have said of Descartes's argument for the separability of pains from the body, it is clear that Descartes did not even share these problematic assumptions. There are no *pains* for Descartes apart from *being in pain* or *having pain*, and there are no *red images* apart from *having a red image* or *seeing red*. Both having pain and seeing red for Descartes are modes of the attribute of thought, an attribute of the soul.

They are not substances or relations between substances. Following Descartes in this respect, the dualist can adopt the ontology of events, replacing references to mental objects with references to mental events. Thus, in the previous statement of Descartes's argument, the conceivability principle (1') can be replaced by (1"), in which reference to things is replaced by reference to events, and (2') and (4') can be replaced by (2") and (4") in which reference to my pain at t is replaced by reference to my having pain at t. From here on, it is this more particular version of the conclusion that I will refer to as the Non-Identity Thesis.

Descartes's Argument, Final Modification

(1") If I can conceive clearly and distinctly of α and ß that they are two events existing apart from each other, then α and ß are two really distinct events in this world.

(2") I can conceive clearly and distinctly of my having pain at t apart from my having C-fiber stimulation at t.

Ergo, (Non-Identity Thesis) my having pain at t = my having C-fiber stimulation at t.

Although this argument is closer in spirit to Descartes's views than the previous one, there is, however, evidence that Descartes would not have endorsed it. Premise (1), Descartes's conceivability principle, is true of complete things--substances --which "depend on no other

thing for [their] existence";[12] events would hardly seem to be examples of substances or complete things for Descartes. Thus, premise (1") is not a special case of (1') for Descartes. I will henceforth make the very un-Cartesian assumption (one perhaps more acceptible to Hume, who believed minds to be simple bundles of loosely connected perceptions) that (1") is true if (1') is true. Descartes himself could probably have endorsed a slightly altered version of the argument. By revising the antecedent of (1") to read that I can conceive clearly and distinctly of <u>each</u> without the other, and by adding a new premise (3"), which bears the same relation to (3) which (2") bears to (2), we get an argument that should have been acceptible to Descartes.

IV. Clear and Distinct Conceivability

Let us now return to the subject with which the chapter began, the relation Descartes assumes to hold between conceivability and possibility. Much skepticism has been directed at conceivability principles like (1), (1') and (1"), but some of it has been misdirected. In this section, I will show how a version of Descartes's conceivability principle can be defended against some of this skepticism.

As I showed above, there are critics of Descartes like

---

12. <u>Principles</u>, pt. 1, sec. 51, at AT VIIIA 24.

Hooker who have claimed that statements like (1), (1′) and (1") are derived on the basis of a fallacy--one of inferring possibility from bare conceivability. The mere fact that I can conceive that such-and-such is the case does not entail, at least not in the general case, that it is possible that such-and-such is the case. Descartes's contemporary Antoine Arnauld was another such critic. Arnauld, for example, argued that some people can conceive of right triangles without the Pythagorean property of having the square on the hypotenuse equal to the sum of the squares on the sides.[13] But although they may be conceivable, it surely does not follow that they are possible. In general any principle claiming that true possibility follows from some mere subjective sense of possibility is problematical at best.

Earlier, I cited textual evidence to support the view that, whether or not Descartes is wrong about the relation he claims to hold between conceivability and possibility, he is not guilty of the fallacy of simply confusing the two things. To see this better, consider Descartes's derivation of the conceivability principle, premise (1").

---

13. In the Fourth Objections at AT VII 201-202. Actually, the argument he criticizes has as a premise that one conceives clearly and distinctly of a right triangle while being uncertain over whether it has the Pythagorean property, but Arnauld takes Descartes to believe this entails that one conceives the right triangle without the Pythagorean property.

(1") If I can conceive clearly and distinctly of α and ß that they are two events existing apart from each other, then α and ß are two really distinct events in this world.

There is considerable textual evidence that he would have regarded it as coming logically from two further premises, which I will label (1A") and (1B").

(1A") If I can conceive clearly and distinctly of α and ß that they are two events existing apart from each other, then there is a possible world in which α and ß exist apart.

(1B") If there is a possible world in which α and ß exist apart from each other, then α and ß are two really distinct events in this world.

Recall Descartes's own words, supporting (1A"), from Meditation VI:[14] "[T]he fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God."[15] In remarks supporting (1B"), Descartes explains in the Principles (pt. I, sec. 60),[16] that "things which God has the power to separate, or to keep in being, separately, are really distinct." But he adds in Meditation VI that God is

---

14. At AT VII 78.
15. I will differ from the Cottingham translation however, in translating as "conceive" what is translated there as "understand."
16. At AT VIIIA 29.

unnecessary for the required separation: "The question of what kind of power is required to bring about such a separation does not affect the judgment that the two things are distinct."[17]

If Descartes is guilty of the fallacy of simply confusing conceivability and possibility, it will thus show up in premise (1A"). To show that he is not, I will consider Descartes's use of the terms "conceive" and "clearly and distinctly." The following account of his use of them, admittedly short of a full explication, is still enough, I believe, to discredit the fallacy objection.

To say that it is conceivably possible that unicorns exist or that Santa Claus delivers toys or that an angle can be trisected with compass and straightedge is not to say that it is possible but, syncategorematically, that it is conceivably so. All that is required is that it be conceivably possible <u>to</u> someone. What the bottom limit on conceivable possibility is, is hard to say. Presumably, there is someone to whom it is conceivably possible that the Pythagorean theorem is false, but it is unclear whether even Descartes, in the deepest depths of his skepticism, ever thought it conceivably possible that $2 + 3 \neq 5$. (It may be that he thought only that he could make errors in even the

---

17. This is Williams' interpretation of the sentence in his <u>Descartes</u>, <u>op. cit.</u>, pp. 106-107--with the further result that the existence of God is not required for the soundness of the Cartesian argument. I follow Williams also in thinking Descartes congenial to this further result.

simplest mathematical calculations). Let me suggest at least this much: for a proposition to be conceivably possible to someone either it must be true or there must be a gap between understanding it and knowing whether it is true.

There is a use for words like "conceive"--and for related words like "imagine," "suppose," and so forth--that is incompatible both with the way I will use them here and, I believe, with the way they are used by Descartes. On this use, a first-person assertion of conceivability can be challenged. If I say, "I can conceive of water being distinct from H_2O," somebody might reply to me, "You cannot. You may <u>think</u> you can. But, in fact, whatever you are conceiving of as distinct from H_2O is not water." Similarly, on this use of "conceive," an anti-Cartesian could assert, "You cannot conceive at all, much less clearly and distinctly, that your mind is separate from your body, or that your pains are distinct from all your bodily states. Whatever you are conceiving of as distinct from your body is not your mind and whatever you are conceiving of as distinct from your pains are not your bodily states, since your mind and your pains are not distinct at all."

These challenges, however, can be forestalled by using "conceive" in a different way--by having a "seemingly" built into it. Clearly, the challenges above are not open to a critic if I say, "I can seemingly conceive water distinct

from H_2O," or, "I can seemingly conceive my mind distinct from my body." Such first-person uses of "conceive" have a degree of incorrigibility that make them immune from such criticisms. I shall use the term "conceive" this way throughout this essay.

Now, the Cartesian thinks that more is needed in order to get real possibility. One needs clear and distinct conceivability. For Descartes, a clear perception is one "present and accessible to the attentive mind"; a distinct perception is one "so sharply separated from all other perceptions that it contains within itself only what is clear."[18] One detects clear and distinct conceptions simply by inspecting the contents of one's mind. But this is not enough to show that these perceptions are clear and distinct.

Bernard Williams has suggested paraphrasing the words "clearly and distinctly" as "however carefully and clear-headedly one considers the situation."[19] A difficulty with Williams' suggestion is that a proposition which is conceivable "clearly and distinctly" is not conceivable "however carefully and clear-headedly one considers the situation" if it is not conceivable at all when one considers the situation with little care or clear-headedness. Descartes writes that he can conceive

---

18. At VIIIA 22.
19. Williams, op. cit., p. 112.

clearly and distinctly of existing without a body, but it does not follow that he can conceive of existing without a body when he considers the situation carelessly.[20]

My own suggested paraphrases for the words "clearly and distinctly" are "with no humanly possible means of explaining away as an illusion ...," or simply "without humanly explainable illusion." According to Descartes's argument understood this way, possibility follows from conceivability if the conception cannot humanly be explained away as an illusion. Descartes's conjecture is that in the case of every one of a particular class of conceivings--that is, conceivings that particular mental items are distinct from particular physical items--the conceiving of something entails its possibility, since the conceiving cannot be identified as an illusion in these cases. This is enough to guarantee real possibility because God can bring into being anything conceived clearly and distinctly. Otherwise, he would have been a deceiver for creating us in such a way as

---

20. In Demons, Dreamers, and Madmen (Indianapolis: Bobbs-Merrill, 1970), p. 135, Harry Frankfurt, suggests the paraphrase "without reasonable grounds for doubting"; a clear and distinct perception that p is a perception without reasonable grounds for doubting that p. I can see two difficulties with this paraphrase. First, for Descartes, there are grounds for doubting even perceptions that are clear and distinct, removed in the Third Meditation only by the proof of a benevolent, nondeceiving God. Might they not be reasonable before the proof? Second, what if there is a paradox, as I claim there is, in thinking about the mind-body problem? In that case, there might be reasonable grounds from the materialist side for doubting dualism even though the case for dualism was clear and distinct. The possibility of paradox cannot be defined out of existence.

to be misled by such apparent possibilities.

There are many cases in which I can conceive that such-and-such is so even though it is not so, sometimes even though it is not even possible for it to be so. The Cartesian would argue, however, that in the psychophysical case it is not possible to explain away such conceptions as false conceptions: that it is not possible to appeal to any kind of illusion to explain away the Cartesian intuitions that mental states are distinct from physical states. The distinction, then, between merely conceiving and conceiving clearly and distinctly is a distinction for the Cartesian between a pre-reflective, merely subjective state and an intuition subjected to exhaustive critical assessment. Contrary to the objectors, only states that measure up to this high standard of assessment are taken to deliver true possibility. Even if there be doubts whether all instances of the conceivability-to-possibility principle (1A") are true, such doubts are distinct from the more mundane doubt about conceivability and possibility I began with.

How is this standard applied? Descartes believed that it was humanly impossible to explain his Cartesian intuitions away as illusions, but he did not need to establish this by eliminating ways he might go wrong one by one. That might be endless. He thought he could give an independent argument for his intuitions, one that already appeared conclusive on its face.

Descartes argued that there were two ways of obtaining a clear and distinct conception. One was through knowledge of all of a thing's properties. Knowledge so complete, however, is rare. The other way relied not on complete knowledge but rather on knowledge of a complete thing. By this, you will recall, he means something needing nothing else beyond itself to exist. According to Descartes, minds and bodies are complete things. Descartes had a clear and distinct conception of himself, he argued, because he had a sufficient conception of himself as a thinking thing to consider himself a complete thing, and thus of something which needs nothing else to exist, even though he did not have complete knowledge of himself. It was not part of that conception that he was bodily although he had a separate conception of his body as a complete thing that itself needed nothing else to exist, such as a mind.[21]

One kind of mistake in interpretating Descartes's argument results, then, from ¬ failure to see that it uses a two-tier conceivability principle. On Descartes's model, one must establish two prior things in order to establish that a conception rises to real possibility. At the first tier, one must produce a seemingly conclusive argument that it is a real possibility, one independent of the inability to explain anything away. Descartes writes of perception here: we might interpret him to mean that there must be a

---

21. Objections and Replies, AT VII 200-202, 220-225.

perceptual relation to the conception's logical relationships.[22] The model is mathematics, where we are supposed to proceed deductively from axioms that we just seem to see the truth of. Although I will not ascribe to this aspect of his epistemology, I will sometimes use perception as a metaphor in connection with this first condition. Not until the second tier do we find that the conception must be humanly unable to be explained away as a false conception of possibility. It must be clear and distinct. On Descartes's account, the first condition entails the second, but it is satisfied independently of the second. The entailment is not a tautology. Someone might take there to be a conclusive argument that something conceivable is a real possibility, while rejecting the second condition.

Here, then, is Descartes's two-tier model of how a mere dualist conception rises to the level of real possibility.

Descartes's Model

To establish a dualist conception as a genuine possibility one must establish that:

(Tier One) it is justified by a seemingly conclusive argument grounded in perceptions of logical relationships; *and*

(Tier Two) it is humanly impossible to explain away as an illusion.

---

22. Here I follow Frankfurt, op. cit., p. 133.

Contrast this with Fig. 2, which summarizes interpretations of the Cartesian conceivability principle which I discuss in this chapter and the next.

	Without a Second Tier of Scrutiny	With a Second Tier of Scrutiny
Bare Conceivability	Hooker Hill	Lycan Dennett[23]
Intuition of Possibility	Levine	Kripke McGinn?
Seeming Perception of Possibility	Williams McGinn?	Descartes

Fig. 2 Accounts of What It Is the Cartesian Believes to Guarantee Possibility

Hooker is, of course, correct that bare conceivability does not guarantee real possibility, but Cartesians do not necessarily believe it does.[24] Joseph Levine is correct that mere intuitions of possibility do not guarantee real possibility, but here again, no Cartesian, including Saul

---

23. See Daniel Dennett, Consciousness Explained (Boston: Little, Brown, 1991), p. 282. Dennett seems to take Descartes's conceivability to be bare imaginability. Although he mentions that it must be clear and distinct to guarantee possibility, Dennett construes this only as a vaguely "higher standard": "The force of such an argument [as Descartes's] depends critically on how high one's standards of conception are." Dennett is underestimating Descartes here; the lesson of Goldbach's conjecture is that it does not matter what you add if the first tier that of bare imaginability, since one may be able to imagine it both true and false even after explaining away all illusions.

24. See also Hill, op. cit.

Kripke, the target of his criticism, has claimed otherwise.[25]

V. A Dilemma for Descartes: Two Cartesian Views

I shall now argue that even when those mistakes have been eliminated the Cartesian faces the dilemma of choosing between two approaches to the conceivability principle that initially appear to have difficulties of their own.

Descartes is not guilty of confusing conceivability with real possibility. But it is a further matter, and quite a different one, whether the conceivability principle he employs is true. Recall the Cartesian hypothesis I set out earlier, which I now will dub Descartes's Conjecture and label as (DC).

> (DC) If I can conceive clearly and distinctly (and thus without any humanly possible means to explain away as an illusion) of a particular mental event α and a particular physical event β that they exist apart, then it is possible that they do.

How should a Cartesian construe the clear and distinct conceivability which (DC) requires of real possibilities?

---

25. Levine, op. cit., p. 356: "For what seems intuitively to be the case is, if anything, merely an epistemological matter. Since epistemological possibility is not sufficient for metaphysical possibility, the fact that what is intuitively contingent [like the statement that pain is the firing of C-fibers] turns out to be metaphysically necessary should not bother us terribly."

Consider two options. The difference between the two options is in how much work there would be for each of the two tiers of scrutiny to do. The first option is Descartes's. It is to let most of the work of the overall dualist argument be done by the "perceptions" of logical relationships and the deductive inferences that follow; the second tier comes in only to guarantee we cannot go wrong (because of God's goodness) in reasoning that cannot be explained away.[26] The second option is different. It lacks faith in conclusive arguments for dualism. Thus, it treats Descartes's evidence as mere intuitions and puts most of the work of discovering whether they represent a genuine possibility on determining whether we can explain them away.

The Cartesian faces a serious dilemma at this point. These seem to be the only two options for a supporter of the

---

26. See Williams, op. cit., pp. 106-108; for a recent critique of Williams on these and related matters concerning the ultimate basis of Cartesian principles and the problem of the Cartesian circle, see Georges Dicker, Descartes: an Analytical and Historical Introduction (New York: Oxford University Press, 1993), pp. 130-133. Williams writes that "the basic content" of Descartes's dualist position is given "at the subjective level": by what he conceives. Its being clear and distinct only guarantees objective truth by God's goodness. Keep in mind, however, that one can be Cartesian about mind and body without believing in God. For nontheist Cartesians, who do not have the general rule linking clear and distinct conception with objective truth, the link must be made piecemeal, on a case-by-case basis. This would make it imaginable that dualist "perceptions" might be wrong even if other "perceptions" were correct about the external world. This would leave a larger gap between "perception" and real possibility than Williams seems to allow but still a smaller one than that between mere intuition and real possibility.

Cartesian form of argument. But the second option appears to depend on a conceivability-to-possibility principle that is still unconvincing. And taking the first option, requiring the dualist conception to be grounded in perceptions of logical relationships, may seem impossible to satisfy. Let me say why.

If one takes the second option, one in effect construes (DC) as asserting that intuitions of possibility not humanly explainable as illusions guarantee real possibility. (DC) will then be unconvincing. Nothing will appear inevitable about (DC). Consider Goldbach's conjecture that every even number greater than two is the sum of two primes. It is easy to imagine a mathematician developing conflicting intuitions about the conjecture's truth value. It is at least imaginable that these could turn out to be impossible to disspell as illusions. The conjecture's truth value is currently unknown. Say that it turns out to be undecidable. But whichever truth value it has is the only one possible for it. There is thus no general guarantee that intuitions of possibility not explainable as illusions are true. It is hard to see why the special case of (DC) would be immune from this problem of conflicting intuitions.

On the other hand, if the "perception" option is chosen, the Cartesian has to provide a story about what it is that gives it prima facie plausibility independently of the absence of humanly identifiable illusion. Why believe

we perceive dualism true--or could deduce it from what we do perceive? It is fair to say that no Cartesian has yet managed to give a widely convincing story.

It seems to me that writers that might be broadly described as "Cartesian" fall into two camps, according to what they would say about this dilemma. There are those like Descartes who would continue to adopt the latter option, of construing Cartesian intuitions as perception-like. This camp is made up of those whom I labeled in the last chapter orthodox Cartesians.

Members of the second camp would select the other option for understanding (DC), that Cartesian intuitions are just intuitions, but with a caveat. The problem associated with that interpretation--that (DC) remains unconvincing--does not arise, because they do not endorse (DC). Instead, they remain noncommittal about (DC). Still, while making no claims of knowledge for (DC), they hold that the intuitions of possibility distinguishing the mental from the physical, namely the Cartesian intuitions of disembodiment and of multiple physical compossibility, have not yet been shown to be illusions. And they hold that until materialism shows this, it cannot be established as true. The members of this second camp are among those I labeled in the last chapter as agnostics.

Agnostics of this sort share with orthodox Cartesians both a commitment to mental realism and a belief that

Cartesian modal intuitions, because of which mental realism is irreconcilable with materialism, cannot currently be explained away as illusions. They differ in one important respect: the orthodox Cartesian does, and the agnostic does not, believe that it follows from an inability to dismiss the Cartesian intuitions that we know they are true, and thus that we know (DC) is true.

It may seem that in giving up (DC), the agnostic has given up the spirit of Cartesianism. But that assessment is premature. The heart of Cartesianism has always been to state clearly and persuasively the dissatisfaction that common sense has with materialism on the basis of intuitions of contingency between the mental and the physical and to force materialism to explain that dissatisfaction away. The agnostic believes that, whether or not one accepts Descartes's Conjecture, it is still possible to put materialism on the defensive in this way.

Whether the agnostic's belief about putting materialism on the defensive is right, however, is another matter. It is not. In what follows I shall argue that orthodoxy, despite serious flaws, is a sounder direction to take.

The Cartesian's problem is to set out a convincing way to link conceivability which is clear and distinct--which is humanly inpossible to explain away--with real possibility. The Cartesian's dilemma is that there is no obvious way to do this. Does the Cartesian have an independent argument for dualism that, as a matter of fact, cannot be humanly explained away as an illusion? Or does the Cartesian argument depend entirely on a conception's inability to explain away? It is unclear how to do the first; nobody has yet provided a widely accepted independent argument for dualism. But the second is equally daunting; there are obvious counterexamples to conceivability principles, and there is little reason to think that Descartes's fares better.

In order to include both approaches, I will call an argument about the relation of the mental to the physical a Cartesian argument if it has the form of Descartes's Argument or conforms to the agnostic alternative mentioned at the end of the last chapter. A Cartesian argument thus has one of two profiles. It may (a) endorse Descartes's Conjecture, (DC), (b) entail that the Cartesian's modal intuitions of disembodiment and multiple compossibility are irrefutable, and (c) entail that (a) and (b) are together incompatible with token physicalism. Or it may simply require materialism to show that no sound argument can

satisfy (a), (b) and (c), and find materialism wanting. Cartesian arguments are thus modal arguments, appealing crucially to intuitions about what is necessary or possible about the mental or the physical or the correlation between them. By my definition, there are Cartesian arguments not Descartes's.

Kripke's anti-materialist argument is one. I interpret Kripke as taking something like the agnostic response to the Cartesian's dilemma. However, although Kripke explicitly disavows Cartesianism, he does not explicitly endorse the agnostic alternative. Still, the view he sets out falls under that label or else is a close cousin to views that do. Unlike Descartes, Kripke does not endorse anything like the Non-Identity Thesis, the conclusion contradicting token physicalism. Instead, he concludes only that the familiar ways of explaining away the intuitions supporting it are unavailable. But this, he suggests, is enough to put materialism on the defensive.

In this chapter, I will compare Descartes's and Kripke's arguments for the premises supporting the Non-Identity Thesis. In the first four sections I focus on the Kripkean approach. In section IV I show that the Kripkean position does not meet the burden of proof against materialism required of it. I argue that despite a misreading of Kripke, Colin McGinn's complaints against his argument demonstrate its failure to meet its burden of

proof. McGinn's attempt to make similar criticisms of Descartes fails, but I go on in section V to make arguments which do defeat the orthodox Cartesian position. Still, I conclude by suggesting that the neo-Cartesian line of thought that I develop in Chapter Four and Chapter Five evolves naturally out of the orthodox Cartesian's view that we have conclusive arguments for psychophysical differences.

I. Kripke's Main Idea

Recall again Descartes's Conjecture.

> Descartes's Conjecture
> (DC) If I can without humanly explainable illusion conceive of a particular mental event α and a particular physical event β that they exist apart, then it is possible that they do.

Except for the caveat noted above regarding mental events, Descartes would have endorsed (DC), as well as the following more general principle that ranges over everything, not just events.

> The General Principle
> If I can conceive clearly and distinctly of α and β that they exist apart from each other, then there is a possible world in which they exist apart.

Recall that he would have done both on the basis of a more general conviction. Since our mathematical reasoning and

much of our metaphysical reasoning seems to be "perception-like" and without any grounds for doubt, then if that were not enough to guarantee truth, God, our creator, would be guilty of deception, something which moral perfection does not permit.

Unlike Descartes, Kripke nowhere endorses anything like (DC) or the more general principle. Thus, there is no reliance in his argument on God's existence or deceptions. General views about the relation between indubitability and truth, or that between conceivability and truth, play no role. Neither does he take our intuitions about mind and body to be "perception-like." He grabs the agnostic horn of the dilemma described at the start of the chapter: he takes our Cartesian intuitions of possibility to be no more than intuitions, but places the onus on the materialist to show why they do not reflect real possibilities. He makes his task easier by restricting the scope of discussion in two ways. First, he focuses on cases relevant to assessing token physicalism--to cases of mental things and physical things--rather than to produce a general account of conceivability. Second, he provides explanations in related cases for why intuitions of possibility would be mere illusions and shows that these explanations do not apply to the psychophysical case, but he is content to do this with just several cases and only one type of explanation.

One <u>would</u> have reasonable grounds for doubting the

possibility of a seemingly imaginable nonidentity of events, according to Kripke's now familiar story, if one could find what Kripke calls an "illusion of contingency,"[1] one arising from our picking out instances of the events by contingent properties of them. If, for some class of events, identifiable illusions of contingency such as this exhaust the reasonable grounds for doubt, then something analogous to a "clear and distinct" conception of distinctness between events of this class would be one which excludes such illusions of contingency.

Kripke's argument is that a familiar way we might go wrong in thinking about theoretical identification is not available to discredit our intuitions that mental states are distinct from physical states. Consider the identity statement "heat = mean kinetic energy." It might seem that it might have turned out false but Kripke argues this to be an "illusion of contingency." The reference of "heat" can be fixed by descriptions of the form "that which causes such and such sensations" or "that which we sense in such and such a way," descriptions referring to the sensations or to the way of sensing which we normally associate with being made to feel hot. These descriptions express contingent properties of heat, since we could be constructed differently and feel something quite different, or nothing

---

1. Saul Kripke, "Identity and Necessity," in Milton Munitz, ed., Identity and Individuation (New York: New York University Press, 1971), pp. 160ff.

at all, in the presence of what normally here and now does make us experience heat sensations. Thus, in seeming to be able to imagine "heat = mean kinetic energy" false, we imagine not heat to be distinct from mean kinetic energy but something else, which happens to be picked out the way we normally pick out heat.

By contrast, there is no such illusion of contingency, we are told, surrounding our seeming ability to imagine "my having pain at t = my having C-fiber stimulation at t" false. The reference of "my having pain at t" is fixed by way not of contingent properties but of essential properties. Heat could fail to feel warm, were we built differently, but having pain, Kripke argues, could not fail to feel the way it does.

Kripke elaborates the argument with considerations reminiscent of Descartes's thought experiment in Meditation One. Just as Descartes does, Kripke asks the reader to compare the actual world to an epistemically similar one. How can the necessity which the physicalist attaches to "pain = C-fiber stimulation," Kripke asks, "be reconciled with the apparent fact that C-fiber stimulation might have turned out not to be correlated with pain at all?" What if we reply by analogy to the case of heat's identity with mean kinetic energy? In that case, there might be beings who are in a qualitatively similar epistemic situation to what we are, picking out something the way we pick out heat, by the

way it feels, even though it's not heat and even though heat is mean kinetic energy. But we cannot by that analogy reconcile the physicalist's psychophysical necessity claim with the contrary way things seem to be. For if we were in a similar epistemic situation, picking out something the way we pick out pain, it would be pain; and if there were C-fiber stimulation without pain, that would contradict the necessity claim. Either way, there is no reconciliation.

Kripke's point is to put materialism on the defensive.

> Someone who wishes to maintain an identity thesis cannot simply accept the Cartesian intuition[s].... He must explain these intuitions away, showing how they are illusory.... Materialism, I think, must hold that a physical description of the world is a complete description of it, that any mental facts are "ontologically dependent" on physical facts in the straightforward sense of following from them by necessity. No identity theorist seems to me to have made a convincing argument against the intuitive view that this is not the case.[2]

Materialists cannot establish the materialist position, according to Kripke, until they show how the Cartesian intuitions of contingency between the mental and the physical are illusions.

---

2. Kripke, Naming and Necessity, op. cit., pp. 148, 155.

II. Lycan's Misrepresentation of Kripke's Conceivability Premise

I want now to compare a commentator's account of Kripke's argument with Kripke's own account. The comparison will help shed light on some of the subtle features of Kripke's proposals.

I shall assume that for all substitutions of singular terms for placeholders ⌜x⌝ and ⌜y⌝, ⌜x is distinguishable from y⌝ is true if and only if it is seemingly imaginable that there is a possible world W at which ⌜x ≠ y⌝ is true. Distinguishability of this sort is a kind of bare conceivability.

Now consider the following schema; if it were to yield true statements for all substitutions of rigid singular terms for placeholders ⌜a⌝ and ⌜b⌝, (DC) would easily follow.

(D) If a and b are distinguishable, then it is possible that a ≠ b, unless:
  (i) someone could be, qualitatively speaking, in the same epistemic situation as the one I now am in vis-a-vis a and b, and still in such a situation a qualitatively analogous statement to the statement that a and b are identical could be false, or
  (ii) there exists some third alternative explanation of the distinguishability of a and b.

Principle (D), with several modifications, appears as a

conceivability premise in William Lycan's reconstruction of Kripke's argument.[3] Even with these modifications that correct for some obvious defects and obscurities in Lycan's account of Kripke's argument, (D) is defective. I will begin with two arguments that show that it does not represent Kripke's position on how to explain away intuitions of psychophysical contingency as illusions. After presenting and defending them, I will present a modified version of (D) which escapes these arguments.

The first difficulty is that substituting "my pain" and "my C-fiber stimulation" for '_a_' and '_b_' yields, for the Cartesian, a true version of clause (i)--allegedly Kripke's account of how _illusions_ of contingency might arise. It is true for the Cartesian since in an epistemic situation qualitatively the same as the one I am now in, a qualitatively analogous statement to the statement "my pain = my C-fiber stimulation," the Cartesian thinks, _could_ be false. In fact, on the Cartesian view it _is_ false. For the analogue to "my pain" picks out the same thing "my pain" does in the present situation, and that is a different thing for the Cartesian from what the analogue to "my C-fiber stimulation" picks out, whatever that is. Since (i) is

---

3. See his "Kripke and the Materialists," _Journal of Philosophy_ 71 (1974), pp. 679; and his _Consciousness_, _op. cit._, pp. 11-12. I take "unless" to pick out exclusive disjunction, although Lycan is not explicit about this himself. Lycan also fails to restrict (D) to rigid singular terms, but without doing so, the principle's clause (i) fails to be true solely of _illusions_ of contingency.

regarded as true, the distinguishability of my pain and my C-fiber stimulation has to regarded as an illusion of contingency. But this, for Kripke's Cartesian, is a paradigm of conceivability entailing possibility, not an exception.

A second difficulty with (D) is that there actually is an illusion of contingency surrounding pain and C-fiber stimulation, one that would make it impossible to derive from (D) the possibility that pain ≠ C-fiber stimulation. The illusion arises not around my sensing of pains but around my sensing of C-fiber stimulations. Analogously to the case of heat, there is a situation qualitatively identical to the present one in which a qualitative analogue of my C-fiber stimulation is not C-fiber stimulation at all. Suppose it is A-fiber stimulation, involving not neurons but shneurons, appearing the same as neurons but with very different microstructures. But there would be an illusion of contingency, (i) would be true and it would be impossible once again to derive the distinctness of pain and C-fiber stimulation.

To overcome these difficulties, (D) must be revised. To alleviate the first difficulty, the condition for being an illusion of contingency must be revised so that the qualitatively identical epistemic situation being compared to the present one is a situation lacking the states at issue--for example, pain or heat. The point to be made

about pain is that, for Kripke, there is no epistemic situation qualitatively identical to the present one which lacks pain.[4] To alleviate the second difficulty, the singular terms must be considered one at a time. Illusions of contingency can arise in connection to each. We have an illusion of contingency like the one we find in the case of heat and mean kinetic energy, the case of illusion which Kripke and Lycan intended to describe, only if the references of <u>both</u> singular terms are fixed on the basis of descriptions that pick out a referent by contingent properties of it.

Thus, I propose the schema (D′) as what Lycan meant by his (D). Principle (D′) incorporates both revisions in the statement of the illusion-of-contingency clause, (i).

---

4. Lycan apparently intends that the second conjunct of (i), the illusion-of-contingency clause--namely, that "in such a situation a qualitatively analogous statement" to the identity statement "could be false"--takes care of this matter. It does in the case of heat; it does not, because of the argument in the text, in the case of pain. The conjunct can thus be dropped in this revision.

(D′) If a and b are distinguishable, then it is possible that a ≠ b, unless:

(i) there is an a′ and there is a b′ such that:

(a) someone could be, qualitatively speaking, in the same epistemic situation vis-a-vis a′ and b, where a ≠ a′ as the one I now am in vis-a-vis a and b, and

(b) someone could be, qualitatively speaking, in the same epistemic situation vis-a-vis a and b′, where b ≠ b′, as the one I now am in vis-a-vis a and b, or

(ii) there exists some third alternative explanation of the distinguishability of a and b.[5]

According to (D′), the possibility that heat ≠ mean kinetic

---

5. Some considerations suggest that clause (i) of (D′) needs a further conjunct, which I will call clause (c). Intuitions of contingency between a and b would be explained as illusions by clause (i) only if (i) included the condition that (c) there is an epistemic situation which is identical, qualitatively speaking, to an epistemic situation vis-a-vis a and b and in which a qualitatively analogous statement to the statement that a and b are identical could be true. Clause (c) is not required for the truth of my (D′), but without (c), principle (D′) would have odd result which Lycan's (D) already has, that the nonidentity of, say, molecular motion and C-fiber stimulation would not follow from their distinguishability. There are many examples of this odd result. The same could be said of heat and color, heat and water, water and color. But in at least some of these cases nonidentity seems to follow from distinguishability. Moreover, without (c), a way Lycan writes of his version of (D) is flawed and the way I write of (D) and my (D′) would be flawed: instances of clause (i) in (D) or in (D′), without (c), would not always constitute "explanations of distinguishability" or "explanations of illusions of contingency" when (i) was satisfied, since in the cases of molecular motion and C-fiber stimulation and of water and color, clauses (a) and (b) of (i) are satisfied but there are no illusions to explain.

energy does not follow from their distinguishability. Someone could be, qualitatively speaking, in the same epistemic situation vis-a-vis some thing heat* ≠ heat, as we are in vis-a-vis heat, and also in the same epistemic situation vis-a-vis some thing mean kinetic energy* ≠ mean kinetic energy, as we are in vis-a-vis mean kinetic energy. There is thus in that case the possibility of an illusion of contingency. We must look at the distinctness of both epistemic counterparts[6] heat* and mean kinetic energy* from their real-life counterparts as (D') requires (deriving an illusion of contingency by satisfying both subclauses (a) and (b) of clause (i)) in order to distinguish this from the case of pain and C-fiber stimulation. Here, there would be only one epistemic counterpart, C-fiber stimulation*, distinct from its real-life counterpart. In contrast to Lycan's (D), my (D') entails that the possibility of pain's being distinct from C-fiber stimulation follows from their distinguishability, unless there is an alternative explanation of it besides the one in clause (i). It is not possible in the case of pain, in contrast to C-fiber stimulation, that someone could be, qualitatively speaking, in the same epistemic situation vis-a-vis something distinct

---

6. Here I am using the term coined by Colin McGinn in his "Anomalous Monism and Kripke's Cartesian Intuitions," Analysis 37, no. 2 (1977), p. 78, where x is an epistemic counterpart to a iff$_{df}$ x is "some entity distinct from a which is such that it puts us in qualitatively the same epistemic state as a does in the actual world" (McGinn's emphasis).

from pain, some hypothetical state pain*, as we are in vis-a-vis pain. There is no pain* ≠ pain.

III. Kripke and Descartes's Conjecture

Let us suppose that clause (i) of my (D′) is an accurate explanation of at least some illusions of contingency: that all seeming possibilities which (i) is true of are illusions of contingency. Let's also assume that Descartes's Conjecture follows from my (D′) by some simple additional principles.[7] Would my (D′) then be true--and, thus, should Kripke endorse it? Does the real possibility that two things are distinct follow from their conceivable distinctness when there are no humanly possible explanations that the conception is illusory?

Lycan does not challenge his principle (D), and I claim that my (D′) improves on (D). But both these principles have difficulties that I have already discussed. Statement (D′), and like it Descartes's Conjecture, would follow from a more general conceivability principle: that every conception that some particular object or event or proposition is possible is the conception of a real possibility unless it can be explained away as a false conception. This general principle would be to possibility

---

7. Such as that (D′)'s clauses (i) and (ii) exhaust the ways of humanly explaining away as an illusion the distinguishability of two events.

just the reverse of what Occam's Razor is to actuality: the former would multiply the possibilities it is rational to posit, while the latter constrains the pieces of actuality it is rational to posit. We surely do not know this general principle to be true, for reasons already stated. I can conceive certain unproven mathematical statements both to be true and to be false. But since there will be among them statements with unprovable truth values, where neither the statements nor their negations can be explained away, it would follow, if the general principle were true, that these statements would be both true and false, which is absurd.

Or assume that some scientific proposition about physical possibility received immunity from reasonable grounds for doubt as the result of some ideal scientific theory, a theory which turned out to be the best one humanly possible. On that basis, we might claim a clear and distinct conception of the possibility of some specific instance of the general proposition. If this is not enough to guarantee true possibility, it might seem that nothing is. But it seems conceivable that even the best human science might get it wrong. It seems entirely possible that the human mind is constructed in such a way that there are intuitions of possibility which are illusory but which are also humanly impossible to explain away. For example, are there faster-than-light velocities in a vacuum? We seem to be able to tell a story, according to some physicists, by

which there are such velocities, consistent with the known laws of physics. Does it follow that there really are such velocities? It seems entirely conceivable that even though it would be humanly impossible for us to explain away the conceivability of such velocities there might no possible worlds in which such velocities occur.

Does it help to narrow the question from that of whether the general principle is true to that of whether Descartes's Conjecture itself is true? It may seem, at least initially, that it does not. It may seem that while mental events seem to be distinct from physical events, perhaps we are just wrong about this. And it surely would not seem to follow in any obvious way just from our being wrong about this that there are humanly possible explanations of why we are wrong. Kripke's argument asserts that the means we employ in explaining away the illusion that heat is distinct from mean kinetic energy are unavailable in the case of psychophysical intuitions because of differences in our modal intuitions about heat and (for example) pain. Although we can imagine feeling hot in the absence of heat, we seem unable to imagine feeling pain in the absence of pain. But we might come to wonder whether we have a good enough command of such modal intuitions to make such judgments. Perhaps, we might think, the human mind is not constituted in such a way for us to know much about the modal properties of pain. Or perhaps we are constructed in

such a way that, even after making every possible relevant consideration, we have firm convictions about the modal properties of our qualitative states which are just wrong. What can be said against such skeptical doubts?

Before continuing, I will distinguish two kinds of doubts one might have about principles like (DC). On the one hand, there are the doubts we have in cases like those of faster-than-light velocities, cases where we may merely seem unable to rule out some epistemic possibility. Here, it is reasonable to think that conceivability does not entail possibility. On the other hand, let us distinguish doubts about those cases from doubts in cases of a very different sort, cases where we seem to be able conclusively to assert the reality of the possibility. For Descartes, this latter one is the case we are presented with by the Cartesian argument. To the orthodox Cartesian, it seems that there is a conclusive argument for the distinctness of mental and physical events, based on the reasoning that we human beings have essential properties no physical bodies could have. This reasoning for the orthodox Cartesian is in a sense just as conclusive as the reasoning that convinces us that 2 + 3 = 5, since both forms of reasoning deliver clear and distinct ideas.

Thus, the orthodox Cartesian views agnosticism about the distinctness of mind and body as just as extreme a position as a corresponding kind of agnosticism about the

proposition that 2 + 3 = 5. As I said earlier, Descartes believes that proof of a benevolent God is required to forestall these kinds of agnosticism, but that is only because they _are_ so extreme. Processes of nondivine creation or creation by a malevolent God could conceivably make us go wrong even in our clear and distinct ideas, he argues, and only belief in a benevolent God could fend off skeptical sources as extreme as this. However, although an extreme position, it is one a God-less Descartes would be forced himself to assume.

Contrast this case with that of faster-than-light velocities. Agnosticism here does not seem so extreme. It may seem strange to think that we could go wrong in even an ideal science about an undemonstrated-but-not-ruled-out theoretical possibility. But neither is so strange as to think that we could seem to have _conclusively_ demonstrated something but still go wrong, something as firm as the proposition that 2 + 3 = 5.

Whether the orthodox Cartesian is right to have this view is the subject of the next section. But this view is different, as I have said, from Kripke's. What is Kripke's view? Does Kripke accept or reject Descartes's Conjecture? I see no evidence in Kripke's published writings of the commitment to Descartes's Conjecture which Lycan reads into them. Certainly the arguments about mind and body in _Naming and Necessity_ do not depend on either. Nowhere does Kripke

state that he has ever identified all the possible illusions of contingency that could lead a conception of possibility astray of real possibility or that we could ever hope to identify them all. Nor does he state that *if* he had identified them all and *if* in some particular case none of them obtained *then* real possibility would be guaranteed. At least in his published works, he remains noncommittal.

Instead, Kripke's position is simply that it *seems that* pains can exist without any brain states and brain states can exist without any pains. This is the Cartesian *intuition*--it is something, Kripke claims, which is *intuitively the case*. And materialism, you will recall, must explain the intuition away. As Kripke writes:

> Someone who wishes to maintain an identity thesis cannot simply *accept* the Cartesian intuition.... He must explain these intuitions [sic] away, showing how they are illusory. This task may not be impossible.... The task, however, is obviously not child's play....[8]

One reason it is not child's play is the argument reviewed in the previous two sections: that the way of showing our intuitions of distinctness between heat and mean kinetic energy to be illusory is unavailable to discredit the Cartesian intuition.

Thus, the threat is not necessarily that unexplainable distinguishability would guarantee dualism by some

---

8. Kripke, *op. cit.*, p. 148.

conceivability principle like Descartes's Conjecture, since Kripke never endorses such a principle, nor even that that would give a strong reason to believe dualism.[9] The threat to materialism for Kripke is that a necessary condition of establishing materialism is to provide an explanation of how the modal intuition cited above is an illusion. For Kripke the materialist has the burden of proof in the face of its being intuitively possible that mental and physical states can exist without each other.

Lycan's error is to make Kripke into too orthodox of a Cartesian. A Kripkean agnostic might eschew commitment to a conceivability principle altogether within the spirit of what Kripke writes. Lycan committed the technical errors which I identified in the last section by jumping improperly from Kripke's specific examples to a rule linking conceivability and possibility. Not only did he stumble while trying to generalize from the examples, but his enterprise of constructing a rule was hopeless from the start. For Kripke, or at least the Kripkean agnostic, there is no rule. His claim is only that the materialist must

---

9. My being able to imagine a headache existing without any brain state, as Nagel interprets Kripke's view, "provides a strong reason for believing" that a headache can exist without a brain state, since "this can't be explained as the imagination of something that only feels like a headache but isn't." See Nagel, op. cit., p. 46. But in the absence of a further story perhaps about the content of what it imagined and why imagining it helps, mere imaginability of a possibility would not seem to provide any reason at all for believing that the possibility is real.

explain away the Cartesian intuition but cannot in the familiar way. If the Kripkean agnostic has such a limited aim, however, then he does not need any conceivability principle at all. He then escapes the difficulties I have shown conceivability principles to have. It will seem to the Kripkean agnostic enough to place the burden of proof on the shoulders of the materialist.

IV. McGinn on the Agnostic and the Orthodox Cartesian

However, it is here that the agnostic falters. For the materialist does not have the burden of proof. If what I earlier called the mind-body paradox is as I have argued--a clash of two independent but contradictory lines of argument, both apparently conclusive, for and against materialism--then it will not do to support the anti-materialist side with a mere inability to explain away the modal intuitions. Materialism has, so to speak, already met a burden of proof; that's what gives us the seemingly conclusive line of argument for the materialist side of the paradox. A mere inability to explain away the modal intuitions would only establish a lapse in us, not a paradox in our very conception of the world. The agnostic never shifts the burden of proof to the materialist. He does not show how our modal intuitions would do that.

These points are closely related to the criticism of

the Kripkean position made in a recent essay by Colin McGinn. He is right in much of his criticism of Kripke, but he is wrong in extending it to the orthodox Cartesian position. It may help the reader understand the differences among these three positions--those of McGinn, the Kripkean agnostic and the orthodox Cartesian--to look closely at the arguments of McGinn.[10]

McGinn defends the view that the mind-body problem arises "because we are cut off by our very cognitive constitution from achieving a conception of that natural property of the brain (or of consciousness) that accounts for the psychophysical link." Our being "cut off" this way is an instance of what he calls "cognitive closure," which he defines as follows: "A type of mind M is cognitively closed with respect to a property P (or theory T) if and only if the concept-forming procedures at M's disposal cannot extend to a grasp of P (or an understanding of T)." McGinn argues that realizing that cognitive closure it at work allows both realism about the mental and a naturalistic solution to the mind-body problem: "cognitive closure with respect to P does not imply irrealism about P. That P is (as we might say) noumenal for M does not show that P does

---

10. See McGinn, "Can We Solve the Mind-Body Problem?", op. cit. A similar argument is made against Kripke (not Descartes) in Levine, op. cit. Here Levine calls psychophysical identities "epistemologically inaccessible," meaning that "we don't have any way of determining exactly which psychophysical identity statements are true."

not occur in some naturalistic scientific theory T--it shows only that T is not cognitively accessible to M."[11]

These views McGinn employs directly against Descartes, Kripke and, by implication, the theorist I have called the "Kripkean agnostic." In response to Cartesian intuitions "to the effect that the relation between conscious states and bodily states is fundamentally contingent," McGinn offers what he calls a "diagnosis." "The reason we feel the tug of contingency, pulling consciousness loose from its physical moorings, may be that we do not and cannot grasp the nature of the property that intelligibly links them.... Not grasping the nature of the connection, it strikes us as deeply contingent; we cannot make the assertion of a necessary connection intelligible to ourselves."[12]

McGinn's account is flawed by a misreading of Descartes and Kripke. There isn't in either writer the connection between Cartesian intuitions and the inability to grasp the psychophysical link which McGinn asserts. Obviously, neither writer holds that it is simply a failure to grasp the connection between consciousness and "its physical moorings"--or that it's simply its striking us as "brute and unperspicuous," as McGinn also writes--that leads him toward (in Descartes's case, to) dualism. But neither does the Cartesian intuition that Descartes and Kripke begin with--

---

11. McGinn, op. cit., pp. 3-4.
12. Ibid., pp. 19-20.

its being intuitively the case that we could exist disembodied--seem to be coextensive with a failure to grasp some hypothetical psychophysical link. Such a failure is neither necessary nor sufficient in any obvious way for our having the Cartesian intuition. It is not obviously necessary, since even if we were to grasp some hypothetical psychophysical link, that would not seem to rule out having the Cartesian intuition--the possibility might still be open for all we knew that in some other world we could exist nonphysically, without the link. It would be one thing to grasp the psychophysical link but a very different thing to know that all possible things are physical. Nor is it obviously sufficient; we can imagine an intelligent-sounding zombie making a Cartesian argument to us on the basis of its inability to grasp the psychophysical link, and we can imagine this even though we imagine the zombie has no consciousness, and thus no Cartesian intuitions, at all.

However, despite the misreading, McGinn has a legitimate complaint against Kripke. Even if we allow that our intuitions of possibility may sometimes arise independently of what is cognitively closed to us, they are no general guarantee of genuine possibility, even if there are no means for explaining them away as illusions. And this is so for McGinn's reason: we may be kept from explaining them away not because they are intuitions of genuine possibility but because explanations of how they are

illusory are cognitively closed to us. Not only are the more general conceivability principles which I have mentioned thus suspect but they provide no support for the more specific case of Descartes's Conjecture. Kripke's Cartesian needs (DC) to refute materialism. If it is unavailable, the Kripkean position collapses.

Not only does an argument like McGinn's rob (DC) of the support of a more general principle but McGinn also supplies an argument against (DC) specifically. He argues that any concept of a psychophysical link is cognitively closed to us, and that since we can explain away intuitions of contingency between the mental and the physical in that way, there is no reason to reject materialism. I shall not dispute the first part of McGinn's argument. The concept of a psychophysical link is cognitively closed to us, the Cartesian would assert, in part for the reason both McGinn and Descartes before him give. Since the concept of a brain state is a spatial concept but the concept of a phenomenal state is not, as both argue, we cannot conceive of something's being both a brain state and a phenomenal state except, at best, in a brute fashion. The concept of a psychophysical link, however, is supposed to make the connection intelligible, not brute; thus, they conclude, no such concept exists.

The flaw in McGinn's account comes in the second part--his claim that our inability to explain away our Cartesian

intuitions is explained by cognitive closure and that we thus have no reason to reject materialism. Suppose that it would turn out that our having these intuitions does not depend upon our having the cognitively closed concept and that our inability to explain the intuitions away is not itself explained by our having the cognitively closed concept. Suppose, that is, that both our Cartesian intuitions and our inability to explain them away arise independently of any instance of cognitive closure. Then the cognitive closure of a psychophysical link would not be any barrier to inferring genuine from seeming possibility.

Now, this is in fact no help to the Kripkean agnostic; as I have portrayed him, he cannot take this position in response to McGinn. This is because the Kripkean agnostic has no faith in our Cartesian intuitions beyond our inability to explain them away. This is just what separates him from the orthodox Cartesian, who believes that the intuitions have prima facie validity independently of our inability to explain them away, which adds very little. McGinn's argument thus works against Kripke and the Kripkean agnostic. But in that case it is overkill. For McGinn's ultimate conclusion is materialist, and as I remarked above, the materialist's argument against Kripke is more direct than this. It is simply that the materialist does not have the burden of proof; the anti-materialist or agnostic does. Making that point does not require what McGinn goes to such

pains to provide, a demonstration of how it is that Cartesian intuitions mislead. The materialist can make the same kind of counterargument that Hume made against the Design Argument in the days before Darwin: we may not know how materialism is true but since we know it is possible and does not require miracles to be true we have no reason to doubt that it is, in some fashion or another.

But this means that McGinn has no argument against the orthodox Cartesian. McGinn notes that Descartes "explicitly argued from (what he took to be) the essential natures of the body and mind to the contingency of their connection." McGinn suggests that if we "abandon the assumption that we know these natures, then agnosticism about the modality of the connection seems the indicated conclusion."[13] It seems to be McGinn's view that we should abandon that assumption, since the essential natures of body and mind are linked in a way we can never know. Actually, McGinn's cognitive-closure argument gives us no reason to abandon the assumption, since it is an open possibility that the assumption was arrived at without any of the closed-off concepts. Imagine somebody--call him McGone--who has brain damage in the area of his brain where he would have developed the conception of the psychophysical link if there were such a link to develop a conception about and if it were open to humans to do that. If there is a psychophysical link it is cognitively closed

---

13. Ibid., p. 20.

to McGone, whether or not it is closed to McGinn and the rest of us unimpaired people. But it would seem to be an open possibility, at least epistemically, that were dualism, perhaps counterfactually, true, McGone could "know the essential natures" of phenomenal states, employing a system of concepts independent of the damaged area of his brain, and convincingly argue for dualism on that basis. Nothing McGinn has argued rules that out.

Still, McGinn's argument places a heavy constraint on any argument against materialism. The presumption must be made that, everything equal, materialism is true, and that if it seems otherwise this is so for McGinn's reason or some other. That should be the presumption, and any dualist or even agnostic view has the burden of showing that Cartesian intuitions cannot be explained materialistically, simply as outcomes of our having a conceptual deficit in thoroughly material brains. Any such view must demonstrate how it that these intuitions arise independently of the conceptual difficulties which McGinn is right to claim we have. The Cartesian attempts to do this. Although I argue in the next section that the orthodox Cartesian does not succeed, I shall argue in the next two chapters that the neo-Cartesian does.

V. Flaws in the Orthodox Argument: The Essential-Properties and the Conceptual-Role Problems

The orthodox Cartesian idea is that it seems possible for me to exist without my body by inspection in my mind of what is needed to be me and what is needed to be my body. And its seeming possible is supposedly not just an intuition but is something for which there seems to be a very good argument.

Recall premise (2") of my modified version of Descartes's argument, the premise that asserts that psychophysical contingency can be conceived of clearly and distinctly.

> (2") I can conceive clearly and distinctly of my having pain at t apart from my having C-fiber stimulation at t.

As we have seen, Descartes's own argument for it requires the existence of a soul, depending upon a premise like (2).

> (2) I can conceive clearly and distinctly of my mind's existing apart from my body.

His argument for (2) is something like this. I am essentially a thinking thing. I know this because I cannot conceive myself without thinking. I know that from directly inspecting my idea of myself, my mind. My body is essentially extended. I know this because I cannot conceive

of it without extension. Again, I know that from directly inspecting my idea of body. Thus, since I perceive by direct inspection that they have different essential natures, I can conceive in a perception-like way the possibility of myself, my mind, existing separate from my body. Descartes assumes here that there cannot be a single complete thing with two essential properties, one mental, one physical.

As I stated at the outset, it would be appealing to keep the Cartesian method of argumentation while avoiding problematic premises like its (2). It would also be appealing to drop Descartes's overbroad application of the concept of thought. An argument that does both things might go like this. My pain at t is essentially a conscious mental state. I know this because I cannot conceive of having it without being conscious. I know that from directly inspecting my ideas of pain and of being conscious. My C-fiber stimulation is essentially spatial. I know this because I cannot conceive of it without its being spatial. Again, I know that from directly inspecting my idea of a bodily process. Thus, since I perceive by direct inspection that they have different essential natures, I can conceive in a perception-like way the possibility of my pain at t existing separately from my C-fiber stimulation at t.

Descartes's main argument for dualism appears in Meditation Six. There, Descartes justifies (2) by asserting

only that "on the one hand I have a clear and distinct idea of myself, insofar as I am a thinking, non-extended thing; and on the other hand I have a distinct idea of body, insofar as this is simply an extended, non-thinking thing." Premise (2) would follow straightforwardly, from the principle of the indiscernibility of identicals. But where do Descartes's clear and distinct ideas of himself and his body come from? They presumably come from the thought experiment of Meditation One and his res cogitans and wax arguments of Meditation Two. It seems possible to Descartes that he can exist in the absence of his body, and he cannot find any source of error in this. Close scrutiny of himself tells him that only thinking is essential to him; that is the only one of his former beliefs about himself that is not subject to doubt. Moreover, thinking about the wax, he sees that despite all its changes, the one thing that does not change is its having extension. This he knows from a mental inspection of the idea of body. Once he knows that, he argues, he knows that he can clearly and distinctly conceive himself apart from his body. His reasoning is that this follows from his knowing that thinking is not essential to his body. Since his idea of a body is essentially the idea of something extended in space, he argues, it does not contain the idea of thinking.

Elsewhere, Descartes also indicates that he needs the assumption that his idea of himself is the idea of a

complete thing. Neither the surface of an apple nor the properties of a triangle are complete things but are ontologically dependent upon the existences of apples and triangles. This is unlike his body and that collection of his private mental states that he has come to call himself, according to Descartes, which he thus concludes can exist, in some sense, in the absence of anything else.

Descartes writes in the passage from the Sixth Meditation that he has a clear and distinct idea of himself as "a thinking, non-extended thing," an idea of his body as "an extended, non-thinking thing." This suggests an argument Descartes might make: My body is essentially an extended thing; I am not; thus, I am distinct from my body. There is a standard response to this argument. It is to reject the second premise on the grounds that it is not possible for anybody to be immaterial. But although this seems conclusive to many materialists, making this response is in fact a bad strategy for the materialist. It leaves the materialist with only a standoff, and the materialist should want more. All the Cartesian needs is the slimmest logical possibility--the possibility in just one possible world. It does not need to be technologically possible or even possible in nature. It's enough, Descartes would insist, that God could do it. But he would also insist that a God is not even needed to do it either, at least conceptually, so that agnostics and even atheists could

become dualists. Thus, the Cartesian will always insist that surely it's at least logically possible, or at least that the materialist doesn't know that it isn't. The Cartesian cannot satisfy a burden of proof on this, but it is hard to see how the materialist could satisfy the Cartesian.

The materialist, however, has a better argument for denying the second premise. It is to allow the logical possibility that somebody might be immaterial but to assert that as a matter of fact Descartes is not. The problem then is to explain Descartes's intuition that he himself, not just somebody, could be immaterial. A Kripkean way would be to explain the Descartes's intuition as the different intuition that there could be an immaterial person whose point-of-view on the world was qualitatively the same as Descartes's--a Berkeleyan counterpart, we might call him. To that account, however, Descartes should insist that he can imagine he himself being immaterial, not just somebody like him. Thus, the materialist counterargument requires the following thought experiment.

Let us suppose that we were to invent a process that makes it possible for people to dematerialize. Let us suppose, moreover, that in the state of dematerialization people can continue to function in many normal human respects. Wells' Invisible Man may come to mind, but I do not mean that it becomes possible just to become

transparent. I mean to suppose that one might lose one's very physicality this way. Suppose that we select as a guinea pig for trying out our process someone I will call Ms. X. We place Ms. X in our dematerialization chamber and throw the switch. At the outset her vision has been directed away from her body so that while she can continue to see during dematerialization she is unable to see whether or not she any longer has a normal human body. Suppose that she is also able to sense the world throughout the process through dematerialized versions of her other four senses.

Our process works by gradually replacing Ms. X's physical features with nonphysical features. At the outset she weighs 120 lbs. After five minutes, she is down to 60 lbs., and after ten minutes, her weight is zero. It is as Fig. 3 shows.

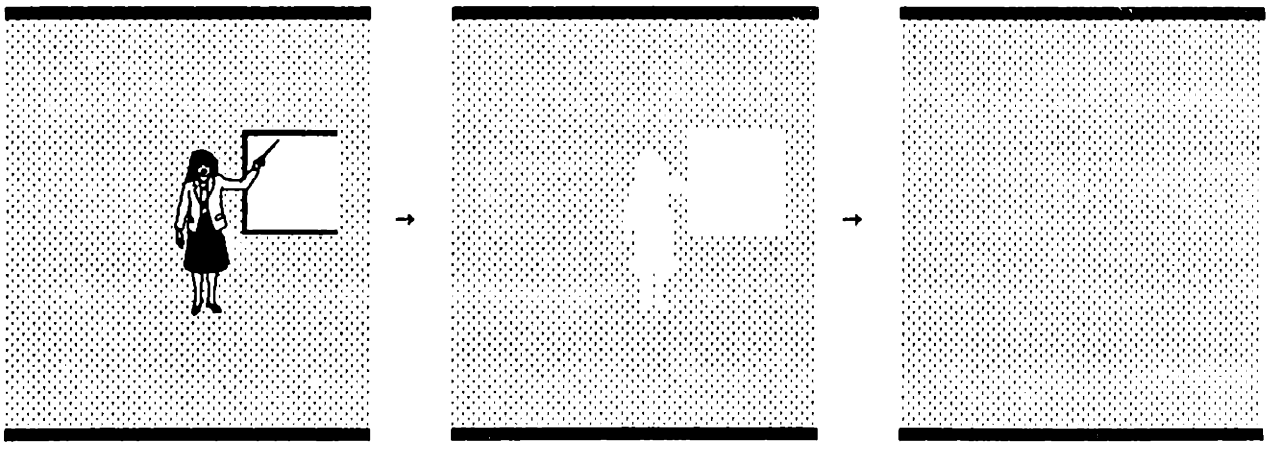

Fig. 3 Ms. X's Dematerialization Adventure

She is now dimensionless, and we can also suppose that she no longer has a location in any normal sense although she seems to have one. She is still in the dematerialization chamber from her point-of-view, waiting for something to happen. After twenty more minutes, Ms. X is returned to normal bodily form.

Ms. X's adventure provides a way to account for Descartes's intuition that he might be immaterial without contradicting the claim that he in fact is not, since Ms. X could be entirely physical in the actual world even if she might become immaterial in some other possible world. The hard-headed materialist (even the soft-headed one!) may balk at supposing that such a process as this is possible, but Descartes should not have any difficulty supposing this. It seems conceivable, clearly and distinctly, that such a process is possible, and God can bring about anything we can conceive clearly and distinctly. And it does not conflict with Descartes's claim that bodies are necessarily extended, since once Ms. X is no longer extended she no longer has a body.

If we shift from talk about mind and body to talk about mental and physiological states, a similar point can be made, although it takes a bit more work. Consider this argument: My physiological states are essentially extended; my mental states are not; thus, my physiological states are distinct from my mental states. Ms. X's adventure does not

so far contradict this argument nor even seem to contradict it. For none of the physiological states she has in the story lose their physical character. Still, it is conceivable that some of Ms. X's states, which just happen to be physiological, could have been nonphysical. For consider the nonactual possible world in which as Ms. X stands waiting in the dematerialization chamber we change our minds and do not throw the switch. Instead, we go out for lunch. I have assumed that Ms. X cannot tell whether we have thrown the switch or not; now assume that there is no experiential difference at all for Ms. X over which it is.

The materialist should reject the second premise here, too. Supposing that Ms. X's and Descartes's mental states are in fact physiological is consistent with the possibility that those very experiences might not be.[14] Let's call the property of giving the very experiences Ms. X has X-mentalizing. Some of Ms. X's actual physiological states X-mentalize, then, even though there may be nonactual immaterial states that also X-mentalize; but this is consistent with all her actual states, including those that X-mentalize, being essentially physical.

---

14. Exploiting Kripke's strategy, McGinn ("Anomalous Monism ...," op. cit.) explains my intuition that my pain at t ≠ my C-fiber stimulation at t as the conceiving that an epistemic counterpart of my pain (≠ it) ≠ my C-fiber stimulation, allowing that my pain = my C-fiber stimulation. But this seems not to do justice to Descartes's intuition that his very states might then and there have been immaterial, not just states like them.

Once again, the materialist may balk at supposing this possible. But the orthodox Cartesian, at whom the argument is directed, should not have any difficulty supposing this. The Cartesian should be able to imagine, for example, that this very pain is both in close union with, and in no union with, a physiological state. But the Cartesian will not be able to find any evidence to distinguish supposing that from supposing that the pain is in fact realized physiologically but might not be.

The orthodox Cartesian assumes that mental properties and physical properties cannot be essential properties of the same thing. Thus they are said to be different properties of different things. The above arguments reject the Cartesian's assumption. But can the argument work without appealing to two separate sets of essential properties which preserve token physicalism but contradict type physicalism? Materialists have adopted two strategies to complete their work against the Cartesian. One is functionalist. The only nonphysical properties are taken to be topic-neutral functionalist ones. But there are well-known difficulties with functionalism, which I argue in this essay to be insurmountable.

The other is the conceptual-role strategy I described briefly in Chapter One and will described in more detail in the next two chapters. It exploits the idea that the styles of representation we use in the case of first-person

reference or mental representation or in the use of phenomenal-concept terms differ from the styles we use in representations paradigmatic of our talk of physiological states. The former ordinarily involve direct reference in ways the latter do not. The two sets of representations differ in conceptual role and may be cognitively independent without always differing in the properties by or to which they refer.

Some Cartesian intuitions can be explained away in this way. The term "I" picks out its referent in virtue of no properties of the referent. Clearly, it fills a distinct conceptual role from terms like "my body." Thus, Cartesian intuitions associated with the terms "I" and "my body" can be explained in terms of their different conceptual roles, without invoking any odd properties. The difference between "my pain" and "my C-fiber stimulation" cannot be explained so easily, however, and here the orthodox Cartesian is on the right track. The illusion that they cannot pick out identical tokens is explained by different conceptual roles, but the terms seem to pick out tokens by way of distinct properties of the tokens. In general, concepts with different roles can refer by and to the same properties, but in the psychophysical case, the very different modes of presentation seem to lead to referents in virtue of different properties. Arguing for this result means leaving the orthodox paradigm for a rather different argumentative

strategy, however, one suggested by Descartes's writing but not exploited by him.

CHAPTER FOUR THE KNOWLEDGE ARGUMENT

According to the Knowledge Argument, there are certain things which you can know everything physical about but not know everything about and which, because of that, make physicalism false. The Knowledge Argument says that your having this extra knowledge--the knowledge about these things which you could have over and above your knowledge of everything physical about them--depends on the existence of special nonphysical properties. These properties, called qualia, are supposed to be properties of your experiences; it is supposed to be in virtue of them that you can say of your experiences that they feel a certain way or look a certain way to you. The Knowledge Argument purports to show that qualia exist. Moreover, it purports to show that qualia are nonphysical and nonfunctional and that, because of this, physicalism is false. The position which it purports to establish I shall call property dualism.

I agree with the critics of the Knowledge Argument that this is an unattractive conclusion; it is hard to know how to fit extra nonphysical properties into the world picture we get from physics and biology. Nevertheless, my aim in this chapter and the next is to show that, despite rumors of its demise and despite the best efforts so far advanced against it by its critics, the Knowledge Argument remains alive and well. I shall focus in this chapter and in the first part of the next chapter on the specific version of

the Knowledge Argument advanced by Frank Jackson. There I argue that although previous criticisms of Jackson's version are unsuccessful, it and the argumentative strategy underlying it must ultimately be rejected. The assumption on which they rely--that simply your knowing everything physical but not knowing everything is enough to contradict physicalism--is false. I shall then in the next chapter develop a different, stronger version of the Knowledge Argument which escapes this and other outstanding objections. According to it, the knowledge you have of some of your mental states could not be about those states unless you picked them out in virtue of properties of them distinct from any physical properties. If this is right, qualia provide routes to your mental states distinct from those provided by any physical properties, contradicting physicalism. The Knowledge Argument is thus a substantial obstacle for any defender of physicalism and, if it is to be defeated, this will happen only with arguments more subtle, and probably more counterintuitive, than any previously made.

I. Descartes's Argument from Doubt

On the basis of a passage in the Discourse on Method,[1] Arnauld attributed to Descartes an argument which he

---

1. In the Fourth Discourse at AT VI 32-33.

paraphrased thus: "I can doubt whether I have a body.... Yet for all that, I may not doubt that I am or exist, so long as I am doubting or thinking. Therefore I who am doubting and thinking am not a body. For, in that case, in having doubts about my body I should be having doubts about myself."[2] As Arnauld points out, this argument is fallacious. It is also, as Arnauld sets the argument out, equivocal. One argument that can be constructed out of Arnauld's first three sentences is as follows.

Arnauld's Representation of the Argument from Doubt
I can doubt whether my body exists.
I cannot doubt that I exist.

Ergo, I am not identical to my body.

Arnauld's fourth sentence is a reductio ad absurdum justification for the conclusion. The argument can be rephrased thus. Assume that the conclusion is false and that I am identical to my body. Then, since the first premise is true and I can doubt whether my body exists, it follows from the law of the indiscernibility of identicals, were the first premise wholly extensional, that I can doubt whether I exist. But this contradicts the second premise. Thus, the assumption is false and the conclusion is proven true.

This argument for the truth of the conclusion is invalid because both premises are partly intensional. Thus,

---

2. In the Fourth Objections at AT VII 198.

the conclusion cannot be derived from the premises on the basis of the law of the indiscernibility of identicals, since the premises do not predicate anything of the particulars referred to in the conclusion. Thus, the reasoning of the purported reductio is unsound. From the falsity of the conclusion and the truth of the first premise it does not follow that the second premise is untrue and that I can doubt whether I exist, since the first premise is intensional and does not allow substitution salva veritate of coreferential expressions that come after the verb.

Arnauld, however, is wrong to ascribe this argument to Descartes. Nowhere does Descartes argue for dualism on purely epistemological grounds. The passage from the Discourse that Arnauld represents as containing this argument actually has a different conclusion from what Arnauld takes it to have. The passage's conclusion is not the conclusion printed above but rather the statement that I can conceive clearly and distinctly of my mind's existing apart from my body, the premise of the Cartesian argument I considered in the last chapter. The passage from Descartes also has a different and more complicated structure, relying not on the indiscernibility of identicals but on the notions of essence and completeness I also discussed there. Descartes, in fact, acknowledges in the Meditations[3] that an inference of the kind Arnauld ascribes to him would be

---

3. In the Second Meditation at AT VII 27-28.

fallacious and correctly asserts that he is innocent of it; Arnauld cites this passage but inexplicably allows his complaint to remain.[4] Descartes's argument for dualism is based upon modal intuitions about mind and body rather than the epistemological intuitions Arnauld claims it to be based on.

This criticism of Descartes has continued into modern times.[5] Although it is not applicable to Descartes, this and criticisms related to it may be applicable to other, more recent anti-materialist positions which, unlike Descartes's, are solely based on epistemological intuitions.[6]

Now consider a slightly different argument. This is not an argument ever given by Descartes but it is suggested by Descartes's initial thought experiment in the Meditations. In the end, it cannot be adequately defended on the basis of Cartesian considerations alone, but I will argue that it suggests forms of arguments that can be

---

4. In the Fourth Objections, op. cit.

5. Peter Geach, in his God and the Soul (New York: Schocken, 1969), p. 8, accused Descartes of committing a fallacy like the "masked man" fallacy discussed by Stoic logicians: that since I know who my father is but not this masked man, my father is not this masked man. In his Descartes, op. cit., p. 112, Williams criticized Geach's assessment of Descartes along lines related to my criticisms of Arnauld's assessment.

6. See Richard Brandt and Jaegwon Kim, "The Logic of the Identity Theory," Journal of Philosophy 64 (1967), pp. 534-535; and Thomas Nagel, "Physicalism," Philosophical Review (1965), pp. 344-345. They do not provide any published examples, however, to which their criticisms clearly apply.

defended against all common-sense counterarguments. The argument runs as follows.

A Cartesian Knowledge Argument
I can doubt everything physical about myself.
I cannot doubt everything about myself.

Ergo, there is something missing from the physicalist story about me.

This, you will recall, is the argument I claimed in Chapter One to survive the scrutiny of Descartes's intuitions about mind and body and to provide a source for the neo-Cartesian's case for property dualism. Notice that its plausibility need not rest on construing the premises intensionally. Interpret doubt here to be a two-place purely extensional relation.[7]

The first premise seems to be validated by Descartes's thought experiment. In Meditation Two, Descartes seems to be able to place in doubt everything physical about himself. It becomes an open possibility for him that God has created for him the delusion that he has "a face, hands, arms and the whole mechanical structure of limbs ... called the body," something having "a determinable shape and a definable location and [occupying] a space in such a way as

---

7. This contrast between intensional and extensional forms of arguments for dualism can be found in Paul Churchland, "Reduction, Qualia, and the Direct Introspection of Brain States," Journal of Philosophy 82 (January 1985), pp. 25-26.

to exclude any other body." He also places in doubt whether he has a material soul, "tenuous, ... permeat[ing his] more solid parts," by which he was "nourished, [and] moved about."[8]

The second premise, on the other hand, might seem to be validated by Descartes's conclusion that, despite these doubts about his physical nature, he exists and is a thinking thing. But it is here that Descartes's support for such an argument as this starts to run out. These things can be doubted. It is not just that the pervasive doubts of Meditation One reasonably introduce in Descartes the suspicion that all his initial beliefs may be false. "So what remains true?" he asks at the outset of Meditation Two. "Perhaps just the one fact that nothing is certain."[9] The more important point is that this is not a mere suspicion. It is a real problem once Descartes realizes at the beginning of Meditation Three "this slight reason for doubt": that "it would be easy for [God], if He so desired, to bring it about that I go wrong even in those matters which I think I see utterly clearly with my mind's eye."[10] This source of doubt is eliminated for Descartes once he proves to his satisfaction the existence of a nondeceiving God. But two difficulties remain. The first is that this will not help the general reader who does not accept

---

8. At AT VII 26.
9. At AT VII 24.
10. At AT VII 36.

Descartes's proof about God. Recall that Descartes does not rely upon God to produce the radical error. For those who suppose that there is no God and that, as he writes, "I have arrived at my present state by fate or chance or a continuous chain of events, or by some other means, ... since deception and error seem to be imperfections, the less powerful they make my original cause, the more likely it is that I am so imperfect as to be deceived all the time."[11] Without his proof about God, Descartes has no means to establish the second premise, nor does the general reader who denies indubitability. The second difficulty that remains is that once Descartes does establish the second premise for himself by proving to himself that God exists, he no longer has reason to accept the other premise, that he can doubt everything physical. For God's existence makes at least some of his physical beliefs about himself reliable.

II. Some Other Knowledge Arguments

Consider now a different but related argument associated with Thomas Nagel. Buried in his writings are the elements of a more successful version of the Knowledge Argument than Descartes's. In a well-known passage, Nagel reminds us that bats seem to have a very different form of experience than we human beings do, something we find

---

11. At AT VII 21.

ourselves unable to fully imagine. This is because they get around by echolocation, a sensory modality we do not have. We do not know what it is like to echolocate. We might be able to imagine some parts of what it is like, but there are gaps.[12] Now imagine a superchiropterist, someone who is not only the world's authority on bats but knows everything a chiropterist could ever possibly hope to know about bats. Since our superchiropterist could not know fully what it was like to be a bat, there would be gaps in even this person's knowledge. Now consider the following argument.

A Knowledge Argument Suggested by Nagel's Account

The superchiropterist knows everything physical there is to know about bats.

The superchiropterist does not know everything about bats.

Ergo, there are truths that escape the physicalist's story.

Since this is not an argument Nagel pursues in this article, I will focus in the remainder of this chapter and in the next on the account by Frank Jackson, who does pursue

---

12. For scientific work on bat phenomenology, see Steven P. Dear, James A. Simmons and Jonathan Fritz, "A Possible Neuronal Basis for Representation of Acoustic Scenes in Auditory Cortex of the Big Brown Bat," Nature 364 (August 12, 1993), pp. 620-623, and the references cited there. Nevertheless, the attempts of Dennett, op. cit., pp. 441-448, and Kathleen Akins, "What Is It Like to Be Myopic and Boring?" in Bo Dahlbom, ed., Dennett and His Critics (Oxford: Blackwell, 1993), to imagine some of what it is like to be a bat, while interesting, are because of the gaps beside the point.

an argument of this sort. Although Jackson's version fails, I will produce a version which succeeds.

Frank Jackson's has probably been the most discussed version of the Knowledge Argument and the most discussed recent argument for property dualism.[13] In two papers,[14] he invites us to consider the unusual experiences of superneuroscientist Mary. Through her neurological research, Mary is as knowledgeable about as much of the physical world as you like--on one version, about every physical aspect of human beings, on another, about the entire physical world. Let me call her neuro-omniscient. Confined throughout her life to a black-and-white room and with access to the outside world only through black-and-white television, Mary has never experienced anything red. On her release, she finally does. On the basis of this story, Jackson makes the following argument.

---

13. Besides Nagel's and Jackson's versions, see also the version by Howard Robinson in his Matter and Sense (Cambridge: Cambridge University Press, 1982), pp. 4-5, and his "Introduction" and his "The Anti-Materialist Strategy and the 'Knowledge Argument,'" in Howard Robinson, ed., Objections to Physicalism (Oxford: Oxford University Press, 1993), pp. 17-18 and 159, respectively. See also the version by John Foster in his The Immaterial Self (London: Routledge, 1991), p. 64.

14. Frank Jackson, "Epiphenomenal Qualia," Philosophical Quarterly 32 (April 1982); and "What Mary Didn't Know," Journal of Philosophy 83 (May 1986), p. 291.

Jackson's Knowledge Argument

(1) Mary (before her release) knows everything physical and functional there is to know about other people.

(2) Mary (before her release) does not know everything there is to know about other people (because she learns something about them on her release).

---

(3) There are truths about other people (and herself) which escape the physicalist-functionalist story.[15]

Henceforth, I will call the argument from the two premises above to the conclusion "Jackson's version." Although I do not accept Jackson's argument myself, I believe that the arguments previously advanced against him have for the most part missed their marks. Before setting out my own objections in section five of this chapter and my alternative defense of the Knowledge Argument in Chapter Five, I will review some of these counterarguments to Jackson's version in next two sections of the present chapter and show why they fail.[16]

---

15. Jackson, "What Mary Didn't Know," op. cit., p. 293. I have added the references to functionalism. While functionalism is part of Jackson's target, he does not make that explicit in the argument I quote here.

16. By contrast, see Robert Van Gulick, "Understanding the Phenomenal Mind: Are We All Just Armadillos?," in Martin Davies and Glyn Humphreys, eds., Consciousness: Psychological and Philosophical Essays (Oxford: Blackwell, 1993), pp. 138-142. Van Gulick reviews many of these counterarguments against Jackson and endorses most of those he reviews.

III. What Mary Can Figure Out and Imagine

The conclusions that follow from the Knowledge Argument--that qualia exist and that they are distinct from physical and functional properties--are inconsistent with the project Dan Dennett defends in his much-discussed book Consciousness Explained. That much is clear. "Are qualia functionally definable?" Dennett asks rhetorically at one point. "No, because there are no such properties as qualia.... Or, yes, because if you really understood everything about the functioning of the nervous system, you'd understand everything about the properties people are actually talking about when they claim to be talking about their qualia."[17]

Dennett's counterargument against Jackson depends upon speculation about Mary's powers of imagination. Like Paul Churchland whom he cites approvingly, he argues that Jackson underestimates the extent of Mary's knowledge and the cognitive powers it gives her.

The counterargument seems to be as follows. If Mary knows everything physical and functional there is to know about other people, then she can at least figure out or imagine what it is like to see chromatic color. But if she can figure out or imagine what it is like to see chromatic

---

17. Daniel Dennett, op. cit., pp. 459-460.

color, she knows what it is like to do so. And since, by hypothesis, she knows everything physical and functional there is to know about other people, then she knows what it is like to see red. Thus, Jackson's premises conflict. Endorsing the first requires giving up the second.

Dennett and Churchland have not had many followers. But at least some of the reluctance to join them in this position has grown out of a mistaken view of what their reductionist position requires. Many critics of the Knowledge Argument would argue that Churchland and Dennett have taken on an unnecessary burden and that the physicalist can settle for much less. In this, they exploit a very natural first reaction to the argument. The idea is this. The physicalist is committed to the view that my having experience is just another physical fact which can be described in paradigmatically physical terms. However, the physicalist, it may seem, is not committed to the view that my <u>knowing</u> all the physical facts about a certain kind of experience will <u>give me</u> experience of that kind. After all, things do not often come into existence simply in virtue of my knowing the principles underlying them.

According to Joseph Levine, the physicalist should not reject, as Churchland and Dennett do, but <u>embrace</u> the idea that Mary cannot imagine or figure out what it is like to see red. "After all," writes Levine, "in order to know what it's like to occupy a state one has actually to occupy it!"

From this general principle it follows for Mary that she "can know which physical (or functional) description a mental state satisfies without knowing what it's like to occupy that state."[18] Levine's argument is that "all Mary's new knowledge amounts to is her new experience," and that it is thus open to the physicalist to hold that this is just a different way of knowing the same thing.

But Levine appears to equivocate. It is almost tautological that in order to know what it's like to occupy a token state one has actually to occupy it. Every token state will have its peculiarities and one will not fully know what occupying any given one is like until one has done so.[19] But Mary's knowing what it is like to have the specific experience she has on her release does not exhaust her red-related knowledge of what it is like, since she also comes to have general knowledge of what it is like to see red, knowledge of what it's like to occupy states of a type. When applied to knowledge of types rather than tokens,

---

18. Joseph Levine, "On Leaving Out What It's Like," in Davies and Humphreys, op. cit., p. 125.

19. In "What Is It Like to Be a Bat?" (op. cit., p. 170), Thomas Nagel claims that it is "beyond our ability to conceive" the "specific subjective character" of the echolocating experiences of bats. In criticism, Owen Flanagan (op. cit., p. 103), remarks that this is a general other-minds problem: "If conceiving of the specific subjective character of the experiences of another means having the experiences exactly as the experiencer has them, then this never happens." But, of course, this is not just a problem about understanding other minds: the experiencer does not herself have the experiences as she has them until she has them.

Levine's principle is not tautological but false.

It would follow, for instance, that since nobody has ever occupied the state of seeing a unicorn or a golden mountain nobody knows what it is like to occupy that state. Not even the most extreme of classical empiricists held that view. According to Hume, we can create complex ideas out of simple ones and can visualize unicorns and golden mountains even though we have never seen such things. And by some other means, we even know what it is like to see shades of blue we have never been exposed to.[20]

Perhaps Levine means to restrict the scope of the principle to this: that in order to know what it's like to occupy states of a simple qualitative type one actually has to occupy one. This is not tautological either, but while it may be true, it is not obviously so, and it would beg the question simply to assume it to be. It would be just what Churchland and Dennett deny.

How could they deny this? It may be helpful to consider a possible analogy. What is like to ride a roller coaster? One perhaps need not have actually ridden one to know. For there may be experiences enough like the various aspects of riding a roller coaster that someone with enough experience could piece together what it is like without actually having done it. Although this would leave out

---

20. Hume's Enquiry Concerning Human Understanding, ch. 2.

knowledge of what any specific rides were like, one might still fully know what it is like in a general way. It is such general knowledge as this of what it is like that Dennett believes Mary to be able to figure out in virtue of her complete knowledge of the physical-functional aspects of human beings. He might agree with Levine's physicalist that she learns nothing on her release but would differ in holding there to be an aspect of her knowledge of what it's like beyond her unique knowledge of the experience: her ability to conceptualize it, to place it in a type. She learns nothing because she can figure out ahead of her release what it's like in this general way, just like she might figure out what it's like to ride the roller coaster.

Now let me single out two of the premises of the Dennett-Churchland counterargument for closer scrutiny.

> *(Neuro-omniscience to Imaginability.)* If Mary knows everything physical there is to know about other people, then she can at least figure out or imagine what it is like to see chromatic color.
>
> *(Imaginability to Knowledge.)* If she can figure out or imagine what it is like to see chromatic color, she knows what it is like to do so.

Both these premises are crucial to Dennett's counterargument, and at least one is false. Let me look at each in turn.

The "Neuro-omniscience to Imaginability" Premise.

Dennett argues that Jackson has given no reason for thinking that, even if Mary indeed has all the neurophysiological knowledge Jackson gives her, she will be surprised when shown a blue object. From this he concludes that Jackson has not shown Mary to have learned anything. Dennett assumes that Mary knows what it is like to see black and white (and presumably gray) objects; the differences between an object's color and properties like its glossiness and luminance; and "precisely which effects--described in neurophysiological terms--each particular color will have on her nervous system." Thus, he writes, the only remaining task for her is to "figure out" how to identify "those neurophysiological effects 'from the inside.'" He suggests this to be possible by her "figuring out tricky ways in which she would be able to tell that some color, whatever it is, is not yellow, or not red" by means of "noting some salient and specific reaction that her brain would have only for yellow or only for red." In this way, she could gain "a little entry into her color space," and from there "leverage her way to complete advance knowledge."[21]

Dennett, however, fails to make it plausible that Mary knows the entirety of what it is like to see chromatic color. Given her wide knowledge, Mary will know most of the effects on somebody of seeing a normal banana. Some of these effects will manifest themselves in thoughts and

---

21. Dennett, op. cit., p. 399.

beliefs of hers I will label "nonchromatic." By that I mean all her thoughts and beliefs except those by which she attributes to herself and others what it is like to visually experience chromatic color. Extending her expertise to bananas, Mary will know all the nonqualitative effects of seeing a normal banana. She may know enough of them "from the inside," to use Dennett's phrase--for instance, through the nonchromatic thoughts she has about the banana--for her to be able to tell that she is seeing something aberrant when she is shown a blue banana. Let me for the moment accept several of Dennett's suppositions about Mary's knowledge. She knows in advance that there is a way things appear, whatever it is, which people label "blue." She knows in advance the thoughts she would have on seeing something appearing this way. She knows on seeing the banana that she is having those thoughts, and that those thoughts are sufficient for her to know that the banana is blue. On the basis of such knowledge, let us say that she recognizes the banana as blue. I will even concede to Dennett that having the recognitional ability to do all this would be sufficient for Mary to know what it is like to see blue.[22] Still, it would not follow that Mary knows

---

22. In his "What Experience Teaches," in William G. Lycan, ed., Mind and Cognition (Oxford: Oxford University Press, 1990), p. 516, David Lewis claims that according to his Ability Hypothesis "knowing what an experience is like just is the possession of these abilities to remember, imagine and recognize." In his Metaphysics of Consciousness (London: Routledge, 1991), pp. 157-158, William Seager

everything. For even though she might have the recognitional knowledge of what it is like in advance of seeing red, she still might lack what we could call imaginative knowledge of what it is like. Having this requires the ability in advance of seeing and recognizing red to anticipate seeing it by calling to mind something which seems to resemble seeing it.

Dennett does not explicitly discuss imagination, but Churchland, whom Dennett cites favorably, does. Supporting his second premise by Mary's color ignorance commits Jackson, Churchland argues, to "the claim that Mary could not even imagine what the relevant experience would be like." Like Dennett, he contends that Jackson has not "adequately considered how much one might know if, as premise (1) asserts, one knew everything there is to know about the physical brain and the nervous system."

> In particular, suppose that Mary has learned to conceptualize her inner life, even in introspection, in terms of the completed neuroscience we are to imagine. So she does not identify her visual sensations crudely

---

argues that one can know what an experience is like without any of these abilities. In a forthcoming review of Seager in the Canadian Journal of Philosophy, Christopher Hill argues that a recognitional ability is required. I will assume here, contrary to Hill's view, that any of the three abilities--to remember, to imagine or to recognize--is sufficient. Consider, for example, someone who, never having seen anything red, is nevertheless wired neurally to imagine seeing red, although she cannot ever hope to recognize anything as red, let us suppose because of visual difficulties.

> as "a sensation-of-black," "a sensation-of-gray," or "a sensation-of-white"; rather she identifies them more revealingly as various spiking frequencies in the nth layer of the occipital cortex (or whatever). If Mary has the relevant neuroscientific concepts for the sensational states at issue (viz., sensations-of-red), but has never yet been in those states, she may well be able to imagine being in the relevant cortical state, and imagine it with substantial success, even in advance of receiving external stimuli that would actually produce it.[23]

But despite all that Churchland supposes about Mary in this science-fiction future, he has not yet given any reason for thinking Mary might have something in her imagination that seems to her to resemble red. What Churchland appears to invent is a possible world in which reference to phenomenal properties has dropped out of the language and has been replaced with reference to objective, public, paradigmatically neurophysiological properties. I will grant that we can imagine such a world. If you were asked in such a world to imagine being in a state characterized physically, one which correlates with sensing red, you could perhaps do so without much effort, if you were the neurophysiologist Mary is and accustomed to characterizing your own occurrent qualitative states in physical terms. But it is a further task to imagine being in a state

---

23. Churchland, op. cit.

conceived of not physically but phenomenally.

Imagine feeling the way one normally does while being in a rapidly dropping roller-coaster car. Can somebody imagine this without having been through the experience of rapidly dropping while riding a roller coaster? In a sense, yes--one need only imagine that one is in a roller-coaster car, rapidly dropping and feeling the way one normally does in such circumstances.[24] Let me call this a case of descriptive imagining--a case of imagining that one satisfies a certain description, "being in a rapidly dropping roller-coaster car and feeling the normal way." But there is a further kind of imagining that seems in order here, which I will call direct imagining--in which one calls to mind something which seems to resemble the feeling of rapidly dropping. Even if one were to do the first kind of imagining, there would still be the task of doing the second kind.

Similarly, we still have reason to think that there is a way in which Mary before her release would be unable to imagine what it is like to see red even if she could also do so in Churchland's way. Labeling it "crude," as Churchland does, does not contradict Jackson's assumption that the normal way is a different way. Only if Churchland can make plausible that this difference in ways of imagining does not entail a difference in what is known does his argument

---

24. I am indebted to Robert Stalnaker for this point.

succeed.

This is perhaps what Churchland intends to do when he suggests that sensations of color might be analogous to musical chords. He suggests that both are "structured sets of elements" and that Mary might be able to imagine red just as musicians gain access to musical chords they have never heard before by constructing them "in auditory imagination."[25] If that were right, then by descriptive imagining alone Mary could literally gain access to something that seems to resemble red and thus could fully know what it is like to see it without actually seeing it. However, there is an obvious disanalogy which Churchland must contend with. The musician has heard the elements out of which musical chords are structured, or he at least has a way of generating the elements out of what he has heard. Even if there is, as Churchland writes, "excellent empirical evidence to suggest that our sensations of color are indeed structured sets of elements" (his emphasis), still Mary, raised since birth away from color, has not experienced enough such "elements" to generate any structures of color-sensation.

The "Imaginability to Knowledge" Premise. But there is a deeper problem with the argument. Suppose, contrary to common sense, that she has access to "elements" out of which she can imaginatively construct color-sensations. Then

---

25. Ibid., pp. 26-27.

consider the case of another superneurophysiologist, Marilyn, who does not. Suppose that Marilyn is blind, although she later learns what it is like to see red when she acquires vision. Why believe that the blind Marilyn, who knows everything physical, has access to elements out of which she could construct what it is like to see red in "visual imagination"? For any putatively structured sensation, it would always seem possible, hypothetically, to come up with someone who (1) masters all the propositions of neurophysiology but also (2) lacks enough raw elements of experience to generate the sensation structures which Churchland supposes to exist. Such a someone might even be Mary herself--say, in some possible world in which she is congenitally blind. In fact, the Knowledge Argument does not require that it is always possible to come up with such a person; one case is enough.

This insight defeats Dennett's counterargument. One might fail to notice that it does by equivocating between the two premises of Dennett's counterargument to Jackson which I have been examining. The equivocation is on the phrase "can imagine." Construe it to mean "<u>can</u> have the ability to imagine," and the "Neuro-omniscience to Imaginability" premise might well be judged true by considerations like Churchland's. But then the "Imaginability to Knowledge" premise is surely false. By the argument above, merely having the ability to imagine red

in some possible world does not entail having any particular knowledge of red in the actual world. Mary can hypothetically have the ability to imagine red in some world, one in which her mental powers are intact, without having knowledge of what it is like in the actual world, let us suppose because of the congenital blindness. On the other hand, construe the phrase to mean "would have the ability to imagine" and the second premise is true: if Mary would actually have the ability to imagine red, then she would know what it is like. But then the first premise is surely false (or at least unsupported by anything Dennett or Churchland argue): neuroscientific omniscience does not alone entail that one would have the ability to imagine red.

For people's powers of imagination vary; rank the powers of imaginatively bringing to mind the sight of red, consider the worst case of such powers in an otherwise normal human being, and select the possible world in which such worst-case powers happen to be Mary's. It is surely plausible that there are worst cases of these powers that are consistent with Jackson's assumption that Mary does not know before her release, and therefore was factually unable to have imagined, what it is like to see red.[26]

---

26. Owen Flanagan (op. cit., p. 104) offers this possibility about how Mary might grasp what it is like to see red without actually having seen it: "Suppose that she discovers a novel way to tweak the red channel. She discovers that staring at a black dot for a minute and then quickly downing a shot of brandy produces red hallucinations." In his Color for Philosophers

A related point can be made about Dennett's "figuring out." Dennett may seem to have improved on Churchland's argument by eliminating the flaw just noted. Mary's vast neuroscientific knowledge may seem to guarantee that she would be able to figure out everything that can be figured out, even if (as I just argued) it does not guarantee that she would be able to imagine everything that can be imagined. Any improvement, however, is illusory. As I claimed before, if figuring out what it is like to see red in Mary's situation does not require imagining it, then figuring out will not be sufficient to gain her complete knowledge of what it is like. Some form of acquaintance with the appearance of red is required and perception of red objects is unavailable to her. If, on the other hand, figuring out does require imagining, then Dennett confronts the same problems I argued to face Churchland.

Clearly Dennett needs more than just that Mary <u>can</u> figure out or imagine. Even if Mary could figure out what

---

(Indianapolis: Hackett, 1988), pp. 91-92, C. L. Hardin offers other ways in which one might do this. But, again, all this is beside the point. A friend of the Knowledge Argument can acknowledge the possibility that in creatures like us the neuroscientific expertise Mary has would <u>enable</u> her to grasp phenomenal red, even without seeing it exemplified in objects. The critic of the Knowledge Argument, however, must take the position that her neuroscientific expertise would not just enable her to do this but would <u>constitute</u> the grasping of phenomenal red, and this is implausible. For it seems easy to imagine a person in Mary's shoes, someone perhaps unlike Mary biologically, who doesn't have the powers of hallucination Flanagan supposes but about whom we would say the things Jackson says of Mary.

it was like from Dennett's reasoning, still she has to figure it out. That means that her knowledge of what it is like to see red is something over and above her neurophysiological knowledge of the factors Jackson and Dennett cite. That is all Jackson needs to reach his conclusion. Dennett's conclusion that she learns nothing new requires that Mary must know what it is like given her physical knowledge--in fact, that her physical knowledge constitutes her knowing what it is like. But then, there is no longer any need for her to figure anything out.

Dennett's position ought to be that the neurophysiological omniscience Jackson assumes of Mary requires that she has already figured out what it is like to see red. But what she must already have figured out includes not just recognitional but also imaginative knowledge, and it is this that makes Dennett's view untenable. Suppose that Mary were able to imagine what it is like to see red due to her neurological knowledge. More is needed. Just as Dennett's argument requires that she must have figured it out, it also requires that she must have been able to imagine it.[27] Only this, on the story Dennett must make, would complete Mary's knowledge.

---

27. If Churchland is correct that Mary can sometimes construct what it is like in her visual imagination, however, then it is not true that Mary must imagine or must have imagined it, no more than she must imagine what it's like to see a golden mountain in order to know what it's like to do so.

Churchland, however, does not argue that she must have been able to imagine it, only that she can have been, and it is hard to see how he could make the stronger claim plausible. Once again, it seems possible for there to be Marys who are poor at imaginatively calling to mind the ways things appear even though they are good, in fact omniscient, at gaining explicitly physical knowledge of things.[28]

---

28. There may well be a further difficulty. Previously I distinguished between recognitional knowledge and imaginative knowledge of what it is like to see red. A point similar to the one I make for imaginative knowledge may well equally be made against Dennett's case for believing that Mary has complete recognitional knowledge. The fact that Dennett requires that Mary, in order to know what it is like to recognize red or blue in advance of seeing chromatic color, "figure out a way of identifying ... neurophysiological effects 'from the inside'" suggests that there is a first-person aspect to the concepts she uses. She thus must correlate what she conceives "from the inside" on the basis of introspection with what her objective neurophysiological theory tells her about the workings of the human nervous system apart from any such correlations. But then the same point works against the recognition argument: Mary may be poor at correlating her "inside" with her "outside" while still knowing all the neurology Jackson supposes. Dennett ought to hold that Mary can recognize red or blue in virtue of her third-person neuroscience alone. If she could do that, she would not need any "tricky ways" of correlating third-person talk with secondary qualities, which she has an impoverished view of, but would instead correlate it with primary qualities, which she conceives normally. Of course, just what the boundary is between primary and secondary qualities is part of what is at issue here, but Dennett's position surely loses much of any initial plausibility it had as the "inside" Mary can use as a basis to figure out the rest shrinks.

IV. Knowing How, Knowing That and Knowing About

I have devoted as much detail to the Dennett-Churchland position as I have because I believe it to be an extremely important one. My counterargument shows that unless there is a defect in the mechanics of the Knowledge Argument or a deep flaw in our common sense about Mary's experiences, then the standard positions about the nature of the mind are untenable. This is true not only of physicalism but of functionalism. The intuitions that Mary will not know what it is like to see red just on the basis of knowing physical properties of herself seem clear-cut. But the intuitions against analytic functionalism ought to be just as strong. The analytic functionalist asserts an a priori connection between mental terms or properties and functional characterizations, but it would seem that Mary might know every functional characterization without knowing what it is like to see red. Functionalists have developed ingenious strategies against the standard anti-functionalist arguments, such as the argument from the possibility of spectrum inversion. They have argued, for example, that our dispositions are intrinsic to our color experiences and that because of that spectrum inversion is impossible.[29] Even

---

29. Dennett argues this, op. cit., pp. 375-389. See also C. L. Hardin, "Reply to Levine," Philosophical Psychology 4 (1991), pp. 41-50.

if they were right about this, Mary's experiences show that functionalism is still *prima facie* untenable, since knowing everything about these dispositions (and any other causal properties you include) seems insufficient for generating knowledge in Mary of what it is like to see red. Again, the truth of functionalism (or physicalism) would require that our common sense about Mary is deeply flawed or that there is a defect in the very strategy of appealing to Mary's experiences.

Consider now a complaint against the Knowledge Argument different from Dennett's and Churchland's. This is that the argument goes wrong when it assumes in the second premise that Mary before release is ignorant of information--something *propositional* or *intentional*. Sometimes called the Ability Hypothesis, this view has it that Mary, before her release, is not ignorant of information but rather lacking in ability--that Mary lacks know-how that she could not get just by obtaining information. On this view, Jackson's premise (2) is unsupported. The Ability Hypothesis plays on the intuition that some of our knowledge--such as Roger Clemens' knowledge of how to throw a 95-mile-per-hour fastball wherever he wants it to go--is not knowledge of some body of *information*, or "knowledge *that* ...," but rather is *ability*, or "knowledge *how*...." This does not mean that Clemens is inarticulate about his fastball--in fact, he has plenty to say about it. It means

rather that his accurate throwing of the fastball is not causally dependent, at least not entirely, on the kind of prior information he might report when speaking it.[30]

Although I share Jackson's view that the Ability Hypothesis is an incorrect assessment of Mary's situation, Jackson's own counterargument to the Ability Hypothesis is unsatisfactory. He contends that since Mary not only gains knowledge of what it is like to see red _for herself_ but comes to know more about the experiences _of others_ as well, it follows that she gains more than abilities. If she were a skeptic about other minds and doubted that she had gained knowledge of others, he argues, she would not be doubting abilities, which "were a known constant throughout."[31] But why can it not be said that, in such a case, what Mary doubts is just another ability--her ability to peer into other minds, since she doubts there are any?

The real problem with the Ability Hypothesis is the intuition underlying it that draws a firm line between "knowing _that_" and "knowing _how_." Sometimes the two forms of knowledge are much closer than the intuition allows. Generally, if Roger Clemens knows _how_ to throw his accurate fastball, then he will have "knowledge _that_" he can report with sentences like, "I know that my accurate fastball is

---

30. See Lawrence Nemirow's review of Thomas Nagel's _Mortal Questions_, _Philosophical Review_ 89 (1980), pp. 475-476; and Lewis, _op. cit._, esp. pp. 514-518.

31. Jackson, _op. cit._, p. 293.

thrown like this," demonstrating with an accurate throw. If I know how to play golf, then generally I have knowledge I can report in forms of words like, "I know that golf is played like this." The main exception to the general claim that "knowledge how" entails "knowledge that" is the kind of case in which I fail to realize that this ability is the ability to play golf or in which Roger Clemens, suffering amnesia, forgets what his pitching ability is for. In such a case the exception is shown by the fact that I can say of Clemens, for example, that he still knows how to throw a fastball for a strike over the inside corner of the plate but he no longer knows that his ability can be described this or any other way. But this exception is irrelevant to the case at hand. It makes no sense to say of someone that she knows how to recognize color but has forgotten that recognizing color is like that. As Brian Loar writes, "Knowing how a state feels is knowing that it feels a certain way."[32] The claim that Mary before her release lacks information--that is, lacks "knowledge that"--and subsequently gains it thus remains untouched, since the Ability Hypothesis does not offer a genuine alternative to it.

---

32. Brian Loar, op. cit., p. 85. This way of assimilating knowing how to knowing that does not commit the fallacy, which Ryle rightly points out, of assuming that intelligent performance requires observance of rules or application of criteria; see The Concept of Mind, op. cit., p. 29.

Now consider an argument along a different line. Churchland asserts that Jackson's argument is "a prima facie case of an argument invalid by reason of equivocation on a critical term." The term he questions is "knows about."

(1) Mary (before her release) knows everything physical and functional there is to know about other people.
(2) Mary (before her release) does not know everything there is to know about other people (because she learns something about them on her release).

---

(3) There are truths about other people (and herself) which escape the physicalist-functionalist story.

Premise (1), he writes, is "plausibly true," given Jackson's story about Mary, "only on the interpretation of 'knows about' that casts the object of knowledge as something propositional, as something adequately expressible in an English sentence." Premise (2) is plausible only on the interpretation casting the object of knowledge "as something nonpropositional, as something inarticulable, as something that is non-truth-valuable."[33]

But are there really two such separate interpretations of "knows about"? Churchland gives us no reason to think there are, besides pointing out two kinds of knowledge. But that is no more reason for thinking "knows about" equivocal

---

33. Paul Churchland, A Neurocomputational Perspective (Cambridge, Mass.: M.I.T. Press, 1989), p. 68.

than the existence of paperbacks and hardbacks is reason for thinking "book" equivocal. Even were I to grant that, if premise (1) is true, then it is true in virtue of Mary's having mastered something propositional and articulable, and that, if premise (2) is true, it is true in virtue of Mary's missing something nonpropositional or inarticulable, Churchland's point would not follow. Why are these not just Mary's having some "knowledge," understood univocally, and lacking same other "knowledge," understood the same univocal way? Why should we believe "Mary" is the subject of a different verb in premise (1) and premise (2)? After all, it is perfectly intelligible to claim that Mary comes to know about other people both every physical characterization true of them and what it is like for them to experience chromatic color. Following Loar's suggestion, we can regiment this claim to read that, for all other people, Mary knows two sorts of one-place open sentences to be true of them: that such-and-such a physical characterization is true of them, and that experiencing such-and-such a chromatic color is for them like that. (Here the demonstratum is a paradigm experience of the such-and-such chromatic color at issue).

Churchland challenges Jackson to provide a univocal interpretation of "knows about" that makes the premises plausibly true at the same time. Churchland constructs his own nonequivocal argument, replacing Jackson's premises with

(1′) and (2′) and Jackson's conclusion with (3′).

Churchland's "Nonequivocal" Argument

(1′) For any knowable x and for any form f of knowledge, if x is about humans and x is physical in character, then Mary knows by f about x.

(2′) There is a knowable x and a form of knowledge f such that x is about humans and Mary does not know by f about x.

(3′) There is a knowable x such that x is about humans and x is not physical in character.[34]

The "nonequivocal" argument is unsound, Churchland tells us, because "there is something about persons (their color sensations, or identically, their coding vectors in their visual pathways), and there is some form of knowledge (an antecedently partitioned prelinguistic taxonomy), such that Mary lacks that form of knowledge of that [physical] aspect of persons." This is supposed to be what it is for her to be unacquainted with what it is like to see red. Of course, she purportedly has another form of knowledge, knowledge by description, of this same physical aspect of persons. Thus, premise (2′) is true and premise (1′) is false.

Initially, Churchland's supposition that Mary has knowledge by acquaintance and lacks knowledge by description of one and the same thing may not seem troubling. After

---

34. For ease of exposition I replace Churchland's expression "knows(f) about" with the more conventional "knows by f about."

all, consider a characterization of the following sort.

(4) To notice that a tomato is red has <u>that</u> property.[35]

At least initially, one can make sense of the claim that Mary knows (4) to be true by description although not by acquaintance: to specify the referent of the demonstrative, one goes on to give a detailed physio-anatomical description. And one might think that Mary could know one and the same thing by acquaintance (instead of by description) by specifying the referent of the demonstrative through producing a paradigmatic experience of seeing red. If the first demonstration could pick out the same state as the second demonstration, then it would be possible to say that the same object of knowledge was picked out by different forms of knowledge in virtue of the two demonstrations.

But it is not enough for Churchland to produce only a single case like this. Churchland's account of why the "nonequivocal" argument is unsound requires that <u>everything</u> knowable by acquaintance be knowable, in principle, by description. What, then, is it like to know (5) by

---

35. As Michael Tye does in his "The Subjective Qualities of Experience," <u>Mind</u> 98 (July 1986), pp. 12-13, footnote 19. Tye's example is closely related, and his account of it, a response to Horgan (see below), is much like the one that I entertain here.

description?

(5) To be in <u>that</u> state (demonstrated by giving a physio-anatomical description) is to be in <u>that</u> state (demonstrated by a paradigm phenomenal experience).

As a materialist, Churchland must hold that on her release Mary will know (5) to be true. But for Mary to know (5) to be true is not for her to know by a new form of knowledge something she already knew by physio-anatomical description. She must know something new, and not just by a new form of knowledge.

The point is a familiar one since Frege.[36] Let "R" be a referring expression the reference of which I fix with a physio-anatomical description. Let "S" be a referring expression the reference of which I fix by ostending, so to speak, a phenomenal experience as a paradigm. The following two statements have different cognitive significance.

(6) R = R

(6′) R = S

It is a matter of some dispute why (6) and (6′) differ in cognitive significance, but there is no dispute that to know

---

36. See Gottlob Frege, "On Sense and Reference," in P. T. Geach and Max Black, eds., <u>Translations from the Philosophical Writings of Gottlob Frege</u> (Oxford: Blackwell, 1952), p. 56.

(6) true is to know something in some sense empty, while to know (6′) true is to know something substantive. In any case, to know (6)
true and to know (6′) true is to know two different things and not to know the same thing by different forms of knowledge.

Thus, Churchland's first "nonequivocal" premise is false as Churchland claims but for a different reason. It is false that Mary before her release knows every physically characterized item of information by every form of knowledge, for there is a physically characterized item of information--different from any (not, as Churchland claims, the same as at least one) of which she has knowledge--which she does not know by acquaintance.

Still, while Churchland is correct that his argument is unsound, he overlooks a different "nonequivocal" argument which derives the same conclusion (3′) from two true premises. Before her release, Mary knows every item of information characterized physically by at least some form of knowledge, whether by description or by acquaintance. But there is an item of information she does not know by any form of knowledge. Thus, (1") and (2") are both true, and (3′) follows.

A Sound "Nonequivocal" Argument

(1") For any knowable x there is a form f of knowledge such that if x is about humans and x is physical in character, then Mary knows by f about x.

(2") There is a knowable x such that for every form of knowledge f, if x is about humans then it is false that Mary knows by f about x.

(3') There is a knowable x such that x is about humans and x is not physical in character.

V. The Real Problem with Jackson's Conclusion

The real problems with Jackson's argument concern not Jackson's premises but his anti-physicalist conclusion. Some of these are problems first identified by Terence Horgan, and I will build here on his original presentation. If we take Jackson's premises, contrary to Churchland, to be about the presence and absence in Mary of knowledge of different things, we can take these different things to be different items of information. In that case, we have an argument for a dualism of information into paradigmatically physical information and the introspective information Mary comes by on her release. But, as Horgan argues, a dualism of information does not guarantee a dualism of properties, since distinct items of information can be about the same property. Thus, he argues, we have no reason to conclude that physicalism is false because of an excess of

properties.

Horgan illustrates his case by considering two statements: "Superman can fly" and "Clark Kent can fly." These statements express different information even though they predicate the same property of the same individual. Horgan argues that it is similarly open to the physicalist to allow that statements made in language paradigmatically mental (and, thus, in language not paradigmatically physical) express different information from statements that explicitly predicate physical properties and relations of wholly physical entities. For Horgan holds that the mental-language statements still predicate the same physical properties and relations of the same wholly physical entities as do the explicit statements. Horgan illustrates his point with the following statement which I label (7).

(7) Seeing ripe tomatoes has this property.

Assume that Mary, in using (7) to express new knowledge acquired after her release, uses the demonstrative "this property" to designate a color-quale, a phenomenal property, instantiated in experience contemporaneous with her statement. There may be a question, given that Mary has all relevant physical information, how a physicalist can make sense of further information expressed in language which (a) is not paradigmatically physical but (b) predicates physical properties and relations of physical entities. From where

would the further information come? Both statements about Superman, even though they express different information, express physical information after all, constructed of language paradigmatically physical. Any uncontroversial pair of statements Horgan could find to illustrate his point about the intentionality of information would be constructed of language paradigmatically physical or topic-neutral. Doesn't it just beg the question simply to assert, without argument, that the same point can be made across modalities of information--that two statements expressing, respectively, mental and physical information can predicate the same properties and relations of the same ontities? But according to Horgan, it is not merely open to the physicalist to assert that the entities referred to and the properties and relations expressed by (7) are physical--it is true.

> Sentence [(7)] expresses new information because Mary has a new perspective on phenomenal redness: viz., the first-person ostensive perspective. Her new information is about the phenomenal color-property as experienced. Thus she could not have had this information prior to undergoing relevant experience herself. But these facts are compatible with Physicalism; there is no need to suppose that when she acquires experiential awareness of phenomenal redness, she thereby comes into contact with a property distinct from those already countenanced in her prior physical

account of human perception.[37]

The account he provides of his assertion that Mary could not have had the information she expresses by (7) until being released from the room is inadequate. Horgan's explanation that one cannot have information about a phenomenal property "as experienced" before experiencing it is false. Before experiencing it, Mary can, at least by being told this, know of the phenomenal property of seeing ripe tomatoes--the "phenomenal property as experienced"--that it is like the phenomenal property of seeing bright sunsets. What Horgan should have written is that Mary cannot have knowledge by acquaintance before having the relevant experience--knowledge, that is, by what he calls the "first-person ostensive perspective." Mary can know that the phenomenal property as experienced has certain properties but she cannot know, if Horgan is right, what it is like to be acquainted with the phenomenal property as experienced before experiencing it. Or, to put it differently, Mary cannot, before experiencing it, have the first-person knowledge of the phenomenal property she expresses by directly referring to the property, if Horgan's account is correct.

---

37. Terence Horgan, "Jackson on Physical Information and Qualia," Philosophical Quarterly 34 (April 1984), pp. 150-151. Similar points are made by Flanagan, op. cit., pp. 98-99.

It seems to me that Jackson has no reply to this position. In fact, I do not see any possible reply within the argumentative strategy behind Jackson's version of the Knowledge Argument. Where exactly, then, do the formal versions of Jackson's argument which I have discussed--one with (1) and (2) as premises, the other fuller representation with (1") and (2") as premises, both with Jackson's (3) as the conclusion--go wrong? As I have argued already, not in the premises.[38] Thus, the problem must lie in the inference to the conclusion. It should be obvious that both versions are enthymemic. Let me focus on the fuller version from (1") and (2") to (3).

(1") For any knowable x there is a form f of knowledge such that if x is about humans and x is physical in character, then Mary knows by f about x.

(2") There is a knowable x such that for every form of knowledge f, if x is about humans then it is false that Mary knows by f about x.

Ergo, (3) there are truths about other people (and herself) which escape the physicalist story.

The terms "truths about other people" and "the physicalist story" appear only in the conclusion, not in the premises. Thus, if a conclusion is to be derived from (1") and (2")

---

38. Flanagan, op. cit., p. 99, makes a mistake about this. He insists that the error lies in premise (1), but elsewhere he accedes to Jackson's assumption that Mary knows the truth of every relevant statement that is explicitly physical, which is all Jackson says (ibid.) he means by (1).

using only principles of logic, either (a) it must be different from (3), or (b) one or more further premises must be added to the argument.

First, consider option (a), that of altering the conclusion. The strongest conclusion that can be derived from (1") and (2") on the basis of principles of logic alone is Churchland's (3').

> (3') There is a knowable x such that x is about humans and x is not physical in character.

But on its face, (3') is not strong enough to accomplish Jackson's anti-physicalist aims. It is open to a critic such as Horgan to insist that although phenomenal knowledge is not characterized physically--that is, is not explicitly or paradigmatically physical--it does not provide contact with any nonphysical property. If the critic is correct, then there is a gulf between what follows from logic alone, (3'), and Jackson's anti-physicalist conclusion (3).

Thus, Jackson is left with option (b). In fact, Jackson means the argument to have a further premise.[39]

> The Further Premise, Version One
> If physicalism is true, then if you know everything expressed or expressible in explicitly physical language, you know everything.

Jackson's fullest argument for Version One goes as follows.

---

39. Jackson, op. cit., p. 291.

> Physicalism is not the noncontroversial thesis that the actual world is largely physical, but the challenging thesis that it is entirely physical. This is why physicalists must hold that complete physical knowledge is complete knowledge simpliciter. For suppose it is not complete: then our world must differ from a world, W(P), for which it is complete, and the difference must be in nonphysical facts; for our world and W(P) agree in all matters physical. Hence, physicalism would be false at our world [though contingently so, for it would be true at W(P)].[40]

From what Jackson writes here, it is evident that he means by Version One something like the following.

> Version One, Jackson's Interpretation
> If physicalism is true, then if you know everything expressed or expressible in explicitly physical language, you have all states of knowledge you could have.

Jackson's error should now be apparent. The reductio does not succeed. Jackson asks us to suppose, contrary to Version One, that physicalism is true but that, as I have already argued, complete physical knowledge is not complete knowledge simpliciter. Contrary to what Jackson writes, it is not then true that our world must differ from a world W(P) which agrees in all matters physical with our world and in which complete physical knowledge is complete knowledge.

---

40. Ibid.

There is no world, W(P), even moderately similar to ours physically in which complete physical knowledge is complete knowledge. For in any world containing "knowledge how" not characterized physically--at least part of Clemens' knowledge of how to throw his fastball, for example, is not physically characterized knowledge--complete physical knowledge comes up short of complete knowledge. Even after Clemens might learn everything explicitly physical about throwing a 95-mile-per-hour fastball accurately, there would still be something further for him to learn--the doing of it. This would be so whether or not such "knowledge how" is a form of "knowledge that." It begs the question then to suppose that this incompleteness of physical knowledge is incompatible with physicalism. No reason has been given yet to suppose that Clemens' knowledge of how to throw the fastball is not simply some physical state of his brain; to suppose otherwise would be to assume the dualism that the advocate of the Knowledge Argument seeks to demonstrate.

A recent defense of Jackson's argument might be thought to overcome this difficulty, but actually it stumbles in a similar way. Geoffrey Madell argues that criticisms of Jackson which rely on a distinction between knowledge by description and knowledge by acquaintance, as do the writings of the Churchlands and others, are self-defeating, since "it must be clear this is not a distinction which is open to the physicalist to make." This distinction, he

argues, "amounts to the claim that knowledge must be grounded in something which eludes description." Whatever such a thing is, Madell writes, the physicalist must hold that "it is some configuration of physical elements, and as such it must be describable. The physicalist cannot therefore accept that even the most complete physical description one could give would nevertheless fail to capture an aspect of what is described."[41] Nor, therefore, if this were right, could the physicalist even distinguish a form of knowledge by acquaintance. This criticism, if it is sound, is general in its impact, defeating Horgan and Lewis as well. But it is not sound. The Churchlands et al. accept that states of knowledge by acquaintance are configurations of physical elements and thus "describable," to use Madell's term. The contrast with knowledge by description is made not on the basis of differences in the physical describability of the havings of the two forms of knowledge. Having knowledge by description and having knowledge by acquaintance, according to physicalists like the Churchlands, are both physically describable. The contrast with knowledge by description is instead made on the basis of differences in the ways in which the two forms of knowledge represent their objects. Roughly, we can say

---

41. Geoffrey Madell, "Neurophilosophy: A Principled Sceptic's Response," Inquiry 29 (1986), p. 155. See also his Mind and Materialism (Edinburgh: Edinburgh University Press, 1988), pp. 80-83.

that knowledge by description represents objects of knowledge in virtue of definite descriptions, whereas representations employed in knowledge by acquaintance refer directly and express singular propositions. Or, to use the Churchlands' different formulation, knowledge by description is mastery of a "set of descriptive propositions," knowledge by acquaintance is "prelinguistic representation." In asserting that knowledge by acquaintance must be "describable," Madell fails to distinguish between the assertion that the state itself must be representable in virtue of definite descriptions, with which the Churchlands would agree, and the assertion that the state's style of representation must be that of definite descriptions, which is false. Because it is false, there is a prima facie distinction between the two forms of knowledge.[42] If one forgets about this prima facie distinction and assumes that this exhausts our forms of knowledge, as Madell seems to, then one arrives at Version Two.

---

42. Related forms of criticism have been directed at the Knowledge Argument. Christopher Hill notes, rightly, that two items of knowledge might have different character (in Kaplan's sense--see Chapter Seven), as Mary's items of knowledge do, but have the same content. See his review of Seager, op. cit. Christopher Peacocke notes, rightly, that indexical knowledge can differ from non-indexical knowledge but have the same propositional content; see his "No Resting Place: a Critical Notice of The View from Nowhere, by Thomas Nagel," The Philosophical Review 98 (1989), pp. 70-71.

The Further Premise, Version Two

If physicalism is true, then all knowledge is knowledge by description.

So Madell is also off the mark when he assesses physicalism by Version Two. Nor is it a given--and it is something that neither Jackson nor Madell choose to demonstrate--that the physicalist has any problem with distinguishing knowledge by acquaintance from knowledge by description in virtue of the former state's style of representation.

To this objection to Madell, there is a natural but unsuccessful response. If complete knowledge of physical theory does not give Mary knowledge of what it is like to see red, it will not give anyone else knowledge of what it is to have the state of knowledge of what it is like to see red. The Madell partisan may object that the same deficit that exists in Mary's knowledge will be duplicated in all higher-level knowledge of her knowledge and that the physicalist can never overcome that deficit. This response is unsuccessful because while it may be true that if there is a deficit in Mary's knowledge there will also be one in all higher-order knowledge of it, it begs the question to suppose, without further argument, that the physicalist must overcome that deficit. It ought to be the physicalist position that he or she is no more required to do that than to derive knowledge of how to throw the fastball from complete knowledge of physical theory.

CHAPTER FIVE PROPERTY DUALISM ARGUMENTS

Even though Jackson's argument begs the question and Madell's is unsound, a version of the Knowledge Argument can be constructed that succeeds against the physicalist where these two fail. It is, however, an argument somewhat different in form and inspiration. This final section will be taken up with constructing and defending this alternative version.

In the first section, I set out the main idea behind this successful version of the Knowledge Argument, drawing on the argument for property dualism first discussed in print--and rejected--by J. J. C. Smart. In the second and third sections, I try to set out a version of the argument Smart rejected that is immune to counterexamples. In the fourth section, I complete my defense of the Knowledge Argument in terms of this argument. And in the fifth section, I reply to common-sense objections against both these forms of argument.

I. The Knowledge Argument and the Property Dualism Argument: The Main Idea

I have suggested that some knowledge provides routes to the objects of knowledge distinct from every route provided by knowledge by description and that at least some of these cases, too, are entirely compatible with physicalism. These

are the cases of knowledge by direct reference that I have just discussed in connection with Jackson and Madell. I shall presently focus on some cases of knowledge by direct reference which, by contrast with those just discussed, appear not to be compatible with physicalism at all.

The point that I shall make is a familiar one. In a slightly different form, it goes back at least to Smart's 1959 essay "Sensations and Brain Processes," and Smart contends there it originated with Max Black. Black's point was something like this. If singular terms are to pick out referents, they must do so in virtue of properties of those referents. If the concepts expressed by two singular terms cannot be known a priori to co-refer (whether it is because they do not co-refer or because they co-refer a posteriori), then the singular terms must pick out their referents in virtue of different properties. Concepts expressed by singular terms referring to things paradigmatically mental cannot be known a priori to co-refer with concepts expressed by singular terms referring to things paradigmatically physical. It follows, according to this argument, that they refer in virtue of different properties--it follows, that is, that mental properties are not physical properties and that physicalism is thus false. I will call this argument a Property Dualism Argument.[1]

---

1. The clearest and fullest statement of the objection is that of Stephen White, "Curse of the Qualia," Synthese 68 (1986), pp. 351-353. Labeling it with the name "the

By illustration, consider first a nonmental case. The definite descriptions "the 41th President of the United States" and "the last governor of Arkansas" refer to the same person, Bill Clinton. Each description refers to Clinton in virtue of properties of him: he satisfies the first description in virtue of being the 41th President of the United States and he satisfies the second description in virtue of having been the last governor of Arkansas before the current one. Since the concepts expressed by these descriptions cannot be known a priori to co-refer (they are known to co-refer, of course, but only a posteriori) they refer in virtue of different properties of him. Thus, I will say that they follow different routes to the referent. In this way they differ from the descriptions "the 41th President of the United States" and "the President following the 40th President of the United States." Here, the concepts expressed by the two descriptions can be known a priori to co-refer and pick out Clinton in virtue of the same properties of him.

Nothing so far requires us to abandon physicalism about persons, presidents or Arkansas governors. It is consistent with what I have written so far that every one of the

---

Property Dualism Argument" is due to White. It was first reported by J. J. C. Smart and linked to Black in Smart's "Sensations and Brain Processes," in V. C. Chappell, ed., The Philosophy of Mind (Englewood Cliffs, N.J.: Prentice-Hall, 1962), pp. 166-167.

properties in virtue of which these three descriptions refer is a physical property and that each of these descriptions follow one of two different physical routes to the referent. Contrast this case with the case of a description which picks out something mental. Assume that both "Clinton's headache at time t" and "Clinton's C-fiber stimulation at t" uniquely refer and that, moreover, they refer to the same thing. Since the concepts expressed by the two descriptions cannot be known a priori to co-refer, it seems reasonable to conclude that they refer in virtue of different properties. But the concept expressed by the first cannot be known a priori to refer to the same thing as any concept expressed by any description referring to things paradigmatically physical (and in this way differs from the concept expressed by "the 41th President of the United States"). Thus, it follows that the properties in virtue of which the mental description refers are distinct from any properties in virtue of which any paradigmatically physical description refers. Unless the mental description has a topic-neutral translation, the properties in virtue of which "Clinton's headache at t" refers are not physical at all but irreducibly mental. And prima facie, the topic-neutral option is unavailable: not only does it seem not be a priori that mental descriptions co-refer with topic-neutral translations, but it also seems, on the basis of other qualia-based counterarguments to functionalism, to be false

on independent grounds.[2]

I shall argue that something like this is going on in the case of the Knowledge Argument. The knowledge Mary had before she had first seen red and the knowledge Mary came to have on first seeing something red represent, I will say, distinct routes to the same thing: the neurological process of seeing red. Because these two forms of knowledge are not linked a priori, there must be independent routes to their common referent in virtue of separate properties of the common referent. The Knowledge Argument is correct because of the soundness of a Property Dualism Argument.

First, however, I want to look more closely at Property Dualism Arguments to see why they work when they do.

II. Black's and White's Versions

Smart represented Black's position this way:

> Now it may be said that if we identify an experience and a brain process and if this identification is, as I hold it is, a contingent or factual one, then the experience must be identified as having some property not logically deducible from the properties whereby we identify the brain process.... If the property of being the author of Waverly is the analogue of the neurophysiological properties of a brain process, what is the analogue of the property of being author of Ivanhoe? There is an inclination to say: "an

---

2. See the references cited in footnote 18.

irreducible, emergent, introspective property."[3]

Why are psychophysical identities assumed by Black and Smart to be contingent? It is because they are identities a posteriori, discoverable by science. Black is assuming as premises to the argument principles of the following forms (where "A" and "B" are placeholders for singular terms).

> First form. For all A and B, if it is not true a priori that A = B, then it is false or contingently true that A = B.
>
> Second form. For all A and B, if it is false or contingently true that A = B, then being A is a different property from being B.

The remainder of the argument must go something like this. The statement 'A = my having pain at t' is a posteriori, if true, for any substitution of a singular term for the placeholder 'A' that refers in virtue of paradigmatically physical properties. Thus, being a case of my having of pain at t either is, or at least involves having, a nonphysical, irreducibly mental property.

There are two difficulties with Black's argument on Smart's version. The obvious one, the one which got Smart's attention, was the inference that experiences have irreducibly mental properties. All this argument shows, if

---

3. J. J. C. Smart, Philosophy and Scientific Realism (London: Routledge and Kegan Paul, 1963), p. 94.

anything, is that experiences have nonphysical properties. It does not yet show that all nonphysical properties are irreducibly mental; it is at least conceivable that some may be "topic neutral" between the mental and the physical. A weaker conclusion is thus called for.

The second problem Smart did not see. It is the idea that psychophysical identities are contingent. This is, at best, controversial. The argument provides no support for the idea, since not all statements of the first form are true. There are identity statements which are not a priori but which are neither false nor contingent, such as the fact that heat is mean kinetic energy, which is necessarily true but known a posteriori.[4]

A different and more subtle presentation of this kind

---

4. Block once suggested (although he apparently no longer believes this) that Black's argument depends on the existence of mental objects and that it can be escaped by replacing reference to mental objects with reference to mental events, in a manner such as I used in sec. III of Chapter Two. See Ned Block, "What Is Functionalism?", in Ned Block, ed., Readings in the Philosophy of Psychology, vol. 1 (Cambridge, Ma.: Harvard University Press, 1980), p. 182. On this point, Block cites Jaegwon Kim, "Phenomenal Properties, Psychophysical Laws, and the Identity Theory," Monist (1972), pp. 177-192. It may be true that there is no need to posit a mental property of sharpness, which no brain state would seem to have, if we reject mental realism about pains and confine our ontological commitments to the havings of pains. Those can be identified with the havings of brain states without our dangling mental properties of sharpness. Kim and Block are right in this. But it does not follow that all dangling of mental properties can be dispensed with this way. For it remains true, prima facie, once we have ascended to realism about havings, that we identify events as havings of mental states differently--by different routes--from the ways we identify them as havings of physical states.

of argument--one that escapes these difficulties--is made by Stephen White.[5] White writes the following.

> We are assuming, for simplicity, that a person's qualitative state of pain at t, say Smith's, is identical with a physical state, say Smith's brain state X at t. Even if this is the case, however, not only do the sense of the expression 'Smith's pain at t' and the sense of the expression 'Smith's brain state X at t' differ, but the fact that they are coreferential cannot be established on a priori grounds. Thus there must be different properties of Smith's pain (i.e., Smith's brain state X) in virtue of which it is the referent of both terms.
>
> The general principle is that if two expressions refer to the same object, and this fact cannot be established a priori, they do so in virtue of different routes to the referent provided by different modes of presentation of that referent.... [T]he natural candidates for these modes of presentation are properties....
>
> Let us stipulate that a property which is neither physical nor mental is topic neutral. Since there is no physicalistic description that one could plausibly suppose to be coreferential a priori with an expression like 'Smith's pain at t', no physical property of a pain (i.e., a brain state of type X) could provide the route by which it was picked out by such an expression. Thus we are faced with a choice between topic neutral

---

5. Stephen White, op. cit. Other presentations of this kind of argument appear in Richard Rorty, "Incorrigibility as the Mark of the Mental," Journal of Philosophy (June 25, 1970), p. 399; Ned Block, op. cit., pp. 179, 182; and William Lycan, Consciousness (Cambridge, Mass.: M.I.T. Press, 1987), p. 9.

and mental properties....

> This argument ... shows that unless there are topic neutral expressions with which mentalistic descriptions of particular pains are coreferential _a priori_, we are forced to acknowledge the existence of [irreducibly] mental properties."

White depends on a premise something like the following one. For a premise with a slightly different form but a very similar role, Brian Loar uses the label, the Semantic Premise, and I will borrow that label here.[6]

> The Semantic Premise, Version One
> For all referring expressions R_1 and R_2, if R_1 and R_2 are coreferential but not known to be _a priori_, then there exists a property ρ by which we pick out the referent of R_1, which is distinct from any the properties by which we pick out the referent of R_2.

Actually the Semantic Premise is derived from two separate further principles.

One of them is found in Frege and is closely connected to the Fregean point about cognitive significance used in Chapter Four against Churchland. It can be put as follows.

> The Fregean Premise
> For all referring expressions R_1 and R_2, if R_1 and R_2 are coreferential but not _a priori_,[7] then R_1 and R_2

---

6. Loar, "Phenomenal States," _op. cit._, p. 83.

7. Notice that this is untrue if we replace the _if_-cluase with "if R_1 and R_2 are not coreferential _a priori_." For then it would be inconsistent with Twin Earth cases. Let us assume that Twin Earth is a world (1) evidentially

pick out their referent by different modes of presentation.

In Chapter Four, I pointed out that statements of the form 'R = R' ordinarily express different pieces of knowledge from statements of the form 'R = S', where 'R' and 'S' differ. This is so, Frege wrote, because the object referred to is ordinarily picked out by two different modes of presentation. When the referring expression 'R' differs from the referring expression 'S' "only ... by means of its shape," he writes, then the cognitive value of the two statements is "essentially equal." In that case, the mode of presentation is the same. A difference arises "only if the difference between the signs corresponds to a difference in the mode of presentation" of the referents.[8]

Frege's account of what modes of presentations themselves are is left a bit sketchy, and I will leave it that way, too. Sometimes I will follow Evans and speak of "ways of thinking" about an object, but we might still

---

like our world in every respect and (2) materially like our world in every respect except in the physical composition of some of its substances. There thus might be a Twin Earth word "water" which (1) would be phonologically identical to our word "water" and which (2) has its reference fixed in virtue of the same superficial properties as those fixing our word but which (3) would be satisfied by a substance with a physical microstructure different from H_2O. Thus, the Twin Earth referring expression "the water in the solar system" and the phonologically identical Earth expression pick out their referents in virtue of the same properties, but they are not coreferential. Of course, Frege was not aware of examples of this kind.

8. Frege, op. cit., pp. 56-57.

wonder what those are.[9] It is sufficient to say that a mode of presentation is whatever explains the difference in knowledge between knowledge of the form 'R = R' and knowledge of the form 'R = S', where 'R' differs from 'S' in more than shape. Another way to put it is this. Assume that a subject has a belief of the form 'R ϕ's' and disbelieves or withholds belief from a statement of the form 'S ϕ's', where 'R' and 'S' corefer. We might say that a mode of presentation is whatever explains the subject's difference in attitude without attributing irrationality to the subject.

Frege allows that even knowledge a priori of the form 'R = S' such as we find in mathematics will use different modes of presentation, so long as it is not of the form 'R = R'. Since pieces of knowledge a posteriori of the first form will always differ from knowledge of the form 'R = R', according to Frege's account, the a posteriori will always introduce distinct modes of presentation.

The argument needs another principle beyond the Fregean Premise to derive the Semantic Premise. Only with a further principle can there be a property difference. That principle, which White does not spell out, might go like this.

---

9. Gareth Evans, The Varieties of Reference (Oxford: Oxford University Press, 1982), pp. 14-22.

> The Property Difference Premise, Version One
> For all M_1, M_2, if it is true but not a priori that M_1 and M_2 are different modes of presentation of some object Ó, then there exists a property Φ by which M_1 is a mode of presentation of Ó distinct from any property by which M_2 is a mode of presentation of Ó.

White's argument for this assumes that if there were two modes of presentation of some object not linked a priori, then there is a possible world in which subjects epistemically identical to us might find the modes presenting different objects. But if there are different objects, they must be picked out in virtue of distinct properties, which in the actual world belong to the same object. The two modes of presentation associated with "heat" and "mean kinetic energy," for example, can come apart in a world epistemically like ours and lead to two different things. That means that in the actual world heat has distinct properties in virtue of which the different things in the other world could be picked out separately.

The remainder of this version of the Property Dualism Argument, then, goes like this. No physicalistic description is coreferential a priori with the term "my having pain at t." But if the term "my having pain at t" picks out its referent in virtue of properties of the referent, though not physical ones, then it picks it out in virtue of either mental or topic neutral properties. Thus,

if there is no topic-neutral description coreferential a priori with the term "my having pain at t", I pick out the expression's referent in virtue of irreducibly mental properties.

This version of the Property Dualism Argument does not have the difficulties I pointed out in Smart's account of Black's argument. It follows Smart in recognizing that the initial argument leaps too quickly to the conclusion that nonphysical properties are irreducibly mental, since there are "topic neutral" properties that are nonphysical but not irreducibly mental. Further, Smart's contingent-identity premises are replaced with the one Semantic Premise. Thus, by this modification, there may be singular terms not linked a priori which are coreferential by necessity, yet pick out their common referent by way of distinct properties. This is enough for the property dualism argument Black intends.

III. Direct Reference and the Property Dualism Argument

Now, a physicalist line of counterargument to the Property Dualism Argument goes this way. Why, it might be objected, must one of the independent Fregean routes to mental pieces of the world be by way of nonphysical properties? Why can't the mental and physical routes be distinct but both physical, reaching their referents via mental and physical properties which are distinct but both

physical properties of the world? Why isn't it open to the physicalist to claim that there *is* a physicalistic description coreferential *a priori* with the referring expression "my having pain at *t*"--namely, "my having pain at *t*"?

The objector may grant that "my having pain at *t*" is not a *paradigmatically* physicalistic expression; it does not *appear* to be a physicalistic description. But the objector argues that it does not follow from the expression's not being coreferential with a paradigmatically physicalistic expression that it is not coreferential *a priori* with any physicalistic expression at all. The proponent of the Property Dualism Argument, according to the objector, neglects two possibilities. One is that paradigmatically mentalistic expressions are a species of physicalistic expressions (though not of paradigmatically physicalistic ones). The other is that mental properties are a species of physical properties. Neglecting these possibilities, the argument begs the question, since the absence of these possibility is, in effect, what the argument tries to show.

The objector can also point to paradigmatic property identities from the physical sciences--such as the identity of heat and mean kinetic energy--as models of how mental properties might just *be* physical properties.

The reply to this objection is that the mental route cannot be by way of paradigmatic physical properties,

whether they be properties identified by theoretical science or by more familiar means. We pick out mental pieces of the world in ways distinct from any of the ways paradigmatic to our picking out physical pieces. The point of the Knowledge Argument can therefore be made with respect to any paradigmatically physical properties you choose. If we try to suppose that the mental properties just are physical, the problem remains. To suppose mental and physical properties identical, we require two separate kinds of routes to them via distinct sets of properties, one set irreducibly mental. Thus, even if "pain" and "C-fiber stimulation" were coreferential on the model of "heat" and "mean kinetic energy," we would still expect what Loar calls "higher order reference-fixers" to provide separate routes to their common referent by way of separate properties.[10]

There is, however, an inadequacy in the account so far. The referential potential of descriptions like "Clinton's headache at t" actually depends on the referential potential of expressions that refer directly--noninferentially, without the mediation of individual concepts or Fregean senses or satisfaction conditions or any other mediating sort of thing. Thus, if the description "Clinton's headache at t" is to pick out a headache then there must be some device for picking out such things[11] directly, as Clinton

---

10. Loar, op. cit., pp. 83-84.

11. Although not necessarily Clinton's headache at t--perhaps Clinton has been struck dumb at t.

himself does in focussing on the feeling running from his left temple to his right and thinking to himself, "This hurts." Otherwise, there would always be, when somebody used a description to pick out something phenomenal, the further question of what that was or of what that was like.

Now, a natural objection to the kind of argument for irreducibly mental properties just advanced is one that parallels Horgan's reply to the Knowledge Argument. Horgan argued that it was conceivable that referring expressions of separate types could refer to the same property. This new objection is that they also might refer by way of either the same properties or no properties of the referent at all. Thus, the two descriptions, mental and physical, would differ not in the categories of properties they refer in virtue of but in the two routes their referring takes, both of which could be wholly physical. It is natural to think that some phase of the routes some singular terms take in referring to qualitative mental states is a direct reference. This will be true of both phenomenal-concept terms like "pain" or "headache" and demonstratives like the grammatical subject of the statement "This hurts." On the other hand, no phase of the routes theoretical singular terms of, say, neurophysiology take in referring to physical states is ordinarily a direct reference. The objector thus asserts that it begs the question to suppose that these routes cannot be wholly physical. The objector explains

that we can be misled to believe that two mutually exclusive sorts of properties exist by the existence of the two very different sorts of routes to the referent.

The objector has located a flaw in even this latest version of the Semantic Premise. So long as two referring expressions represent a referent by different styles of reference, they may pick out the referent in virtue of the same properties of it. Say I look down and to the left, identifying the red pen sitting here, and utter to myself a sentence of the form, "That is D," where the D-position is filled by a definite description. Suppose that the demonstrative picks out the pen perceptually, in virtue of a set of the pen's properties ρ_i through ρ_j. Now let the definite description be the expression that makes explicit these properties ρ_i through ρ_j and picks the pen out in virtue of them. Given normal human ways of knowing such things, it might be *a posteriori* that *that*--the pen--was the thing jointly satisfying the set of properties. Even though I pick it out demonstratively by those properties of it, I do not know explicitly that I do, as I would if the reference were to occur by way of senses or satisfaction conditions. It might be by way of a perceptual gestalt.

Thus, it might be a substantive item of information and not *a priori* that the thing I pick out has them. Still, it would be untrue that the two expressions picked out the common referent in virtue of distinct properties. It might

seem because of the way I pick out the pen perceptually that there could be an epistemically identical situation in which a counterpart definite description could fail to pick out the same thing as a counterpart to my demonstrative, but this is an illusion.

This is an unusual case. Yet the objector's point is well-taken. A wholly physical creature could make separate references, linked a posteriori, to some aspect of itself not in virtue of distinct properties of the aspect but rather in virtue of distinct styles of reference to it--demonstrative, say, versus descriptive. Reference to it might be fixed by a higher-order reference fixer which exploits the same properties of it in picking it out that the creature does explicitly in picking it out descriptively.

While the objector is correct about this general point, however, the case of qualitative states is special. While it may be true that demonstrative reference to qualitative states does not depend upon Fregean senses or intervening concepts or any other mediating entities, it does not follow that we can understand it without appeal to phenomenal properties. Not only are phenomenal properties required to make sense of direct demonstrative reference to qualitative states but irreducibly mental properties are prima facie required.

The success of demonstrative reference depends upon the

demonstratum's being picked out for demonstrator and audience by a mode or manner of presentation. Following a standard account of how this happens for demonstrative reference, I suggest that the presentation must have at least three aspects. These are a scene of which the demonstratum is a part, a directing intention on the demonstrator's part for what is to be demonstrated in the scene, and an externalization of this directing intention for conveying it, such as a pointing. Cases where arguably no mode or manner of presentation is needed to fix reference, such as standard uses of the pure indexicals "I," "now" and "here," are not uses of "true demonstratives," to use Kaplan's words. Such cases would seem to be irrelevant anyhow to questions concerning normal demonstrative reference to qualitative states.

In those cases in which demonstrative reference picks out a demonstratum by a mode of presentation, this is possible only in virtue of properties of the demonstratum which the demonstrator indicates to an audience.[12] Demonstrative reference to a public audience would not be

---

12. The term "manner of presentation" and most of the rest of the terminology that appears in this paragraph comes from David Kaplan, "Demonstratives" and "Afterthoughts," in Joseph Almog, John Perry and Howard Wettstein, eds., Themes from Kaplan (New York: Oxford University Press, 1989), pp. 489ff., 514f., 526f. and 582f. My one addition is my technical use of "scene" in place of Kaplan's non-technical use of the term "picture." His use conceals, I think, the fact that the demonstratum is normally demonstrated in virtue of a perspective on it in the world, not indirectly, in virtue of some representation of it.

successful--or would at least be faulty--unless the demonstratum were (1) part of a scene, (2) the object of a directing intention, and (3) the target of an externalization. None of these three conditions could be satisfied except in virtue of properties of the demonstratum. In the case of direct demonstrative reference to a qualitative state,[13] where demonstrator and audience are identical, the properties of the state in virtue of which the demonstration individuates it must be mental properties, since only mental properties are available to do this. It is the state's mental properties in virtue of which the state is part of a scene and is singled out in that scene by a directing intention.

That these properties are irreducibly mental is shown by an argument like the one employed above with respect to descriptions. It follows prima facie from the fact that modes of presentation associated with direct demonstrative reference to qualitative states cannot be shown a priori to be of the same demonstrata as modes of presentation of things paradigmatically physical or modes of presentation employing topic-neutral forms of demonstration. In the latter case, I observe again that this appears not only not

---

13. As opposed to indirect demonstrative reference to a qualitative state, in which case the demonstrator need not have knowledge by acquaintance of the qualitative state and does not display such knowledge in the demonstration, as when I point toward a grimacing Clinton and say, "That headache."

a priori but, given the absent-qualia argument against functionalism, false.

The physicalist might object that it is not helpful to compare demonstrative reference of a phenomenal sort and demonstrative reference of a physicalistic sort, as the prima facie argument above does, since demonstrative references of the latter sort could never even rise to the level of demonstrative references of the former sort. The reason the two are not linked a priori, the physicalist might object, is not that direct demonstrative references to things paradigmatically physical take place in virtue of distinct properties; the reason, it is said, is that there are no direct demonstrative references of a physicalist sort at all.

The reply to the physicalist's objection is that the physicalist has the direction of explanation reversed. There are no direct demonstrative references of a physicalist sort because direct reference of the normal sort makes use of phenomenal properties beyond those employed in physical reference. There could not be direct reference of the familiar sort without access to phenomenal states in virtue of special properties of them distinct from their physical properties.

IV. A Different Interpretation of the Knowledge Argument

The defense of the Knowledge Argument now goes like this. We know our qualitative mental states by acquaintance, picking them out by direct reference as states "like <u>this</u>," so to speak, ostending to ourselves some occurrent state.

If physicalism is true, then all routes to the referents of the singular terms we use to refer to aspects of ourselves run via physical properties or topic-neutral properties, properties neither physical nor mental. Thus, if physicalism is true, then all routes from the knowledge we have of our qualitative mental states to the states themselves that are the objects of that knowledge run via physical or topic-neutral properties of those states. That is, if physicalism is true, then all routes from such states of knowledge to the objects of such states of knowledge are of one kind--let me call it a <u>physical-functional</u> kind. But there are at least two kinds. Besides routes of a physical-functional kind, there are also routes that run via irreducibly mental properties.

Thus, while Horgan is correct that the intentionality of Mary's knowledge is consistent with the possibility that the physical knowledge she has before her release has the same objects as the phenomenal knowledge she comes to have after release, he is wrong that this is compatible with

physicalism. Mary could only come to possess the "first-person ostensive perspective" on her qualitative states--the new perspective in virtue of which she makes the discovery of what it is like to see red and can refer to it directly--if she could pick those states out in virtue of irreducibly mental properties of them. The physicalist cannot at this time say how it is that Mary comes to have, if she does, a second kind of knowledge via a second referential route to her phenomenal states--that is, why it is that there is a "first-person ostensive perspective" at all--and it is hard to see how the physicalist could ever say how.

It is not because there is a form of knowledge beyond the paradigmatic forms of scientific and everyday knowledge we have of physical things that the physicalist story is incomplete--there are forms of such knowledge fully compatible with physicalism. Nor will it do to claim, as Madell does, that the physicalist cannot distinguish the knowledge by description we have of physical things from the knowledge by acquaintance we have of our phenomenal states--these differ at least in their forms of representation. Rather, physicalism runs aground for different reasons.

If Mary has phenomenal knowledge that picks out its objects in virtue of different properties from any in virtue of which her physical knowledge picks out its objects, then physicalism is false. The fact that Mary discovers what it is like to see red after knowing everything physical about

seeing red assures that Mary's phenomenal knowledge picks out its objects in virtue of different properties--irreducibly different ones. Thus, physicalism is false. This is the true version of the Knowledge Argument. It requires as a premise not the claim that Mary, despite her neuroscientific omniscience, lacks some knowledge or other but that she lacks the very specific knowledge of what it is like to see red. The further premise that is required by the Knowledge Argument, then, is something like this.

> The Further Premise, Final Version
> If physicalism is true, then if Mary knows everything expressed or expressible in explicitly physical language about what it is like to see red, she knows everything about what it is like to see red.

And this is true because of a Property Dualism Argument. It is true because physicalism would require that her knowing everything expressed or expressible in explicitly physical language about what it is like to see red would be knowing by every route there normally could be--that is, the only route there normally would be--about what it is like to see red. But, of course, there is at least one other route for Mary, as Mary discovers on her release.

This answers critics who would argue that whatever difference between knowledge by description of matters physical and knowledge by acquaintance of matters mental leads some to infer a difference in types of facts and

properties is nothing more than a difference in styles of representation. It is true, as such critics suggest, that somewhere along the reference chain there is a divergence in styles of representation, matters physical being represented in virtue of definite descriptions, matters of a mental nature in virtue of direct reference. But these forms of representation would not succeed in picking out their referents, the objects of knowledge, unless there were a further difference. The form of representation picking out qualitative mental states does so in virtue of properties distinct from the paradigmatically physical properties in virtue of which neurophysiological descriptions refer.

Now consider the argument of Brian Loar, who attempts to meet this objection to physicalism by giving a detailed description of how one might refer to or have knowledge of physical properties of one's own experiences through what Horgan calls the "first-person ostensive perspective." He suggests that we can be led through the first-person ostensive perspective of what he calls "recognitional/ imaginative concepts" to the very same physical properties of our brains that we are led to by way of the third-person perspective of the theoretical concepts of neuroscience.

> Given a normal background of cognitive capacities, certain recognitional or discriminative dispositions suffice for having specific recognitional concepts, which is just to say, suffice for the capacity to make judgments that depend specifically on those

> recognitional dispositions. Simple such judgments have the form: the object (event, situation) a is one of that kind, where the cognitive backing for the predicate is just a recognitional disposition, i.e. a disposition to classify objects (events, situations) together, that often but not inevitably is linked with a specific imaginative capacity.

If a recognitional/imaginative concept is linked to the ability to class together things with the same objective property, Loar says that the property "triggers" applications of the concept. In that case, Loar writes, "the property that triggers the concept is the semantic value or reference of the concept; the concept directly refers to the property, unmediated by a higher order reference-fixer" (my emphasis). And nothing, he argues, prevents the property picked out by some theoretical concept also triggering some recognitional/imaginative concept, "so the two concepts can converge in their reference despite their cognitive independence...."[14] They would do this without introducing separate properties, since there would be no higher-order reference-fixer on the phenomenal-concept side at all to introduce new, further properties. This, if Loar is right, would refute the Semantic Premise.

We can think up cases for which Loar's point is well-taken, but these are cases very different from those that generate the mind-body problem. Imagine someone, for

---

14. See Loar, op. cit., pp. 84, 87-88.

example, who can, without physical evidence, report and categorize many of her own brain states, even states that lack qualitative character. Call her Marcy. For Marcy, there will be states lacking qualitative character, toward which she has what Horgan calls "the first-person ostensive perspective" and which she may refer to, both to herself and to others, using demonstratives. She reports her own brain states, but she does not do so in virtue of physiological or phenomenal evidence. Imagine also that Marcy sometimes goes through very disunified states of mind. In these states, Marcy may be undergoing brain states associated with pain but unable to report the pain by its feel. Still, in these states, let us suppose, Marcy can employ her ability to report her brain states without evidence to report her being in pain. In these cases, she reports without physical or phenomenal evidence being in pain.

Marcy has access to her states through separate routes that could create the illusion of dualism. Connections between states lacking qualitative character but demonstrated from the "first-person ostensive perspective" and states picked out through explicitly physical properties of the person could only be known a posteriori. Here it would not follow from someone's having two distinct forms of knowledge, forms not linked a priori but following distinct routes to the object of knowledge, that each route of knowledge would pick out its object in virtue of entirely

distinct properties of or facts about the object.

In this case, the direct reference would not pick out the object of reference, the brain state in question, in virtue of any properties of the referent of which the subject is aware. Instead, Marcy would "just know" she was having the brain state and would not make reference to it in virtue of evidence about it. The direct reference might succeed in picking out its object by physical properties of the referent of which the subject is unaware, properties that might also figure into the route by which some neurophysiological descriptions pick out the object. Still, this would be a separate route from the route by which neurophysiological descriptions picked out states of her, and a separate route still from any phenomenal one.

To use Loar's terminology, Marcy has a recognitional/imaginative concept that is triggered by, and thus has as a semantic value, the very same property referred to by some theoretical concept of neurophysiology. The two concepts are linked a posteriori. But it would be wrong to conclude they introduce distinct properties. Clearly they do not. This example, however, is very much unlike us and thus beside the point when it comes to understanding creatures like us. The case of Marcy and the kind of case where Loar's argument most obviously works are cases with "just known" routes, in which predicates are applied in virtue of no properties of the referents. In those cases, we do not

require "higher order reference fixers." But the normal case of demonstrative self-reference to one's own phenomenal states or phenomenal properties would seem to require higher order reference-fixers--those phenomenal properties in virtue of which one picks out one's own states, properties distinct from any other properties. Although it is different perhaps for other recognitional concepts, so-called "phenomenal/recognitional concepts" could not pick out referents at all unless they did so in virtue of properties distinct from the physical or functional properties by which so-called "physical/functional theoretical concepts" refer.[15]

Notice that the critic's line of argument is analogous to a counterargument that actually does defeat the modal argument. On this non-Fregean story, our Cartesian modal intuitions of a contingent connection between the phenomenal and the physical are explained not by ontological differences in parts of the world but rather by distinct styles of representing the world--that is, direct, demonstrative reference versus descriptive reference. This

---

15. Loar, op. cit., pp. 97-98, seems to concede that there might be someone like Marcy and that not all recognitional concepts are phenomenal concepts. What more is required, then, to be a phenomenal concept? It seems to be "the ability to re-identify and perhaps to imagine a feeling of a certain type, for example, feeling like this." But now Loar has conceded too much: phenomenal concepts can only refer this way in virtue of a mode of presentation that re-introduces phenomenal properties. (I owe this point to Ned Block.)

form of counterargument worked against the modal argument. In that case, one could explain the appearances of contingency that seemed to make it possible to pull mentalistic and physicalistic representations apart by the possibility that two different kinds of reference were picking out the same things. By contrast, the Knowledge Argument makes no claim of contingency between pieces of the world, only of nonidentity between properties of the world. Thus it does not work in this case.

Assume, for the sake of argument, that it is not possible to pull the two kinds of representations apart, that they have the very same referents in common. Still, the very distinction between direct, demonstrative reference and descriptive reference in cases we are most familiar with--our introspective, phenomenal knowledge and our nonintrospective knowledge of the things we identify in paradigmatically physical ways--entails the existence of distinct sorts of properties.

What makes this initially plausible is the clear-cut intuition that we directly pick out referents in introspection at least partly in virtue of their <u>looks</u> and their <u>feels</u>, and that these are different from any of the paradigmatically physical properties in virtue of which we pick out aspects of our brains. It does not matter whether we do this descriptively or demonstratively. My argument in this chapter, however, does not simply rely on what is

merely plausible but undertakes the work of creating an explanatory model that guarantees that the distinctness of the direct, introspective route cannot be accounted for on other grounds.

Thus, by ruling out cases of just-known referents, not picked out in virtue of any properties of them, and by requiring similarity in styles of reference, we arrive finally at a common-sense argument for property dualism. This argument makes use of a version of the Semantic Premise which withstands the objections raised above and can be set forth as follows.

> The Semantic Premise, Final Version
> For all pairs of referring expressions R_1 and R_2 of the same style of reference, except those R_1 whose referents are just known independently of properties of the referents,[16] if R_1 and R_2 are coreferential but not a priori, then there exists a property $\emptyset$ which we pick out the referent of R_1 in virtue of and which is distinct from any the properties we pick out the referent of R_2 in virtue of.

There are thus three sorts of routes to referents, as Figure 4 shows. Suppose that there were someone who was like Marcy in actually picking out internal states of hers in each of these three ways. She might state identities in this way.

---

16. Here I include pure indexicals, such as "I," if we refer with them independently of properties of their referents.

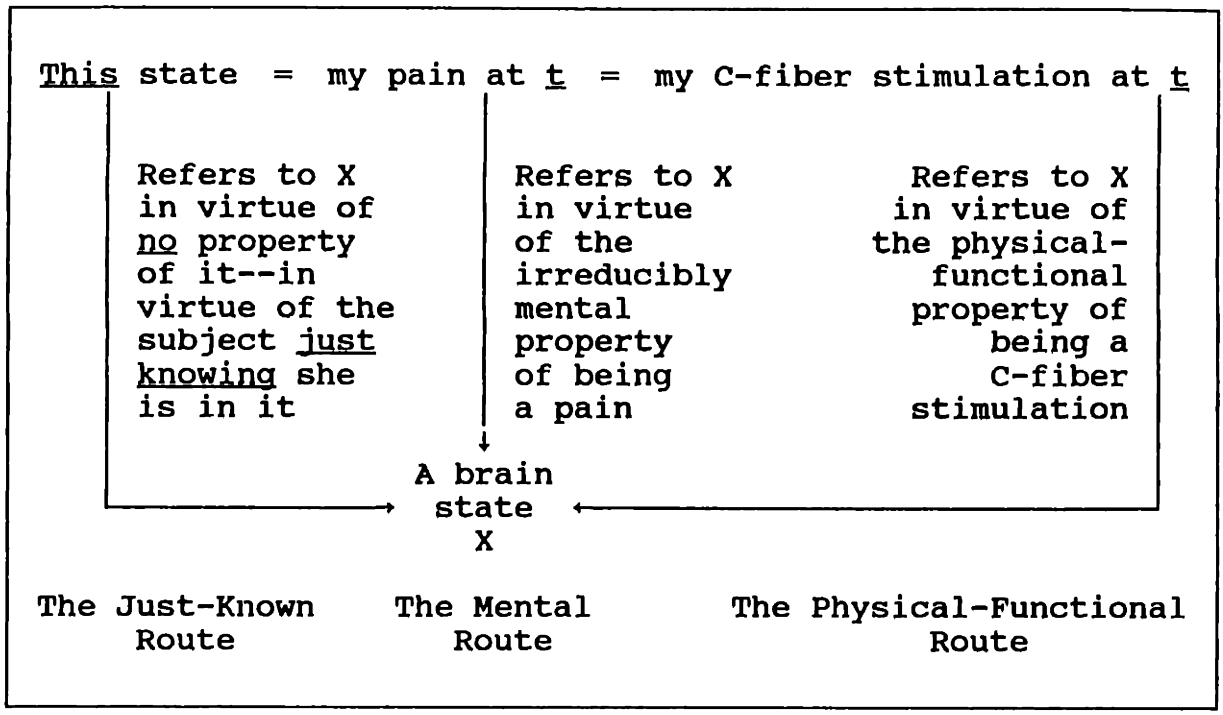

Fig. 4--Three Kinds of Route to the Referent

If we exclude picking out referents in ways that are just known and if we exhaust ways of picking them out along physical-functional routes, then any ways of picking them out that remain must do so in virtue of irreducibly mental properties.

This, then, is the fundamental basis of the Property Dualism Argument, and thus of the Knowledge Argument. It is the source of what is right about the Cartesian approach to the mind-body problem. And it provides some sense to Descartes's idea that our Cartesian intuitions represent perceptions of distinct mental properties. For the phenomenal routes we have to our mental states are routes, in contrast with what is just known, which run by way of

properties that impact on our experience, but also impact differently from any physical-functional route. Thus let me make this conjecture.

A Conjecture

Refuting the Knowledge Argument would be tantamount to solving the mind-body problem.

Phenomenal consciousness provides us with evidence of our internal states, in contrast to the no-evidence route of what is just known. By the Knowledge Argument, this requires irreducibly mental properties over and above our physical and functional properties. The Knowledge Argument, then, raises many questions. How can anything provide evidence by this route of phenomenal consciousness? Why aren't the contents of our internal states just known. How can there exist something which provides us with evidence? Moreover, why is phenomenal consciousness needed at all? Why can't something else provide this evidence? Why, for example, can't there be a physical mechanism causing self-knowledge of the contents of brain states by way of evidence but nonqualitatively--that is, by physical means that reproduce the effects of phenomenal consciousness? Or, to return to the even more extreme Cartesian intuition, why can't the body do everything it does by way of its physiological properties alone--why are further phenomenal properties required?

It is not a weakness of the Knowledge Argument that it

does not provide answers. The Knowledge Argument only sanctions the Cartesian side of the mind-body problem; it does not solve the problem. It is clear that very different considerations, probably far outside common sense, will be required to do that.

V. No Common-Sense Way Out

Arguments like the Knowledge Argument and the Property Dualism Argument have been thought to be vulnerable to several kinds of common-sense objections. But none of them seem adequate to the arguments advanced in the preceding section.

One option for the physicalist might be like the one that McGinn argued against Kripke. McGinn never discusses Jackson, but it is clear how this form of argument might go. On this view, psychophysical reductions are "cognitively closed" to us, and the best the anti-reductionist can do is to show that phenomenal properties are physical but noumenal, beyond any humanly possible psychophysical reduction. But this is no more successful against the neo-Cartesian strategies of the Knowledge Argument and the Property Dualism Argument than it was against the orthodox Cartesian. We do not accept the Semantic Premise on the grounds an argument from cognitive closure would require. There is no place in the argument where we jump to property

dualism merely by an inability to come up with an alternative. Rather, the argument proceeds from very general considerations about knowledge and reference that have wide application. It is based on clear, well-tested, theory-based intuitions. Nothing seems to be cognitively closed. Thus, there is no more reason here than with the orthodox Cartesian to think that the central ideas are plausible only because of the cognitive closure of them.

This leaves the two options I cited at the beginning of Chapter Four, section IV. One was that there was a problem with the mechanics of the Knowledge Argument. I argued that there was such a problem with Jackson's version. In this chapter I have tried to construct a version in which there are no such problems. Against this version, the physicalist might continue to pursue Loar's strategy of showing that phenomenal-concept terms pick out referents without higher order reference-fixers. I have not discussed all the ways someone might follow this strategy, but none seem very promising to me. A suggestion I have not considered, for example, is that these terms behave more like pure indexicals, contributing to truth values in virtue of the appropriateness of their circumstances of use. This option, however, runs up against the fact, noted in the literature and discussed in Chapter Seven,[17] that pure indexicals are

---

17. See, for example, Stephen Schiffer, "The Basis of Reference," Erkenntnis 13 (1978), pp. 171-206, and the reply by David Austin, What's the Meaning of "This"? (Ithaca:

very different from demonstratives and descriptions.

The other common-sense option against the Knowledge Argument has been to argue that although there are nonphysical properties they are functional properties. On this view, they are characterizable entirely by reference to causal relationships among stimuli, behavioral responses and internal psychological states--and thus are not irreducibly mental. This line of counterargument is defeated by the absent-qualia argument defended in the next three chapters. However, as I have argued, it is already independently defeated by the Knowledge Argument and the Property Dualism Argument. They advance considerations that, unlike those of the modal argument against physicalism, weigh against functionalism as well.[18] Even if some functional description were to pick out the same mental state as does some mentalistic description, it seems that it could only do this by way of a distinct referential route, reaching the common referent by distinct causal properties of the

---

Cornell University Press, 1990), ch. 3. See also Michael Bennett's view set out in Kaplan, op. cit., pp. 527-528.

18. The development out of Smart's response to Black's argument of one line of functionalist writing is traced by Ned Block, "What Is Functionalism?", in Ned Block, ed., Readings in the Philosophy of Psychology, volume 1 (Cambridge, Mass.: Harvard University Press, 1980), p. 179. The difficulties functionalism faces in accounting for qualia are well-known. They are reviewed, among other places, in David Lewis, "Mad Pain and Martian Pain," and Ned Block, "Troubles with Functionalism," both reprinted in Ned Block, ed., op. cit., pp. 216-222 and 268-305, respectively. Once again, the considerations against functionalism that come out of the Knowledge Argument are additional ones.

referent, separate from any of its phenomenal properties. Mary might know everything knowable by a physical-functional route to the brain about what it is like to see red without knowing by the phenomenal route. By the previous argument, this is to know by properties distinct from those by which she lacks but gains knowledge. The analytic functionalist's claim of an a priori link between functional and phenomenal-concept terms, one that makes us pick out mental aspects of the world by way of a single type of properties, is thus inconsistent with common sense.[19]

This suggests that any solution to the mind-body problem will carry us far outside common sense.

---

19. This seems to be the lesson, though unintended, of Ned Block's homunculus-headed and China examples. In the first thought experiment, we create a system that produces behaviour like yours through a Turing-machine-simulation of your psychology, using many little men located in its "head" each responsible for one state instruction from the Turing table the system uses, corresponding to your psychology; in the second, we convert the government leaders of China to functionalism and convince them to enlist that country's billion-plus inhabitants to realize a human mind for an hour. See Block, op. cit., p. 276. Block's conclusion that there is no consciousness in either system begs the question against the functionalist, a point correctly noted in Lycan, op. cit., pp. 26-27. Nevertheless, although one cannot conclude from Block's examples, as Block appears to, that consciousness does not supervene on the two systems, it is quite clear that if there were any phenomenal properties that did supervene on functional properties of the systems, they would prima facie be distinct properties from any functional properties, since there would surely be no a priori connection between the phenomenal and the functional properties. Nothing we could be told about either system in functional terms would by itself constitute telling us that it had phenomenal properties. This is enough to undermine functionalism.

CHAPTER SIX FUNCTIONALISM AND SKEPTICISM

According to the functionalist, terms denoting qualitative mental states can be defined descriptively, as those states which have such-and-such causes and such-and-such effects. The Knowledge Argument contradicts functionalism, since it requires that functional properties, just as physical properties do, provide distinct routes to referents from the ones phenomenal properties provide. In case the functionalist thinks that there are ways out of that argument, the Cartesian will want to bolster the case against functionalism by a further argument.

If sound, the Knowledge Argument would establish the nonidentity of functional and phenomenal properties. The further counterargument to functionalism I will now develop shows that phenomenal properties do not even supervene upon functional ones. The strategy is familiar: to argue that there might be a state which would have all the causes and effects of a typical qualitative mental state but would lack qualitative character altogether. Such a state, were it to exist, would be said to be an absent qualia or ersatz state. Ersatz counterparts for such types of genuine qualitative mental states as pains and the havings of red after-images would said to be ersatz pains and ersatz havings of red after-images.

A fairly obvious kind of functionalist reply is as follows. It might seem, were the counterargument sound,

that we could not know we were having the qualitative mental states we do actually have. For if there were absent-qualia states, it might seem, they would have no causes or effects to make it possible for a subject to distinguish having them from having their non-ersatz counterparts. Surely if we know anything, we know that we are having the qualitative mental states we are having. Since we do know that, it might seem, absent-qualia states are impossible and the argument does not work against functionalism.

This defense of functionalism against the possibility of absent qualia need not be verificationist, and thus it need not suffer the faults usually associated with verificationism. So far, it makes no explicit mention of meaning, nor need it rely on covert assumptions connecting meaning and verifiability.

It goes awry in a different way, however, by relying on a false theory of knowledge. The argument seems to depend on some version of the principle, suggested by some of Descartes's reasoning in the Meditations, that an individual cannot know a proposition to be true unless the individual has evidence to distinguish the case of its truth from all cases of its falsity. The problem with such an epistemological principle as this is that it would make impossible (at least without something like Descartes's eventual certainty of the good intentions of a morally perfect Creator) various kinds of knowledge that we in fact

possess. In the case of our knowledge of the external world, for example, we never have comprehensive evidence distinguishing actual perceptions of the external world from those cases in which an evil deceiver causes qualitatively identical perceptual states that bear no relation to the external world. Still, we do have knowledge of the external world. It seems to be enough that our beliefs result from belief-producing mechanisms that reliably discriminate the truth from relevant, although not necessarily all, alternatives to it.

I shall argue that the anti-skepticism argument against the possibility of absent qualia, which is directed at our common sense, in fact goes well beyond common sense. This is because it appeals to a supposedly common-sense epistemological principle that makes our knowledge of our own qualia depend upon evidence more comprehensive than common sense sanctions. The argument makes such knowledge depend upon our distinguishing the actual qualia our mental states have from more alternatives to our qualia than common sense tells us our knowledge of them depends.

The case I am making in this chapter against the anti-skepticism argument I intend to be a modest one. It is important for the reader to bear that in mind to prevent misunderstanding. I do not deny that for all we know we might come up with <u>empirical</u> evidence supporting the epistemological principle on which the anti-skepticism

argument relies. Rather, I make only these two different claims: that if we were to find such evidence it would contradict common sense, and that the anti-skepticism argument therefore cannot rely merely on common sense. In the final chapter, I shall argue that the anti-skepticism argument is, in fact, correct that if absent qualia states were possible we could not distinguish genuine qualitative states from their ersatz counterparts. But this is true, I shall argue, not because of the general epistemological principle at issue here but because of more peculiar reasons that invalidate the remainder of the anti-skepticism argument.

In the present chapter, after setting out some background in sections one and two, I will show in sections three and four that current versions of the anti-skepticism argument appearing in the literature have this difficulty. I will argue that it is not at all obvious that our evidence is as complete as these published versions require. Our evidence would be so complete as this if we had transparent access to the phenomenal properties of our own mental states, but I will argue that there is a strong common-sense case that we lack it in crucial ways. In the final section, I shall argue that no improvement of the anti-skepticism argument will escape this problem and that any apparent virtues to the anti-skepticism argument that remain are illusory and can be explained away.

I. The Anti-Skepticism Argument

As a first approximation, I will represent the anti-skepticism argument against the possibility of absent qualia as follows.[1]

The Basic Argument

(1) If absent qualia are possible, then we cannot know of our qualitative states that we are having them.

(2) We do know of our qualitative states that we are having them.

Ergo, absent qualia states are impossible.

How might a defender of the argument support its premises? Sydney Shoemaker and Earl Conee provide considerations of the sort discussed above.[2] If absent

---

1. This kind of argument was first advanced in print by Sydney Shoemaker, "Functionalism and Qualia," Philosophical Studies 27 (1975), pp. 291-315; this is reprinted as chapter 9 of his Identity, Cause, and Mind (Cambridge: Cambridge University Press, 1984), pp. 184-205. Shoemaker continues to endorse his original conclusions; see his "Qualia and Consciousness," Mind 100 (1991), p. 507. It has also been advanced recently in Georges Rey, "Sensational Sentences," unpublished manuscript, July 1989, and in Dennett, op. cit.

2. A very different way of defending the first premise is considered and correctly rejected by Ned Block in his "Are Absent Qualia Impossible?", Philosophical Review, April 1980, pp. 257-274. The view considered is that (1) the possibility of absent qualia would entail the epiphenomenality of qualitative character to the mental states having it, but that (2) that would entail, by the causal theory of knowledge, the unknowability of qualitative character. Block's response is that (2) is only non-question-begging if one's picture of a property's being epiphenomenal to a mental state is such that genuine states

qualia were possible, it might seem that the presence or absence in a mental state of qualitative character would make no difference to the functional causal role of the mental state--that is, to its causal relations with perceptions, behaviors and other mental states. If it would make no difference, then it would seem to lack distinctive evidence for the presence of qualitative character. You could not tell if you had it. But if we know that some p is true, then we do have distinctive evidence for its truth, distinguishing p's truth from any possible case of p's not being true. From this, it would follow that the argument's first premise would be true--that if absent qualia were possible, then we would not know of our qualitative states that we are having them.

From these considerations, Conee, a critic of Shoemaker, reconstructs Shoemaker's argument against the possibility of absent qualia along the following lines.[3]

---

and ersatz counterparts have all the same effects (not just all the same psychological ones), but then (1) is question-begging against a physicalist story about how to create ersatz states.

3. Earl Conee, "The Possibility of Absent Qualia," The Philosophical Review, July 1985, pp. 345-366.

Conee's Reconstruction of the Anti-Skepticism Argument

(EP) If ersatz pain is possible, then it is not possible to distinguish cases of genuine pain from cases of ersatz pain.

(K) It is possible to distinguish cases of genuine pain from cases of any possible state lacking qualitative character.

Ergo, ersatz pain is not possible.[4]

One brief comment about distinguishing. For the purposes of this chapter, I will suppose that if one can do the distinguishing (K) requires one can correctly sort a set of occurrent states into genuine pains and states lacking qualitative character. In the next two chapters, however, an ever looser condition on distinguishing will be consistent with what I say. There, all that is required is the detection of a difference, one that might even fail to be transparent to the subject.

Conee's accounts of both why K is needed and how it can be supported are flawed.[5] First, consider the former.

---

4. For ease of exposition, I have eliminated parts of Conee's versions of EP and K that are irrelevant to my account.

5. As stated above, Conee's version of the argument is a reconstruction. Shoemaker acknowledges that, while he accepts it himself, the very step from EP to the conclusion, whatever premise is employed, is unconvincing to some. Its controversial character motivates him to propose a second argument against the possibility of absent qualia that does not depend on this step or any other epistemological premise. The second argument is based on the idea that since anyone functionally identical to us will use "pain," for example, with the same meaning as us (since the causal

Conee is correct about the need for something like K but wrong about the reasons. Why not derive the impossibility of absent qualia from EP by a modus tollens argument such as the one above but by taking as the minor premise of the argument K*, the denial of EP's consequent, instead of K?

(K*) It is possible to distinguish cases of genuine pain from cases of ersatz pain.

(K) It is possible to distinguish cases of genuine pain from cases of any possible state lacking qualitative character.

K entails K*; K ranges over much more distinguishing than does K*. Conee incorrectly maintains that it is not open to Shoemaker on logical grounds to derive the impossibility of absent qualia from K* and EP. He writes that K* implies "that it is possible to distinguish cases of genuine pain from cases of ersatz pain" and that "[t]his is something that Shoemaker would have to deny, since he holds that there

---

stories connecting reference and referents with be the same), a functional duplicate will refer to the same thing we do in using the term. But this is to assume erroneously that functional identity determines identity of meaning and reference. Putnam has cast doubt on this view; see "The Meaning of 'Meaning,'" in K. Gunderson, ed., Language, Mind, and Knowledge (Minneapolis: Univ. of Minnesota Press, 1975). Other counterarguments appear in Stephen White, "Curse of the Qualia," op. cit., pp. 340-350, and in Conee, op. cit., pp. 364-366. Nowhere does Shoemaker explicitly endorse K as the step from EP to the conclusion, but Conee has argued both that K is defensible along the lines set out above and that the anti-skepticism argument requires something like it. He reserves his criticisms for EP.

cannot be any case of ersatz pain."[6] But Shoemaker no more must deny this than he must deny that it is possible to distinguish cases of elliptical pegs from cases of round square pegs; both sorts of distinguishing are possible, even though there cannot be any case of a round square peg and even though he argues there cannot be any case of ersatz pain.

There is a rationale for this step in the argument but it is different. It is this: the case against believing absent qualia possible does not rest on an epistemological premise limited to difficulties with ersatz pains alone. It must be principled, drawing on more general facts about mental states that would include facts about ersatz pains, if there were any. Thus K. The aim of the opponent of the possibility of absent qualia is to find a step from EP to the impossibility of absent qualia which distinguishes genuine pains from every one of a set of mental states wide enough to include ersatz pains but not so wide that the principle lacks plausibility.

II. The Theory of Knowledge the Argument Depends On

Let us now consider more closely the theory of knowledge on which the argument depends to determine whether K is, in fact, defensible, as Conee claims it to be. The

---

6. Conee, op. cit., p. 351.

step from EP to the anti-absent-qualia conclusion might seem to depend on the problematical principle cited above in defense of the anti-skepticism argument. By that principle, a subject cannot know that a proposition p is true if there is no evidence that distinguishes, in principle, p's truth from p's falsity.

Such a principle conflicts with the natural idea that knowledge is possible because a subject knows a proposition to be true, if it is true, in virtue of using reliable belief-forming processes, processes that reliably produce true beliefs. It is possible for there to be reliable belief-fixing mental processes even if there is no evidence so comprehensive as to distinguish actual perceptions from all possible cases of deception. This kind of view is known as reliabilist. All sides would seem to accept reliabilist constraints of some type on an adequate theory of knowledge, although they might differ in the details. I will follow them in this respect.[7]

Thus, a straightforward refutation of this way of justifying the anti-skepticism argument is that it is no more plausible than parallel arguments against skepticism about the external world. I seem to be wearing a wristwatch as I write this at 3 p.m. on July 4, 1989. A skeptic might

---

7. The reliabilist approach is in large respect due to Alvin Goldman; see his "What Is Justified Belief?" in George Pappas, ed., Justification and Knowledge (Dordrecht: Reidel, 1979).

say that it is possible I am not wearing one. It is surely not an adequate reply to the skeptic to assert that that cannot be so since then I could never know I am wearing a wristwatch, even if I were. It is not a good reply because in the case of wristwatches, and in the case of the external world more generally, the possibility of knowledge is compatible with the possibility of unremediable deception. For there are processes of belief fixation that reliably generate the belief I am wearing a wristwatch when I am (and fail to generate the belief that I am wearing a wristwatch when I am not), and thus provide justified true belief, or knowledge, to that effect. Absent wristwatchhood--satisfied by states of affairs that have all the evidentiary relations of normal wristwatches in the absence of any actual wristwatches--is not ruled out because it would lead to the impossibility of knowledge. Why should absent qualia be ruled out on a similar basis?

This has suggested a different strategy to opponents of absent qualia: searching for a more limited epistemological claim anchoring the step from EP to the anti-absent-qualia conclusion but not sanctioning skepticism about the external world. According to this line of thought, the case of K is a very special case since the difficulties with distinguishing appearance from reality that undermined the problematical general principle above need not plague K. The truth of K requires no distinguishing of appearance from

reality but only the distinguishing of appearance from appearance. Conee suggests that K is supported by the principle R*.

> (R*) For all b and for all S such that b is a report of the content of S's experience to the effect that it has phenomenal quality phi, then S knows b true only if S has evidence that distinguishes S's experience's having phi from any possible case of S's experience not having phi.[8]

Taken together with some further Knowledge Premise like the following--

> Knowledge Premise
> We know our experiences to have the qualia they do.

--R* would provide adequate support for K's claim that we can distinguish our pains from states lacking qualia. Taken together with EP, the anti-absent-qualia conclusion would follow directly.

To Conee, supporting R* is a more manageable task. R* does not lead to skepticism by setting unsatisfiable conditions on knowledge of the external world, since it does not apply to such knowledge. R* makes a more modest epistemological claim. It is based on the insight that the reality/appearance distinction collapses when it comes to my

---

8. Conee, op. cit., p. 353. What I call R* Conee calls R, but I reserve this latter label for the revision of R* which is the subject of much of the rest of the chapter.

own experiences. Here, the knowledge is direct, unmediated by further appearances that could potentially deceive me about an independent reality. Surely, it may seem, there is a difference between being in a certain phenomenal state and not being, and surely we are directly aware of this difference.[9] Thus, R* requires what I will henceforth call "comprehensive evidence"--evidence which distinguishes the truth of a known proposition from all possible cases in which the proposition does not obtain--but only in cases of knowledge in which it may seem unproblematical to do so.

In fact, R* does not fully satisfy the aim of shutting out the problems. It leaves the door open just a crack on a world external to appearances where these skeptical difficulties continue to arise.

Consider Jones, who sees on a cerebroscope (a biofeedback device giving him contemporaneous representations of states of his brain) that he is having experiences with a particular phenomenal quality. Jones knows this because the cerebroscope, while not telling him which phenomenal quality it is, does show neurological activity in an area of his brain characteristically associated with his phenomenal experience. There is a pattern to what the cerebroscope displays to him about this phenomenal quality of his experience, one that repeats itself several more times; Jones introduces the name "phi"

---

9. Ibid, p. 353.

to refer to the phenomenal quality picked out by this pattern exhibited on the cerebroscope. Throughout this period of time, as Jones watches phi on the cerebroscope, he notices introspectively several phenomenal properties--A, B, C, D--that phi might be. However, Jones does not know of these phenomenal qualities that he introspectively finds in his experience--A, B, C, D--which one is the one he has identified with the cerebroscope as phi. Jones thus has a peculiar kind of knowledge of the phenomenal quality of his experience, a kind of knowledge that is not direct knowledge.

Let b be Jones' report to friends of the content of his experience to the effect that it has phenomenal quality phi. Jones has fixed the reference of the term "phi" for his friends in the same way as he had previously done for himself, but b does not, let us suppose, convey to his friends any more than it does to Jones himself about what it is like to go undergo experience having phi. Surely Jones' knowing b to be true does not depend, contrary to R*, on his having comprehensive evidence distinguishing the case of his experience having phi from every possible case of his experience not having phi. Surely he need not distinguish it from the case in which there is no cerebroscope but only a hallucination of one caused by an evil deceiver. Here the difficulty with R* is that it quantifies over all beliefs about the phenomenal qualities of experiences. Yet some of

those beliefs, even though knowledge, are not immune from the possibility of deception about the external world, since they pick out the phenomenal qualities nonphenomenally.

Now consider a case in which Jones sees on a cerebroscope that someone is having experience with phenomenal quality phi although he knows not whom. By now, let us suppose, he knows what it is like to have phi-experiences; thus, his belief ascribing experiences with phi to someone picks out the phenomenal quality phenomenally. He does not know whom the cerebroscope is reporting to have phi-experiences but he is able to pick the person out descriptively, whoever is it, as the person being monitored by the cerebroscope. In fact, but unbeknownst to him, the person is himself. Now let b be a qualitative belief of Jones' about Jones' experience to the effect that it has phi. Surely, once again, Jones' knowing b to be true does not depend, contrary to R*, on his having comprehensive evidence distinguishing this case of Jones' experience having phi from the case in which the cerebroscope is a hallucination. The trouble with R*, once again, is that it allows the beliefs to which it applies to depend upon information about the external world--this time, information from the cerebroscope concerning whom is being monitored.

These two kinds of difficulties with R*, however, are not faced by the following principle. R quantifies not over third-person public reports--which a subject can be wrong

about even when they refer to the subject's own experience, in much the same way as the subject can be wrong about the experience of others--but over first-person beliefs by acquaintance, which a subject cannot be wrong about in that way.

> (R) For all b and for all S such that b is a self-ascribing qualitative belief of S's about the content of S's experience to the effect that it has phenomenal quality phi, then b is knowledge for S only if S has introspective evidence that distinguishes S's experience's having phi from any possible case of S's experience not having phi.

Beliefs of the kind to which R applies cannot misfire over extraordinary facts about the external world. The difficulties over knowledge about oneself and knowledge about the phenomenal qualities of one's experiences that plagued R* are outside the range of R, since the knowledge to which R applies is all direct and by acquaintance with one's qualia and one's ownership of them. The door on the outside world has been closed all the way.

Before proceeding, let me review the structure of the argument based on R. First, recall the Knowledge Premise.

> Knowledge Premise
> We know our experiences to have the qualia they do.

If R and the Knowledge Premise are true, supporting K, then since ersatz pains, if they exist, are among the possible

states lacking the qualitative character of genuine pain, EP's consequent is contradicted. Recall EP.

(EP) If ersatz pain is possible, then it possible to distinguish cases of genuine pain from cases of ersatz pain.

R and the Knowledge Premise together directly contradict EP's consequent; K is, in fact, unnecessary to the argument.

As stated earlier, these facts are true in part because of the relation between R, K and EP's consequent. In those cases of phi-experiences which make the Knowledge Premise true, R and K depend on their truth upon, respectively, D_R and D_K.

(D_R) It is possible to distinguish experiences having phi from every experience lacking phi.

(D_K) It is possible to distinguish states having phi from every state lacking qualitative character.

The denial of EP's consequent likewise depends on D_{EP}.

(D_{EP}) It is possible to distinguish phi from every ersatz state.

The logical relationship among these propositions is as follows.

$D_R \rightarrow D_K \rightarrow D_{EP}$

Stated more simply, what is ersatz lacks qualia, what lacks qualia lacks any given phenomenal property phi.

III. Counterarguments to the Theory of Knowledge

But even R is too strong; and without R, K and the anti- skepticism argument as a whole are unsupported. Contrary to R, there are cases of my knowing that I have experiences of a particular phenomenal character where I don't have evidence distinguishing my experiences with that character from any and all experiences without it. In fact, it would seem that no experience is such that I have evidence distinguishing it from any and all experiences which lack the phenomenal character it has, even though it surely does not follow that I never know the phenomenal character of my experiences.

The point can be illustrated by the following case.

> Case 1. Misconfirmation of pain. A sensation prompts me to say, "I am in pain." That judgment, however, is incorrect, caused by a sudden sensation of extreme cold in the context of an expectation of pain.[10]

---

10. Keith Lehrer, Knowledge (New York: Oxford University Press, 1974), p. 96; Stephen L. White, "Transcendentalism and Its Discontents," in his The Unity of Self (Cambridge, Mass.: M.I.T. Press, 1991), p. 129.

In the normal course of events, my expectations of what I will feel are confirmed or disconfirmed by the feeling that follows; misconfirmed expectations are neither. A case of misconfirmation is like confirmation except for relating me to a false proposition, and like disconfirmation except in requiring my belief in the proposition's truth; in this, it is akin to misremembering or to misperception. Now, would it be correct to say of me, on other occasions when I report being in pain accurately and with justification, that I do not <u>know</u> I am in pain, merely because a counterfactual case exists in which I do not have evidence that distinguishes pain from cold, such as in this odd case?[11]

---

11. In related cases, phenomenal qualities are non-transparent not because we get them wrong but because we cannot decide which they are; such cases also refute R because these odd cases of non-transparency do not jeopardize normal cases of phenomenal knowledge. I taste something I have never tasted before. Moreover, it tastes like nothing I have ever tasted. I am unable to categorize it. It is not transparent to me that this experience falls into the phenomenal category it does; thus, I may be unable to distinguish it from experiences of a different phenomenal category, even though I do normally know, contrary to R, when I am having experiences of this latter category.

In tests for what is called color agnosia, subjects are given skeins of wools of different colors and asked to sort them according to color categories. The assortments produced appear to be random. Yet we may satisfy ourselves on other grounds that such subjects have normal color perception and are impaired neither in the ability to recall object colors nor in the ability to use color names. In such cases, which several investigators report finding in connection with brain damage, we can justifiably say that the subjects have no transparent access to the phenomenal color properties they see. Again, subjects may be unable to distinguish what they see from experiences of a different phenomenal category, even though I, contrary to R, know when I am having experiences of this latter category.

The locus classicus on color agnosia is O. Sittig,

This is a counterexample to R. Surely I can knowledgeably ascribe a phenomenal quality to my experience even though I cannot distinguish my experience's having that quality from every single case of my experience not having the quality, such as this odd one. It will not do to defend R by claiming that, even in the odd case, the sensation of pain is "evidently different" from the sensation of cold if that means that the existence of a difference is supposed to be, as Conee writes in support of R, "evident on the basis of immediate awareness."[12] That there is a difference is not evident on the basis of immediate awareness. If it were, it would have to be something in immediate awareness, but all that is in immediate awareness in the odd case is cold, not both pain and cold, and thus never anything making evident a difference between them. Since sensations of pain differ from sensations of cold, they do have different properties, and whenever pain or cold is in awareness, some among these differing properties are in awareness and are evident from awareness. But it would be wrong to conclude from this that the fact that there are different properties is also evident from awareness.

"Störungen im Verhalten gegenüber Farben bei Aphasischen," Monatsschift für Psychiatrie und Neurologie, vol. 49 (1921), pp. 63-68, 169-187. Sittig distinguishes Farbennamenanmesie, which is referred to by (among others) Jules Davidoff, Cognition through Color (Cambridge, Mass.: M.I.T. Press, 1991), as "color anomia," from Farbenagnosie, or color agnosia.

12. Conee, op. cit., pp. 353-354.

We do not have a reality/appearance problem here as in the case of the Gedankenexperiment concerning the evil deceiver; one cannot be misled about the reality by the appearance of it since, in a sense, the appearance is the reality. Still, anti-skeptical intuitions paralleling those that weigh against taking the possibility of an evil deceiver too seriously as a threat to knowledge of the external world seem to have some force here as well in protecting the possibility of knowledge of appearances. It would seem that here, as before, it is possible to have knowledge, even if there is always the possibility of going wrong. The reliabilist's intuition is that stable and reliable processes of belief fixation make it possible for there to be justified beliefs, and thus knowledge, even in the face of hard cases in which what is known is indistinguishable from other things. That intuition seems as applicable here as it is in hard cases about our knowledge of the external world.

The reliabilist can say of the odd case described above that it is fully consistent with having knowledge in non-odd cases since the mental process that fixes belief in the odd case is a different mental process than that operating in most other cases. In the odd case, a pain-belief is triggered by feeling cold while expecting pain. In non-odd cases, pain-beliefs are caused by feeling pain while expecting pain or by feeling pain independently of

expectations. In non-odd cases, knowledge is possible since there is a belief-producing process which reliably produces true beliefs.

The argument against R also works because the odd case represents a failure of transparent access to our mental states, and R requires some transparency to our mental states. I can only distinguish cases of some phenomenal property phi from all incompatible alternatives to it, as R requires for me to have knowledge of phi, if it is transparent to me for each alternative that it is not a case of phi. Thus, in the present case, if R is true, then when I know I am in pain I must be able to distinguish my being in pain from incompatible alternatives to my being in pain, such as my misconfirming being in pain on the basis of feelings of cold. But this would require me to have transparent access to my failure to feel pain in this case, and I do not have such transparent access. Thus R is false.

An opponent of this common-sense argument might try to preserve transparency in one of two ways. First, the opponent might argue that it does not follow from your calling a state of yours "painful" that you believe you feel pain and not cold. This route does not seem promising. Of course, it is possible to misspeak, but it is easy enough to modify the case to rule that out, so that I not only say, "I am in pain," but also wince, make efforts characteristic of being in pain to eliminate the source of the feeling, and

display other pain-related thoughts and behaviors. Moreover, it is a common-sense principle that you believe what you sincerely express, and surely it is no more reasonable to give up that common-sense principle than it is to give up the principle of transparency.

Second, the opponent might claim that if I say I am in pain and wince and register other pain-related thoughts and behaviors, then I _am_ in pain, that that is what it _is_ to be in pain. This, it might seem, is what the functionalist _ought_ to hold. But this is a functionalism that would not even allow the possibility of a failure of transparency, and that is going too far. An anti-skepticism argument must appeal to principles that rule out absent qualia while at least allowing the common-sense possibility of a failure of transparency. It could turn out _a posteriori_ that transparency was always preserved, but appealing to such a principle would go outside common sense. The anti-skepticism argument, however, is presented as an argument from common sense.[13]

---

13. Stephen White, in his "Transcendentalism and Its Discontents," _op. cit._, p. 119, constructs a functionalist theory that allows failures of transparency. White takes his departure from David Lewis' suggestion (in "Mad Pain and Martian Pain," _op. cit._) that the functionalist define psychological terms by their roles in the folk-psychological theory consisting of common-sense platitudes about human psychology. If among these platitudes was that transparency fails, then the functionalist could allow that. Such a move, however, would force an advocate of the anti-skepticism argument to find an epistemological principle that rules out absent-qualia cases while allowing other transparency failures in, and I am unconvinced that this is

One might believe that Descartes ignored the skeptical problem that functionalism is alleged to solve. While Descartes is the source for our appreciation of the threat of skepticism in modern philosophy and reviewed many imaginable ways in which our belief-producing mechanisms could and do go wrong, he never claimed a skeptical threat to knowing the phenomenal contents of one's own mental states. Still, despite what some writers have claimed,[14] the evidence is not conclusive that Descartes believed judgments about phenomenal contents to be incorrigible.

He does allow another imaginable kind of deception about something as seemingly incorrigible as qualitative character: deception concerning the truth of simple mathematical propositions. How could God perpetrate that deception? one might wonder. Descartes does not say. One does not have to assume, as some readers of Descartes do, that it is by His making the world one in which the simple mathematical formulae we take for granted fail--e.g., by His making a world in which $2 + 3 \neq 5$. Rather, one can imagine that it is by His making our minds go wrong when we contemplate mathematics. He might make us going wrong without knowing it when we add 2 and 3. Perhaps he might

---

possible to do.

14. See Rorty, op. cit.; Flanagan, op. cit., pp. 30-31; Dennett, op. cit., pp. 67, 363; Georges Rey, "A Reason for Doubting the Existence of Consciousness," in Richard J. Davidson, Gary E. Schwartz and David Shapiro, eds., Consciousness and Self-Regulation, vol. 3 (New York: Plenum, 1983), pp. 3-4.

also make it hard for us to compare what we think with what is said and written by others by making us go wrong in the meanings we attach to mathematical language.

In this regard, Descartes himself asserts[15] that "many people do not know what they believe, since believing something and knowing that one believes it are different acts of thinking, and the one often occurs without the other." Might Descartes, on grounds similar to this rejection of the transparency of belief, also have rejected the transparency of qualitative character?

There was also for Descartes always the general possibility of God's deception in matters that seem clear and distinct, even apparently including the cogito. This suggests that it would have been the existence of a benevolent God that ultimately for Descartes would have ruled out the possibility of being wrong about one's occurrent states. The skeptical problem about the contents of one's own mind is thus a real one for the Cartesian and the non-Cartesian alike. It can be solved, however, without embracing theism or functionalism.

IV. Two More Kinds of Failure of Transparency

Before turning to the two additional arguments against transparency, let me say something more about transparent

---

15. At the outset of the Third Discourse, at AT VI 23.

access itself. I have transparent access to properties of my mental states (so it is enough to say for present purposes) if I believe that my states have them when they do. Normally, this is enough for knowledge of them. Transparency fails when there are mental states I am in without believing I am. This failure could conceivably come about through an incorrect belief or simply through an absence of belief. Both kinds of failure would seem to be possible, and both kinds, as I will show, are relevant to refuting R.

Transparent access is thus distinct from incorrigible access; for present purposes, if I have incorrigible access to my mental states, then when I believe a mental state of mine has a property, it does. It is controversial whether we have incorrigible access to properties of our mental states, but questions of transparent access can in some cases be less controversial. Jackson's Mary provides such an example. Before her release Mary's phenomenal states are non-red phenomenally but she lacks knowledge by acquaintance that they are. This is a transparency failure through an absence, rather than an error, of belief. There is no counterpart failure of incorrigibility here, although there are, of course, other transparency failures in other cases which do entail failures of incorrigibility.[16]

---

16. Of the kinds of transparency failures I claim to contradict R, only the first, Case 1, invariably involves a failure of incorrigibility.

Drawing on further kinds of failure of transparency, I will now make two more objections to using R against absent qualia. In the first case, I will provide a kind of transparency failure which, like misconfirmation, contradicts R's requirement of comprehensive distinguishability of alternative qualia without jeopardizing our ability to know the qualia we have. However, it differs from misconfirmation in how it reconciles such knowledge with the indistinguishability of alternatives. In the second case, the transparency failure does not contradict R but rather the further premise that we know our mental states. Thus, we are in no position to use R to support K, since doing so requires the truth of both R and R's antecedent, the claim of knowledge.

Complexity. It would seem possible to lack knowledge about the qualitative character of a visual appearance just because of the complexity of the appearance.

> Case 2. Missing Waldo. A subject is presented with two pictures of the kind found in picture books like Where's Waldo?[17] These books are filled with very complex cartoon drawings, and the puzzle for the reader is to locate the cartoon character Waldo in the drawings. In the case at hand, Waldo is the only figure in the drawings with any purple. In one picture, his watchband is purple; however, in the other, his watchband is red. Other than this one

---

17. Martin Handford, Where's Waldo? (New York: Little, Brown, 1987).

> difference, the pictures are identical. There are enough nearby colors on the color spectrum like blue and red that the purple is of little help to the subject in locating Waldo or seeing that the pictures differ. The subject notices the purple on the one picture but never notices its absence on the other. The subject never finds Waldo.

Here is a case in which it is does not seem transparent to the subject that the picture lacking purple does lack it, even though the subject's visual field, while the subject stares at the picture, lacks the phenomenal property of being purple. At the same time, there would seem to be no barrier to the subject's noticing purple in the picture that has it that the subject is unable to distinguish the picture that lacks it. Thus, contrary to R, the subject may know that his visual field has phenomenal purple even though, due to this failure of transparency from complexity, he is unable to distinguish from it the case of a visual field lacking phenomenal purple. For he may be having such a visual field without believing himself to be.

One way of rejecting these intuitions is to hold that it is not determinate that a subject's visual field is as of purple or as of red at Waldo's watchband until the subject notices. Again, I grant that we should allow the possibility of deciding a posteriori after sufficient data collection and theory construction that, all things considered, such a thing is indeterminate until the subject

notices. The anti-skepticism argument against absent qualia, however, requires a priori, or at least common sense, reasons for rejecting the Waldo intuitions, but common sense supports these intuitions. One can surely imagine being made to notice the red watchband and sincerely saying, "That's it--that's what made it look different!"[18]

Subthreshhold phenomenal change. Now consider the following sort of phenomenon.

> Case 3. Pain Change. A subject is introduced to pain from an external pain machine that can be increased in intensity. Intuitively, we know that increases in intensity correlate with increases in level of pain. However, it is observed that there are increases in the cause's intensity that are not noticed. They fall below a increase-rate theshhold required for being noticed. With each subthreshhold increase, a subject will attest to no change, even though cumulatively the increases eventually cause severe discomfort and the kinds of behavior associated with it.[19]

It may sometimes be natural in such a case to say that there are phenomenal increases in the subject's pain even though the subject does not believe there to be. One might well be troubled morally by having increased the pain machine's output even if the subject professes not to notice an

---

18. Ned Block suggested this example to me.

19. For discussion of this case, see Parfit, Reasons and Persons (Oxford: Oxford University Press, 1984), pp. 78-79, and R. M. Hare, "Pain and Evil," Proceedings of the Aristotelian Society, Suppl. Vol. 38 (1964).

increase in pain. This sort of case is a counterexample to R's consequent: one is unable to distinguish no increase in pain from a subthreshhold increase. Here, however, we do not ascribe to ourselves the knowledge either. Thus, R survives the example. But the Knowledge Premise does not. Thus, again, the case for K, and against absent qualia, is undermined.

Thus, R can be used to discredit the possibility of absent qualia. I sometimes know a state of mine has a certain phenomenal property even though I do not, because of these failures of transparency, have comprehensive evidence for the property that distinguishes the state I have from every state lacking the property. The relevant alternatives I must discriminate it from in order to know the phenomenal property to be instantiated are fewer than this. And even when R is true, there are times, as perhaps in the case of subthreshhold pain change, when R's antecedent is false, and again the requirement of comprehensive evidence fails. Thus, the conjunction of R and the proposition that we do know our qualia entails that our knowledge requires distinguishing that is in fact irrelevant to having knowledge. If ruling out the possibility of absent qualia requires that my knowledge of phenomenal properties must be grounded on evidence so comprehensive that the number of relevant alternatives is the maximum conceivable, then the case against absent qualia collapses.

V. Reliabilism, Relevance and Saving the Argument

Can R be improved? Is there any way out of these difficulties for the defender of an argument against the possibility of absent qualia based on a threat of skepticism? Just as reliabilist intuitions about the compatibility of knowledge of the external world with the possibility of error undermined the epistemological principles we considered prior to introducing R, so further reliabilist intuitions have undermined R itself. It may help to begin with those intuitions.

What constraints does a plausible reliabilist account place on a principle like R? There are many reliabilist proposals about the nature of justified belief and knowledge, and they differ along a number of dimensions.[20] I shall not try here to adjudicate among them; instead, I shall review several issues that are pertinent to the reliabilist analysis of the special case of our knowledge of our qualitative mental states.

Any account of knowledge must recognize that many cases of knowing <u>that</u> entail knowing <u>which</u>.[21] My knowing that my friend will meet me in Harvard Square entails my knowing

---

20. Alvin Goldman reviews the literature on this approach in his <u>Epistemology and Cognition</u> (Cambridge, Mass.: Harvard, 1986), esp. chs. 3, 5.

21. See Alvin Goldman, "Discrimination and Perceptual Knowledge," <u>Journal of Philosophy</u> 73 (1976), pp. 771-791.

which square Harvard Square is. The knowledge-producing mechanisms to which the reliabilist draws our attention are mechanisms by which I know which things my knowledge is about. If I know which things satisfy a predicate, I can distinguish those things which do from many alternatives to them. A reasonable requirement, then, on the reliabilist's account of my knowing that something is psi is that my knowing-producing mechanisms distinguish something's being psi from relevant alternatives to something's being psi. My belief that my friend will meet me in Harvard Square is knowledge only if the mechanism producing this belief in me reliably distinguishes being in Harvard Square from being in relevantly alternative locations.

By way of illustration, contrast this condition on knowledge with the weaker condition I previously discarded, which I will call the Maximal Evidence Principle.

> The Maximal Evidence Principle
> For all p, if S knows that p is true, then S has evidence that distinguishes the truth of p from any possible case of p's not being true.

For me to know that my friend will meet me in Harvard Square, the Maximal Evidence Principle requires that I distinguish the event of my friend's meeting me in Harvard Square from all possible alternatives to that event. This includes such alternatives as my friend's being ill and staying home, my friend's meeting me in Inman Square, and my

friend's meeting me at my home. The Maximal Evidence Principle does not, however, require me to distinguish the Harvard Square meeting from a meeting in a square that looks almost like Harvard Square, so long as meeting at such a location is not a possible way for it to fail to be true that my friend meets me in Harvard Square. But the "relevant alternatives" condition might well require this. And it seems right to. It is a plausible condition on my knowledge that I know which square Harvard Square is. Knowing which could require a capacity to distinguish it from relevant alternatives similar in appearance even if our meeting in a lookalike square is not a possible way for my friend and me to fail to meet in Harvard Square.[22]

R is actually consistent with these reliabilist intuitions, and it thus has much going for it. Unlike the Maximal Evidence Principle, R employs the stronger "relevant alternatives" condition. It employs a maximal criterion of relevance: knowledge of having a phenomenal property requires one to distinguish having it from every possible case of not having it. Thus, R does not make the mistake of restricting relevant alternatives to actual alternatives. In this, R is unlike the reliabilist principle which Shoemaker, for example, sees as a possible threat to the

---

22. See Goldman, "Discrimination and Perceptual Knowledge," op. cit., and Epistemology and Cognition, op. cit., pp. 45-46.

anti-skepticism argument.[23]

To see this mistake, consider the following case.

> Case 4. Accidentally Accurate Pain Report. Consider again the mechanism which causes me to say, "I am in pain," on the basis of an expectation of pain and a sensation of cold or pain. Imagine a possible world in which the mechanism does not come into use very often but, on those rare occasions when it does, gets the sensation correct--I am in pain. Now, imagine that this is accidental, the result of some fortuitous correlation between sensations of cold and the absence of phenomenal expectations of various sorts, including expectations of pain. It is in this way by accident that the mechanism gets my pain-states correct; there turn out to be no opportunities for it to get them wrong.

Although the mechanism reliably allows me to distinguish between pain and the alternatives to it actualized in this possible world, the mechanism does not give me knowledge. This lack of knowledge becomes evident when I move to nearby worlds where the fortuitous correlation no longer holds. There, I cannot on the basis of this mechanism distinguish anymore between genuine pain and relevant alternatives such as those sensations of cold which trigger my reports of pain. It is for this reason that the belief-producing mechanism does not give me knowledge. It is a mere accident

---

23. Shoemaker, "Absent Qualia Are Impossible," op. cit., p. 596.

that I am correct in the actual world. For there to be knowledge of pain, the relevant alternatives such a mechanism must reliably distinguish from my genuine pains must go beyond actual states to include such counterfactual states as those I undergo in these nearby worlds.

Despite these virtues of R--its respect for relevant alternatives and its treatment of counterfactual states as relevant--R remains inadequate as a condition on knowledge. Again, R makes <u>all</u> alternatives relevant, and that is too many. We presently lack a theory of relevance, but I have already said enough to suggest some alternatives that are not among them. The case of misconfirmation, for example, shows that it is irrelevant to my knowledge of my pain that I distinguish between sensations of pain and any and all sensations of cold which would trigger avowals of pain. Anyone who argues that the argument I have been criticizing can be saved by modifying R has to show that were there, by hypothesis, to be any ersatz pains, they would be among the relevant alternatives covered by the modified R. That is, they must be among the relevant alternatives to my actual phenomenal states that I must be able to distinguish my actual states from in order to have know of my actual states. I do not have any argument that this cannot be done, but it seems clear that nobody as yet has met this burden of placing ersatz states among these relevant alternatives.

We have no guarantee that ersatz pains themselves are not simply among the counterexamples to R. Granted, they do not fall into the paradigms of counterexamples we have examined so far. They do not involve the same sort of abnormal belief-causing processes misconfirmation does. What makes them ersatz is that they are functionally isomorphic to genuine states; thus they are parts of otherwise normal, belief-causing processes. Nor are they experiences that match the other paradigm, ones which, due to complexity or change that falls below a threshhold, we are unaware of the phenomenal properties of. For ersatz pains are not experiences; nor are they necessarily aspects of complex mental states or subthreshhold phenomenal changes. Still, even though ersatz states do not fall into these paradigms, it is unclear why we should deny the possibility of further paradigms of counterexamples to R. The initial, robust intuition expressed through R is after all unsound. There may, of course, be special cases of R which one cannot give up--for example, the special case of R according to which my knowledge of having a <u>longstanding</u> pain requires the possibility of my distinguishing having it from the alternative of having a <u>longstanding</u> feeling of cold. But I see no reason to believe that the intuitions governing such special cases generalize to cover absent qualia. Only if they do would we have reason to place ersatz states, were there any, among the relevant

alternatives which knowledge of qualia guarantees distinguishability of.

There are thus no general epistemological principles guaranteeing that we can always distinguish our normal qualitative states from nonqualitative states. Nor is there reason to believe that we can distinguish normal states by such principles from a subset of nonqualitative states which would include ersatz pains, if there were any. Anti-skepticism arguments against the possibility of absent qualia which depend on such general epistemological principles therefore fail.

I shall argue in Chapter Eight that even without such general principles, we can nevertheless distinguish our genuine states from ersatz counterparts on other grounds. However, I shall go on to argue that this is consistent with the existence of enough absent-qualia states to undermine functionalism. The reason for this is that not all absent-qualia states threaten the kind of skepticism their functionalist critics have alleged.

To make this plausible, I shall devote most of the present chapter and Chapter Eight to reviewing and rejecting an objection to the functionalist account superficially similar to mine. After showing what it wrong with the superficially similar objection, I shall use the lessons developed to create at the end of Chapter Eight a different, more successful objection.

The superficially similar objection is that it is

enough to forestall the functionalist critic that we distinguish absent-qualia states that share the nonqualitative causes and effects of genuine states. The objector claims that this leaves room for detectable qualitative differences between the two sets of states by way of the beliefs they cause. In the present chapter, I examine and reject the notion that it would be enough for this to occur that the respective beliefs differed solely in their wide contents. This means returning to the question I began examining in the fourth and fifth chapters, that of how qualitative beliefs get the contents they have.

I. The First Objection: The Conee-Shoemaker Version

Recall the specific anti-skepticism argument against the possibility of absent qualia due to Sydney Shoemaker and Earl Conee. As Conee reconstructs Shoemaker's original argument, recall that there are two premises and a conclusion, derived by modus tollens.

Conee's Reconstruction of the Anti-Skepticism Argument

(EP) If ersatz pain is possible, then it is not possible to distinguish cases of genuine pain from cases of ersatz pain.

(K) It is possible to distinguish cases of genuine pain from cases of any possible state lacking qualitative character.

Ergo, ersatz pain is not possible.

Conee argues that premise K is defensible. Although I refuted his argument in the previous chapter, I shall assume here provisionally for the sake of argument that there is some much weaker version of K that would produce a logically valid argument and is defensible. This would push the burden of the modus tollens argument onto EP, according to which the existence of absent qualia would make it impossible to distinguish genuine from ersatz pains on the basis of the presence or absence of qualitative character. It is with EP that Conee finds fault. I will call the kind of case he makes against EP "the First Objection."

For EP to be defeated, the presence of the qualitative character of pain must make a difference in the causal capacities of pain, a difference that would enable us to distinguish genuine pains from ersatz cases. But what could this difference be?

The First Objection, as I have indicated, supposes that it is a difference in the beliefs they cause. Ordinarily, if I distinguish two things, it is in virtue of the beliefs caused by them differing. Can the beliefs caused by two functionally equivalent states differ, however? In a sense, yes. Recall Earth and Twin Earth. Assume that I have a doppelganger on Twin Earth who has functionally equivalent beliefs to mine. He believes he is sitting in front of a word processor; I believe I am sitting in front of one. No difference yet. But he believes he had a glass of XYZ this

morning, I believe I had a glass of H_2O, even though we each formulate our recollections with the word "water." Here we do have a difference, one in virtue of the different Russellian propositions expressed by each of us in reporting our beliefs, or as I will say here, in virtue of the different wide contents of the respective beliefs.[1]

Now imagine a person Smith who feels a genuine pain, g, with qualitative character, c, and then undergoes an ersatz pain, e, with no qualitative character. By introspecting the genuine pain's character, c, Smith could come to know:

(Bg) The state I am in which I believe to be pain presents this character to introspection (making direct reference to c).

Can Smith accomplish something similar with his hypothetical ersatz pain? Can he know that his ersatz state presents some character to introspection? Conee writes: "Giving attention to a mental state one is in, attempting to introspect some qualitative character, and failing to find any, is a mental process that includes experience of a certain phenomenal character--one has the seeking-and-finding-no-feeling sort of experience." Call this an experience of character n. While in e, Conee asserts, Smith could introspect and come to know:

---

1. Of course, there is also the wide-content difference in his beliefs being about him, my beliefs being about me.

(Be) The state I am in which I believe to be pain presents this character to introspection (making direct reference to n).

Conee argues: "Suppose that Smith introspects these things and that as a result he gains beliefs constituting knowledge of Bg and Be. Generating these two different items of knowledge is a causal difference between g and e. By knowing Bg and Be, Smith distinguishes the genuine case of pain from the ersatz."

If he is correct, this is a counterexample to EP, since according to EP the possibility of ersatz pain precludes a causal difference enabling one to distinguish between it and genuine pain. And with such a refutation of EP, the conclusion of the modus tollens argument, that ersatz pain is not possible, would also be refuted.[2]

Conee's case rests in part on what the functionalist is committed to concerning functional definitions of phenomenal states. The functionalist claims not just that mental states can be interdefined but that phenomenal character can be explained away. That means that any functional definition which refers in any way to phenomenal character can only be provisional. The functionalist must in principle be able to produce functional definitions that eliminate explicit reference to phenomenal states

---

2. Conee, op. cit., pp. 354-356.

altogether. Thus, the anti-functionalist has only to show the possibility of an ersatz state that satisfies a nonqualitative functional characterization of a genuine state.

Conee's argument tries to exploit this fact. It may seem that the relation between pain and pain-belief is too close and the character of some pain-belief too phenomenal to allow the possibility of a state that would cause genuine pain-belief but itself have no phenomenal character. An ersatz state characterized nonqualitatively, one might believe, would allow more room. Conee's point is that to defeat functionalism it is enough that it be possible to produce functional isomorphs of pain and pain-belief that are qualia-free in certain ways. It would be enough to produce a qualia-free isomorph of pain that has causal relations with a state isomorphic to pain-belief but directed toward the qualia-free state in the way pain-belief is directed toward pain.

Conee contends that this is enough to defeat the anti-skepticism argument. With their functional equivalence, the two beliefs would be similar enough to preserve the ersatzness of the ersatz state. But with their wide-content difference, they would be different enough to distinguish genuine from ersatz. His claim is that it is possible to do this within the constraint against skepticism because the very qualitative difference between the two beliefs, one

genuine and the other qualia-free, makes it possible to distinguish them.

II. Why the First Objection Won't Work--A Summary

In this section I will summarize my counterargument to the First Objection before going on to set out my counterargument is considerably more detail. At the end of this section I will set out the substance of the Second Objection, which I will defend in more detail at the end of the chapter.

My argument against the First Objection, in summary, is that there is no ersatz state having an effect like belief in Be that could both distinguish the ersatz state from a genuine counterpart while, at the same time, preserving the functional isomorphism between the ersatz and genuine states. The reason is this. *Any belief like Be in virtue of which such distinguishing could, by hypothesis, occur would require mental processing that would have to go outside the set of normal effects of the genuine state, thus upsetting the functional isomorphism between it and any hypothetical ersatz state causing such a distinguishing belief.*

My counterargument to the First Objection is general. There is no way to repair the First Objection by identifying a different pair of hypothetical effects of genuine and

ersatz pains which would escape the difficulties on which my counterargument is based and distinguish the genuine from the ersatz pains. For the proponent of the First Objection is committed to claiming not only that Smith's e-related belief, Be, can distinguish genuine pain from ersatz pain but also that it must distinguish the two. Even if there were another pair of beliefs caused by Smith's genuine pain and his hypothetical ersatz pain which escaped the difficulties created by the beliefs Bg and Be, still these difficulties would be enough to undermine the First Objection. For if there is at all to exist an ersatz counterpart to genuine pain such as Smith's state e, there has to be a distinguishing counterpart to his belief in Bg, like Be. And I argue there cannot be one.

Now, it obviously would not help the First Objection to construe Bg and Be, Smith's two beliefs in virtue of which he is supposed to distinguish genuine from ersatz pains, as wholly nonqualitative beliefs. Smith's genuine pain, g, and e, Smith's hypothetical ersatz pain, have all their causal relations in common to nonqualitative beliefs, other direct nonqualitative effects and any indirect effects of those beliefs and other direct effects. The ersatz state e and its nonqualitative effects will cause or prevent nonqualitative beliefs of a given type under any circumstances g and its nonqualitative effects would. Thus, there would be nothing between belief in Bg and belief in Be

to distinguish g and e if these beliefs were wholly nonqualitative.

But it also will not help to adopt a certain false picture of what would make these two beliefs qualitative beliefs--that is, of what would make them the kinds of beliefs that would make it possible to distinguish g and e. According to this picture, which I examine in the next section, what makes them qualitative beliefs, and thus the kinds of effects of g and e that would make it possible to distinguish g and e, is that their referring expressions are satisfied by qualitative characters. On this picture, qualitative beliefs are just like nonqualitative beliefs except that their referring expressions are satisfied by qualitative characters. Thus, it is possible to distinguish g and e, on this picture, because the two respective effects of g and e, namely Bg and Be, differ in wide content. The former is satisfied by c, the phenomenal character of Smith's genuine pain, the latter by n, the hypothetical phenomenal character of Smith's hypothetical ersatz pain.

If it were as simple as this, there would be a way out of the difficulties that I describe below that make it impossible both to make room for a distinguishable qualitative difference between the effects of g and e and to preserve the functional isomorphism among the nonqualitative effects of g and e. But it is not as simple as this. For I will argue that if belief in Bg and belief in Be differed

only in what satisfied their referring expressions, it would be possible to know when they did differ only so long as there were some other way to distinguish the referents. The difference in referents between Bg and Be would not be enough by itself to distinguish g from e, genuine from ersatz, which the First Objection requires.

It may seem to help the proponent of the First Objection to make the referring expressions contained in the two beliefs demonstratives and the reference to their objects direct. It does help, I will argue, but still, only if there is something else besides the difference in referents that distinguishes the beliefs.

There is something else. It is this. In demonstrative reference to qualitative character we humans pick it out qualitatively--from the inside, in virtue of what it is like to have mental states with it. We do not pick it out in virtue of public aspects of the qualitative character, such as physical or functional properties, as other creatures might. This, then, is how the proponent of the First Objection would have to distinguish g from e through Bg and Be, if g and e could be distinguished. If there is a difference between the beliefs Bg and Be that makes possible distinguishing g from e beyond the mere difference in referents, it is this. The phenomenal characters of g and e are both picked out qualitatively and the picking-out of g's character differs qualitatively from the picking-out of e's

character.

But I argue that it is not possible to distinguish them this way. In general, the ways qualitative characters get picked out insure that there are qualitative aspects to qualitative beliefs beyond their wide contents. But these further qualitative aspects, because of differing causal relations to them, upset any possible functional isomorphism between a genuine state and any hypothetical ersatz state one might try to construct. This makes ersatz counterparts with qualitative causal relations impossible.

There are two reasons for this. First, suppose that there were a state e distinct from g in Conee's way--because Be was distinguishable from Bg. To construct such a hypothetical e, I will argue, we would be forced to choose between an unacceptible epiphenomenality, in virtue of which qualitative states lack nonqualitative effects, and an impossible irrationality, one that would have to fix false beliefs in the face of conclusive evidence of their falsity.

The problem arises because certain nonqualitative effects that an e would have that would have no counterparts among g's effects. Normally, for example, states like e and Be would cause perplexity. By hypothesis, Be reflects Smith's experience that there is a state he simultaneously believes to painful and feels to be painless. But neither g nor its immediate effects, like belief in Bg, cause such perplexity. Thus, I argue, either belief in Be cannot have

the normal effect of prompting perplexity, or it has that effect at the same time it fills the functional causal role of belief in Bg, its counterpart among genuine pain's effects. In the first case, the result would be unacceptible epiphenomenality, in virtue of which qualitative states have no nonqualitative effects.

In the second case, the result would be the impossible irrationality referred to. In order to maintain the functional isomorphism of e with g, belief in Be must have spurious effects in nonqualitative belief, namely those of belief in Bg, belief in painfulness. But whatever processes fix these spurious beliefs would conflict with the counterveiling evidence of the normal nonqualitative beliefs caused by a painless state. This would lead not to an acceptible form of irrationality but to an impossible state, for the latter, nonspurious beliefs would undercut the processes fixing the former, spurious ones. If it were possible to overcome the counterveiling evidence of the nonspurious beliefs somehow, it would only be so by use of processes outside the functional causal role that g and e share, processes that would not fit the template of causes and effects that e and its effects must conform to.

Second, I argue, if Smith's hypothetical state of belief Be were to exist, it could not have intensional content at all, and thus it could not constitute a kind of knowledge through which Smith could distinguish e from g.

This is because the demonstrative term "that character," which purportedly picks out e's phenomenal character, in fact cannot pick out a demonstratum at all. Suppose that a demonstratum could be picked out in one of the following two ways--either through an introspective search or in virtue of a demonstratum just, so to speak, "calling attention to itself." But a "location problem," as I dub it, makes it impossible to find the term's referent through an introspective search, since there is no counterpart search which is part of g's functional role. This problem, I will argue, precludes assigning the term a referent directly. The belief Bg gets a referent this way, in virtue of c, g's character and Bg's referent, calling attention to itself by filling a specific region of phenomenal space and by standing out in contrast to its phenomenal background. But e has no such qualitative character that could do this.

For there to be such a functional fit between them, g and e would need to have a relationship analogous to that which phenomenal states have in cases of spectrum inversion. In inversion cases, however, counterpart phenomenal states are functionally equivalent and differ only in qualitative character. By contrast, g and any hypothetical ersatz state like e would differ in more respects.

Both of these problems--the epiphenomenality-irrationality problem and the location problem--work against the existence of e and Be for the same kind of reason. No

ersatz state could both conform to the template of normal effects and effects of effects of a genuine state like g and, at the same time, be distinguishable introspectively from that state in virtue of a qualitative difference between the states. There would be side effects running outside the functional template.

All these difficulties depend on functional differences in the qualitative ways g and e would be picked out. There would be no such difficulties if e could be distinguished from g nonqualitatively, merely in virtue of a difference in referents among the beliefs they cause. That is the appeal of trying to distinguish e this latter way, but as I have previously claimed and will argue in more detail below, it is not possible to distinguish e from g in this way either.

For these reasons, the First Objection is unsuccessful. However, a related Second Objection is available that suffers none of the problems that defeat the First Objection. The Second Objection is successful.

Consider Smith's nonsentient, homunculus-headed doppelganger, an entity all of whose mental states lack qualitative content. Between Smith and his doppelganger, distinguishing of the weak sort required by the anti-skepticism argument is only possible in Smith. The doppelganger could not distinguish, since the doppelganger, being nonsentient, could not possibly pick out the qualia-free contents of its states within the constraints of the

functional isomorphism to Smith it by definition has. In a word, it's a zombie at best and doesn't know anything. Now compare Smith and the doppelganger. In Smith, K is true--the presence or absence of the qualitative character of pain most certainly makes it possible to distinguish cases of genuine pain from cases of any possible state lacking qualitative character. For as I have argued in the previous sections, Smith's genuine pain states cause qualitative pain-knowledge but no ersatz pains are available to him to produce anything that he could confuse with genuine pains. (This is so even if, as I argued in chapter six, there is no more general epistemological principle of transparency supporting K.) The existence of the doppelganger does not contradict K, since its states do not have raw feels to be distinguished from the absence of the qualitative character of pain and its ersatz states cause nonqualitative pain-belief-like states that do not constitute knowledge at all. But these facts are compatible with the possibility of ersatz pains, since the doppelganger by definition has them. Thus, the possibility of ersatz pain does not entail, contrary to the functionalist, the inability to distinguish ersatz from genuine states.

III. Distinguishing Ersatz States by the Wide Content of Beliefs about Their Qualitative Character

I shall now consider at greater length the first option discussed in the previous section for characterizing qualitative beliefs. That is the option of construing them, to use the rough approximation there, as wholly nonqualitative but for the satisfaction of their referring expressions by the raw feels of real phenomenal states. With this first option, then, Bg and Be would differ, roughly, only in the referents of their referring expressions and would have no qualitative character except that of the referents. This option would provide a qualitative difference between g and e, but it would do so without the difficulty with the side-effects of qualitative character. Thus, it would provide support for the First Objection to the anti-absent-qualia argument.

How could Bg and Be differ only in their referents? I called this a rough approximation of the option. That is so because Bg and Be surely must differ in some other respect in order to differ in their referents. I glance over to a glass of clear liquid sitting near my word processor and say, "I had a glass of that stuff this morning." My Twin-Earth doppelganger makes a similar glance and utters a similar-sounding sentence. The beliefs he and I express differ in their referents, water and XYZ, but this is so

only because there are further differences--for example, in their demonstrations (my finger versus his), in their relations to previous samples of clear liquid, and so on.

One natural way to make the distinction is to say that what is "in the head" on these two occasions is the same and that what differs is what is "outside the head." Among the things "in the head" in virtue of which the beliefs are similar are their syntactic structures, their constituents, and their causes and effects. Among the things "outside the head" in virtue of which they differ are the two referents--the water and the XYZ--and my two different relations to them.

In an important way, however, my beliefs about the water and his about XYZ fail to parallel Bg and Be. By hypothesis, Bg and Be, unlike the first two beliefs, do differ in what is "in the head." Thus, I shall not discuss what is "inside" or "outside" Smith's head. Instead, I shall characterize this first proposal this way: that because the qualitative character of their referents is the only qualitative aspect of Bg and Be, the wide-content difference between Bg and Be is the only qualitative difference between them. If this were the only qualitative difference, there would be no qualitative side-effects to picking out their referents to upset their functionalism isomorphism and make absent qualia impossible. This is what might make this proposal attractive to the defender of the

First Objection.

I shall now provide two reasons why this proposal would fail to make of Bg and Be effects that could distinguish their hypothetical causes, g and e.

I shall call referential relations to the qualitative characters or raw feels of g and e--that is, c and, by hypothesis, n--that would fit such an account as this simple. A simple referential relation would pick out qualitative referents for Bg's and Be's demonstratives nonqualitatively. Thus, as I have said, it would do so without the complicating side-effects of picking them out qualitatively that give rise to the "epiphenomenality-irrationality" and the "location" problems.

Now, let us look for a moment at a way, one I will reject, in which one might interpret the demonstratives of Bg and Be to have simple referential relations to the qualitative characters that are supposed, by hypothesis, to satisfy them. According to this interpretation, first suggested by Stephen Schiffer, demonstratives are a disguised form of definite description, one expressing individual concepts. Say, for example, that Tom believes (1) true of some cup, by the very words of (1).

(1) *That cup* is red.

According to the view under consideration, Tom's belief is identical to some belief that could be expressed in words

that replace (1)'s subject with a definite description, one which expresses an individual concept entirely through general terms and the logically proper "I" and "now." Thus, it has been suggested that (2) expresses the same belief as (1).

(2) *The only object which I am now looking at which appears to me to be a cup* is red.[3]

None of the authors who have taken this view have discussed what one should do with demonstrative beliefs about bodily sensations, such as Bg and Be. Let us speculate about what proponents of the view could say. It need not be a true account; the question before us is whether any such account as this would be adequate for distinguishing Bg from Be and refuting Shoemaker's argument. What we need is a belief that bears the same relation to Bg or Be that (2) bears to (1). Let me propose (3) as a belief that proponents of the descriptivist account of demonstrative belief might take to be fill that role.

(Bg) The state I am in which I believe to be pain presents to introspection *this character* (by hypothesis, making reference to c, or in the case of Be, to n).

---

3. Schiffer's view is set out in "The Basis of Reference," Erkenntnis (1978), pp. 171-206. For criticisms of it, see the next section. Beliefs identical to (1) and (2) are discussed in Kent Bach, "De re Belief and Methodological Solipsism," in Andrew Woodfield, ed., Thought and Object (Oxford: Clarendon Press, 1982), p. 140.

(3) The state I am in which I believe to be pain presents to introspection *the character which I am introspectively attending to now* (by hypothesis, making reference to c, or in the case of Be, to n).

If Bg or Be is construed to express a meaning like that expressed by (3), then the wide-content difference between Bg and Be, I shall now argue, cannot by itself distinguish g from e.

First, suppose that Bg is semantically equivalent to (3) or something like it. Unless there is a further piece of knowledge beyond Bg by which Smith can identify the character which he is attending to as the particular qualitative character it is, not by some further description but directly, then there would be no way for Smith to distinguish g from e on the basis of Bg and Be. For there is by e's very definition nothing between the two beliefs to distinguish them by way of descriptions. Their functional equivalence connects them to the same descriptions.

Suppose that Smith knows (3) to be true and knows it to be true in the very words of (3). Suppose he knows that the state he is in which he believes to be pain has the qualitative character he is contemporaneously directing his attention to. He does not yet, however, know that the state he is in which he believes to be pain has any particular qualitative character--this one or that one--unless he also knows of some particular this one or that one that it is the

one he is contemporaneously directing his attention to. For it is contingent of any particular qualitative character that it is the character that Smith is directing his attention to contemporaneously (from Smith's point of view, now), and it is a substantial piece of further knowledge for Smith that c or n fits that description.

There is a further problem. Descriptions containing psychological terms, like the description in (3), cannot provide distinguishing beliefs, and it is hard to see how to improve on them. Suppose that there exists an ersatz pain of the form Smith is supposed to undergo and that he directs his attention to his bodily sensations to search for the character of the state causing his pain behavior. Suppose that he picks out a qualitative character that makes Be true in virtue of some description, like that in (3) or by employing some other psychological relation besides that of attending to describe the character. Either he must employ a further description to place the character in the psychological relation (for how does it get to be true that Smith attends, for example, to that qualitative character?) or there must be some reason that attending differs from believing in taking one directly to an object, without any description.[4]

But neither option--that of a further description or

---

4. For more on this dilemma, see Brian Loar, Mind and Meaning (Cambridge: Cambridge University Press, 1981), p. 104, and Austin, op. cit., esp. pp. 38-39.

that of a psychological difference--is satisfactory for distinguishing e from g in virtue of Be and Bg. For any further description, on the one hand, there would remain the further knowledge of what the description was satisfied by. Say his search leads Smith to the phenomenal location of the genuine pain counterpart. If it makes sense to characterize Smith's attending to the raw feel at that location as by description, this would be so without his knowing which raw feel it is. For attending to it by description, if there were such a thing, would require attending to it in virtue of its satisfying some description. If some piece of knowing something about it is a case of knowing under some description about its being attended to, then there will always be the further knowledge of its being the thing attended to. This is knowledge de re. Without this latter, there would be no distinguishing of the sort the First Objection requires.

But on the other hand, any psychological difference between attending and believing will still require a direct psychological relation to n, the hypothetical ersatz character. That will create the other set of problems--the location problem and the epiphenomenality-irrationality problem--that we have been looking to the simple referential relation between Bg and c and between Be and n to alleviate. To make our knowledge of qualia depend upon direct attendings or any other direct psychological connection

would be to require the kinds of mental processing that would upset the intended functional isomorphism between the genuine pain and the hypothetical ersatz counterpart. To attend to e's character n directly, for example, would surely require a search for qualia for which there is no counterpart with the genuine state g or its character c.

IV. Direct Demonstrative Reference and Qualitative Belief: Two Incorrect Accounts

Smith's beliefs in Bg and in Be are not merely in causal and referential relations with c, the qualitative character of g, and n, the hypothetical qualitative character of e. Rather, Conee writes, they are also in relations of direct reference. If my arguments are correct, Conee is right to stipulate them to be relations of direct reference. Only with knowledge by direct reference beyond that expressed by nonqualitative descriptions could one hope to find, if there were one at all, a qualitative difference in effects that could distinguish g from e.

By making this stipulation, Conee means at the very least to rule out any account according to which Smith fixes the reference of the demonstrative expression "that character" entirely through the mediation of some definite description, as was entertained in the last section. How then does the reference of the expression "that character" get fixed? I will not construct a full-blown theory of

direct reference, or of the related notion of de re belief. There are many accounts and each is controversial. It is adequate to look at several alternatives to determine if there is a defensible interpretation of Smith's situation fulfilling Conee's goals.

I shall argue in the remainder of this essay that there is not. In this section, I argue that beliefs making direct reference to qualitative character and occurring in normal human beings like ourselves and almost normal ones like Smith require qualitative aspects beyond their qualitative wide contents. If a difference between Smith's two beliefs beyond the wide-content difference made it possible to distinguish his genuine and ersatz states, it would be in virtue of the qualitative characters of the states being picked out qualitatively. And it would be in virtue of the picking out of one differing qualitatively from the picking out of the other. But since, as I argue in Chapter Eight, direct reference employs not the finesse of descriptive reference but force, and force has side-effects, there could be no e functional equivalent to genuine pains.

First, I shall consider two accounts that would lead to the conclusion that both beliefs Bg and Be must have qualitative aspects beyond their qualitative wide contents. I use several previous thought experiments to show that these accounts are false. Then, I argue that direct demonstrative beliefs about qualitative character must have

qualitative aspects beyond their qualitative wide contents when, like Bg and Be, they occur in humans like Smith or in other nonhuman sentient creatures capable of distinguishing qualia.

Bach's account. Consider a position advanced by Kent Bach.[5] Bach opens his argument by properly criticizing the account of direct reference put forth by Stephen Schiffer, discussed above.[6] Schiffer's view, as Bach notes, can be interpreted as the view that the mode of presentation under which a perceptual belief comes to be about some object is an individual concept as expressed by a definite description containing "I" and "now."

Bach and Schiffer agree that there are modes of presentation, but on Bach's alternative account to Schiffer's, modes of presentation are viewed as "percepts" rather than individual concepts. "Intuitively," Bach writes, "the trouble with Schiffer's view is that to believe something of an object one is perceiving does not require thinking of it under any description at all, for it is already singled out for one perceptually.... The content of a perceptual belief, like that of any de re belief, is not a proposition expressed by a closed sentence. Rather, its content is expressed by an open sentence with the percept functioning as a mental indexical."[7]

---

5. In Bach, op. cit., pp. 139-149.
6. In Schiffer, op. cit.
7. Bach, op. cit., pp. 143-146.

To explicate this, Bach introduces the schema "$A_{\underline{m}}f_{\underline{x}}\underline{s}$" to represent a person $\underline{s}$'s being in a certain type of perceptual state of being "appeared$_{\underline{m}}$ to $\underline{f}$-ly," where "$\underline{m}$" ranges over sense modalities (visual, tactile, olfactory, etc.) and "$\underline{f}$" ranges over manners of appearance. It is a schema to represent the contents of perceptual states. According to Bach's account, "the conceptual content of [$\underline{s}$'s] belief, as expressed by the open sentence '$\underline{x}$ is G', applies to an object only if there is an object which is perceptually causing $\underline{s}$ to be in a perceptual state with content '$A_{\underline{m}}f_{\underline{x}}\underline{s}$', in which case the belief is about that object."[8] For someone to have a perceptual belief, then, that person must be, according to Bach, appeared to visually, tactually, olfactorily, or by some other sense modality in which there is qualitative content.

Bach does not mention pain-beliefs or other introspective, phenomenal beliefs, but if Bach's account were true and could be extended to phenomenal beliefs, then this would constitute a picture of how there might be qualitative aspects to Smith's Bg and Be beyond the qualitative characters in their wide contents. But while it normally holds of perceptual beliefs, neither is Bach's account necessarily true of the perceptual beliefs it is intended to cover nor would it be necessarily true of introspective, phenomenal beliefs like pain-beliefs.

---

8. Ibid., p. 146.

If absent qualia are possible, then somebody might have a perceptual belief without there being qualitative character to that belief or to any of its states. Even if there are no absent qualia states, we might imagine somebody deprived of visual and auditory abilities to identify objects in her surroundings. Despite this handicap, imagine that she is able to reliably make true assertions about remote objects that seem to express perceptual beliefs about those objects. Or imagine a machine that is complex enough to be ascribed beliefs about its surroundings on the basis of its measurements of its surroundings. These two cases are more extreme than actual cases of blindsight and artificial intelligence but neither case is unimaginable. Both the person and the machine might make what seem to be demonstrative references to objects and substances in the environment seeming to express de re beliefs even though there is nothing qualitative to the demonstrations, whether in any modes of presentation or in any other aspects of the demonstrations.

As for the case of phenomenal beliefs like pain-beliefs, recall Marcy (from Chapter Five, section five), whose reports of her pains are unconnected to phenomenal evidence of being in pain. Let's imagine that she also does this for others--suppose she is wired up to them in a way that provides her with the ability to demonstratively refer to their pains. Suppose also that she refers

demonstratively to the character of their pains and does so without feeling any pain. Here, then, we have a case, although quite odd, of a belief with qualitative character in its wide content but without further qualitative aspects.

Conee's account. Conee asserts that any belief that uses a constant to refer to a token qualitative character or state is a "qualitative belief" and thus governed by the same principles governing other qualitative states. Since the beliefs Bg and Be have demonstrative expressions as constituents, and these expressions are constants, it is supposed to follow that Bg and Be are "qualitative" in Conee's sense.

But recall Jones (from the previous chapter), who can identify the token qualitative characters of his states not on the basis of appearance but on the basis of physical parameters, using a "cerebroscope" that allows him to observe and decipher his own brain states. He can refer to the token qualitative characters demonstratively, and he can express beliefs whose contents contain demonstrative references to the token qualitative characters, on the basis of observing his "cerebroscope." Again, one can imagine having demonstrative beliefs about features of one's internal qualitative states without being presented with the features qualitatively.[9]

---

9. For Conee's discussion, see Conee, op. cit., pp. 348-349.

The term "qualitative belief" is supposed both to pick out a belief that presents qualitative features qualitatively--as Conee wants Bg and Be to do in order to distinguish g from e for Smith
--and pick out a belief that employs constants to refer to a token qualitative state or feature. But Jones' beliefs show that the term could conceivably be unable to do both.

V. A Sound Argument

Despite these fanciful cases, what cannot occur are beliefs just like Smith's direct, demonstrative beliefs which have no qualitative aspects beyond the raw feels to which they refer.

Smith's beliefs are very different from the states I just discussed which, by contrast, do lack qualitative aspects beyond the raw feels to which they refer. If my previous arguments are sound, then it is conceivable that a nonsentient mechanism, or Marcy or Jones, might be able to pick out c and n in some nonqualitative way. The nonsentient mechanism might have belief-like states about c and n, let us suppose, not inferentially through these raw feels' satisfying physico-functional descriptions but rather through their being picked out more directly than that. The device might be caused to refer by interacting in the right causal way with c and n. Marcy might similarly "just know"

that some state had c or n; Jones might know it cerebroscopically. But Smith, a mostly normal sentient creature except for his ersatz states, does not identify c and n in a nonqualitative way. Smith picks out c and (if he does at all) n in ways different from those of the nonsentient mechanism or Marcy or Jones. Smith picks them out qualitatively. In a normal Smith, there would be qualitative aspects to any direct, demonstrative belief of his to the effect that he was undergoing a mental state with some particular raw feel, such as c or n.

Only if Smith has in his head apparatus that picks out qualitative features on the basis of nonqualitative properties of them (or like Marcy's, on the basis of no properties of them) would Smith's beliefs directly refer to qualitative features while lacking other qualitative aspects. But Smith--like you and I and anyone else biologically similar to us--does not have such apparatus. We do not seem, phenomenologically, to have de re beliefs whose only qualitative features are their referents. By itself, this is all right, since psychology has discovered all kinds of mental states that we never thought we had. The trouble here is that such beliefs would constitute a possible source of skepticism about something that seems indubitable.

Recall Kaplan's picture of demonstrative reference, which I discussed in Chapter Five and which supports these

intuitions. According to it, reference to qualia or any other demonstratum requires what I previously labeled as a scene of which the demonstratum is a part, a directing intention, and an externalization of the directing intention, such as a pointing. In the normal, first-person case of direct reference to the raw feel of a pain, where the demonstrator and the audience are identical and there is no explicit pointing, this would involve situating the referent at a phenomenal location against a phenomenal background and picking it out from its background. As I argued in Chapter Five, this is done in virtue of phenomenal properties of the referent. Even the causal story about such a belief with which the functionalist would <u>begin</u> would be heavily loaded with phenomenal language, the functionalist's hope being that such language could ultimately be dispensed with by reduction.

From Kaplan's picture, it is natural to conclude that there are two ways in which Smith's beliefs Bg and Be have qualitative aspects beyond the qualitative characters to which they refer. One we might call <u>intrinsic</u>. In virtue of their intrinsic qualitative aspects, Bg and (if such a belief were really possible) Be would, let us say, <u>incorporate</u> c and n into themselves. Bg and Be would not only have c and n as demonstrata but as constituents. They would produce direct psychological effects in virtue of the qualitative character of the states whose character they

were about. In both these ways, they would differ from belief-like simulations of Bg and Be that have raw feels like c and n as demonstrata but pick the raw feels out by nonqualitative properties of them. In these simulations, the demonstrata would not be constituents, and they would not necessarily produce effects qualitatively.

The other way in which Smith's beliefs Bg and Be have qualitative aspects beyond the qualitative characters to which they refer I will call relational. In demonstrating c to oneself, according to Kaplan's picture, one would pick c out from a background, and doing that would require relating c to some of the other qualia at phenomenal locations around it. Introspecting Bg, then, one would introspect the qualitative aspects of locating c.

One final issue. Both these ways in which Bg and Be have qualitative aspects beyond the raw feels to which they refer introduce their own special problems when we try to think how Bg and Be could be isomorphic.

So far I have shown that beliefs that normally refer demonstratively to raw feels or qualitative character--that is, in normal, sentient beings like Smith--do so in virtue of qualitative aspects of the beliefs beyond that of the raw feels or qualitative characters to which they refer. It does not seem to follow directly from this that beliefs in normal, sentient humans like Smith that differ in qualitative demonstrata must differ in other qualitative

aspects as well. It might conceivably be the case, for example, that certain types of "inverted spectrum" cases could be set up in such a way that demonstrative reference to a qualitative state and its qualitative inverse were structurally similar. And they might be similar enough so that, although there would be qualitative aspects to the demonstrative beliefs about the states beyond the qualitative aspects of the states themselves, the beliefs would <u>differ</u> qualitatively only in their referents.

Obviously, it would help the First Objection if between Bg and Be there were a weak kind of distinguishability of the sort one would find in the spectrum inversion examples, a sort of distinguishability strong enough, however, to defeat the functionalist's argument. Nevertheless, I will argue in Chapter Eight that Bg and Be are not like this. Because of what the demonstrative references in Bg and Be are like and because of the complexities in the Be case, there are more differences. Thus, for some ersatz pain to be distinguishable from its genuine counterpart along the lines that the First Objection entertains, it must cause a belief that differs in two respects from some counterpart belief caused by the genuine states. First, the ersatz-caused belief must differ from its counterpart in qualitative referents. And second, the two beliefs must have and differ in qualitative aspects beyond that of their referents, aspects in virtue of which they pick out their

referents.

But, as I have previously stated in summarizing the overall argument against the First Objection, this latter kind of difference creates its own difficulties for the First Objection, and in Chapter Eight I will set these out in more detail.

CHAPTER EIGHT DISTINGUISHING QUALIA QUALITATIVELY

We have been examining a counterargument to the absent-qualia argument against functionalism. According to this counterargument, anyone rejecting functionalism by allowing the possibility of absent qualia is committed to hopeless skepticism about the qualitative character of our mental states. The counterargument claims that this opponent of functionalism leaves phenomenal states and any qualia-free functional duplicates they might conceivably have indistinguishable.

The First Objection, you will recall, is that we can sufficiently distinguish genuine from absent-qualia states to forestall the functionalist critic. It is said to be enough that we distinguish absent-qualia states that share the nonqualitative causes and effects of genuine states and, the First Objection claims, that leaves room for qualitative differences between the two sets of states in the contents of the beliefs they cause.

The problem, I have argued, is that there would need to be qualitative aspects to whatever demonstrative beliefs are caused by the two sets of states over and beyond the qualitative aspects of their demonstrata if a subject could distinguish the absent-qualia states from the genuine states in virtue of the beliefs. But there could only be further qualitative aspects to the beliefs of an appropriate sort if the beliefs differed in these further aspects, and differed

in ways that upset the supposed functional isomorphism between the genuine states and any hypothetical absent-qualia states.

As I stated earlier in summarizing the counterargument to the First Objection, there are two kinds of difficulties. One kind is the focus of sections one and two. The second difficulty, which I earlier labeled the Location Problem, I will review in the third and fourth sections. In the fifth and final section, I will offer my own objection to the anti-skepticism argument, the Second Objection, and indicate why it escapes the problems of the First Objection.

I. Distinguishing Ersatz States Qualitatively: Initial Difficulties

Recall the first problem. If some hypothetical ersatz pain e of Smith's were distinguishable from some genuine pain g of his in virtue of different demonstrative beliefs Be and Bg about the raw feels of e and g, respectively, these beliefs would have nonqualitative effects such as his reports about what was felt. These effects would be identical. In the case of e, these nonqualitative effects would include tendencies to say things like "Pain here" which would be contradicted by the painlessness that Smith would be directly aware of, as Fig. 5 illustrates.

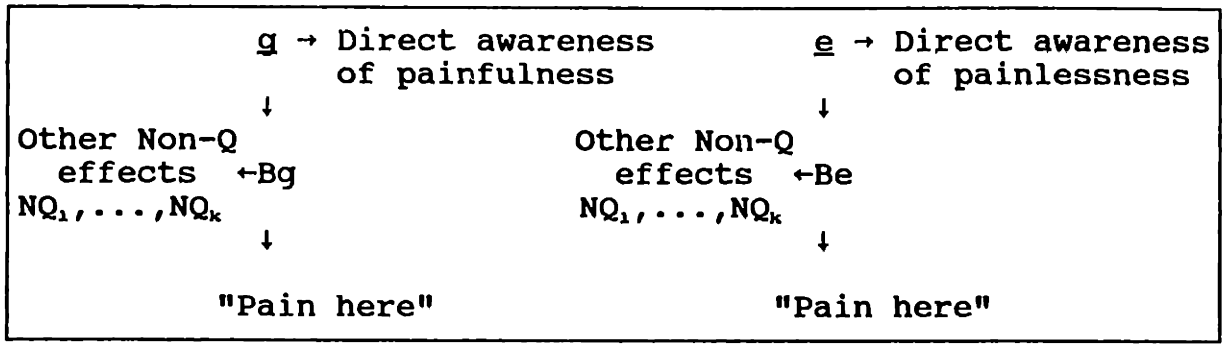

Fig. 5 Ersatz and Genuine Pains Cause Statement "Pain Here"

Normally, direct awareness of this sort associated with e would also have nonqualitative effects, including tendencies to make statements like, "No pain here."

Now, imagine that Be, the demonstrative belief about e's character, produces the nonqualitative effects it shares with g and with Bg independently of e's character and its own character. Imagine, for example, that this is because e produces the effects it shares with g merely in virtue of physical manipulation of Smith and not in virtue of anything phenomenally present or absent in e nor anything else Smith is directly aware of. In this, e would differ from g, as illustrated by Fig. 6.

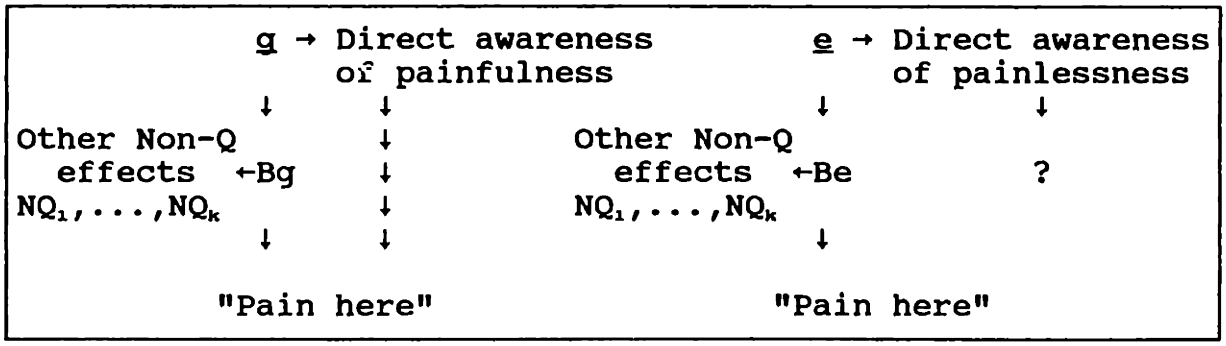

Fig. 6 How the Ersatz Differs from the Genuine State

Unlike what we are imagining the hypothetical ersatz state e to do, the genuine pain g presumably does not diverge between any psychological effects it causes in virtue of its qualitative character--what Smith is directly aware of by his having g--and the psychological effects it causes in virtue of physical manipulation of Smith. For e to accommodate such a divergence, on the other hand, Smith would have to be capable of undergoing a very abnormal psychological condition. Conee portrays that condition in the following words.

> But e is a very strange state--ersatz pain.... Yet the state is not pain because it does not feel any way at all. This lack of feeling also has its inevitable epistemic impact. Very peculiar.... Also, notice that the possibility of ersatz pain does not imply that the beliefs engendered by ersatz pain are rational. The state is by definition one that can cause false beliefs about the presence of qualitative character. Yet anyone subjected to such a state would be aware of the qualitative facts of the matter by direct experience, too. So direct awareness and the causal properies of ersatz pain would work together to bring about a bizarre combination of beliefs. Smith's epistemic condition is highly peculiar. But nothing here establishes its impossibility.[1]

Conee never explicitly states why he takes this to be a condition of irrationality. Actually, what Conee writes is

---

1. Conee, op. cit., pp. 358-359.

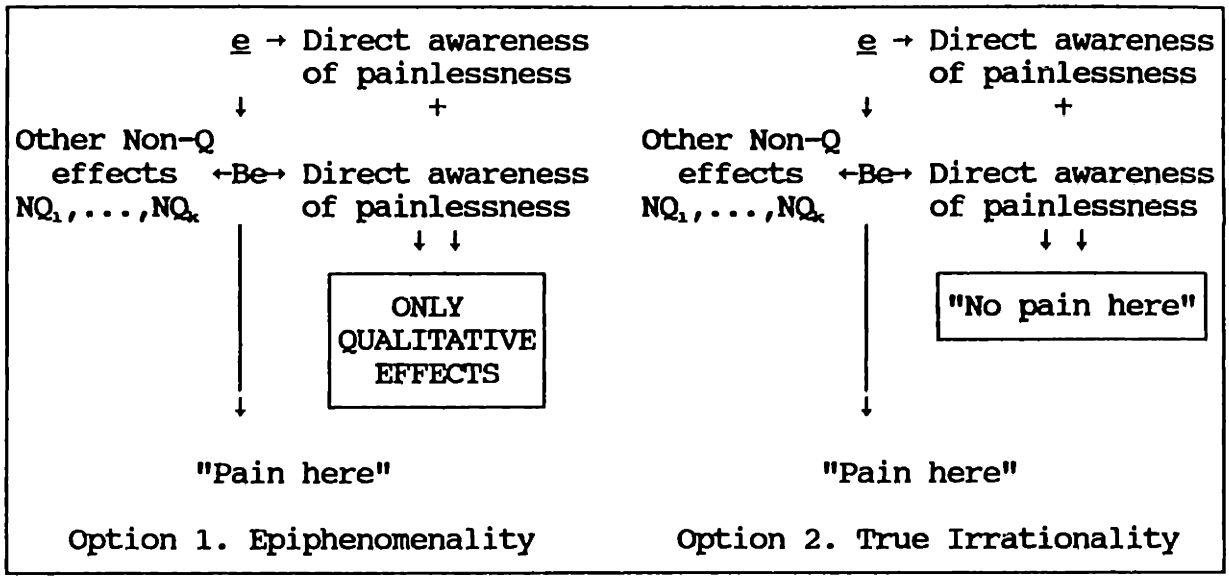

Fig. 7 The Epiphenomenality and Irrationality Options

compatible with two options for how to understand Smith's condition, which I illustrate in Fig. 7. On the one hand, there might be a condition of epiphenomenality, in virtue of which Smith's direct awareness of e's and Be's painlessness would lack nonqualitative psychological effects in him altogether. In this case, Smith might still be rational since he might have no beliefs of the form p and not-p. On the other hand, we can also imagine a condition of true irrationality, one in which Smith's direct awareness of his painlessness while in e and Be does produce nonqualitative effects in his belief and produces beliefs that directly contradict those produced by e and shared with g.

Neither option is possible, however, contrary to what Conee asserts. The kind of epiphenomenality required by the first option is unrealizable. And the irrationality

entertained in the second option is self-defeating. Smith would have to be, as the result of e, subject to the fixation of false beliefs in the face of his belief, caused by his direct awareness, that he was painless, but he could not be this way within the confines of how e is defined.

Consider Option 1 more closely. The argument against the epiphenomenality it appeals to--against the possibility that Smith's direct awareness of his painlessness while undergoing e has not nonqualitative psychological effects--is straightforward. Assume for the sake of argument it is otherwise. In that case there would exist a possible qualitative state that had no nonqualitative psychological effects whatever. Although such a state could be known qualitatively, it could not be the object of nonqualitative knowledge. Thus, no expressions of belief could be true of it; in fact, no beliefs at all could even be expressed about it. This seems absurd. For we do not know what it would be like for there to be a kind of qualitative mental state of which no expressible knowledge was in principle possible. Of course, there are qualitative states about which many of us, as a matter of fact, lack certain kinds of expressible knowledge; among such states are those we train painters, musicians and wine-tasters to become more sensitive to. But surely Smith's direct awareness of his painlessness is robust, not at all like those states. Moreover, it begs the question against functionalism to assume that there are

qualitative states about which we of necessity have no expressible knowledge.

Some kinds of epiphenomenality have had defenders, such as that associated with dual-aspect theories, according to which phenomenal properties are epiphenomenal with respect to physical properties. But the possibility of qualitative states lacking nonqualitative effects is very different. Dual-aspect epiphenomenalism is at least superficially compatible with qualitative knowledge, since it is consistent with the view that phenomenal properties, although epiphenomenal, supervene on the physical-functional properties sufficient for such knowledge. The kind of epiphenomenality illustrated in Fig. 7, however, does not even have this much favoring it. For it is hard to see how there could be qualitative knowledge of being in a state that one was, by hypothesis, directly aware of the phenomenal character of without there also being nonqualitative dispositional aspects of the knowledge beyond its qualitative aspects. This is a lesson of the private language argument, but one need not accept its common behaviorist formulations to understand the lesson.[2] And without even knowledge of what one is supposed to be

---

2. See Ludwig Wittgenstein, Philosophical Investigations (New York: Macmillan, 1968), pp. 88ff.; Norman Malcolm, "Wittgenstein's Philosophical Investigations," in V. C. Chappell, The Philosophy of Mind (Englewood Cliffs, N.J.: Prentice-Hall, 1962); Saul Kripke, Wittgenstein on Rules and Private Language (Cambridge, Mass.: Harvard University Press, 1982).

directly aware of in the state, a truly absurd skepticism of the sort Shoemaker envisions and criticizes would, in fact, confront us.

If this is not a realizable option and we are to develop a scenario for e's producing in Smith the nonqualitative effects of genuine pain independently of any direct awareness of his painlessness, then we must turn to the irrationality option. In such a case, Smith's direct awareness of his painlessness would produce its own set of nonqualitative effects alongside the nonqualitative effects already produced by e. But this result would not be any more possible than the previous option. Many of the nonqualitative effects e shares with the genuine pain g express belief in painfulness. But whatever mental processes produce spurious belief of this sort would be interrupted by the overwhelming, counterveiling evidence from direct awareness of painlessness and from the normal nonqualitative beliefs caused by such direct awareness. This would not be an acceptible irrationality but an impossible state in which the latter, nonspurious beliefs would undercut the processes fixing the former, spurious ones. And if it were possible to overcome the counterveiling evidence of the nonspurious beliefs somehow, this would become so only by use of processes outside the functional causal role that g and e share, processes that would not fit the template of causes and effects that e and

its effects must conform to.

Some of the effects of genuine pain are like reflexes. Automatically, Smith in pain winces and tries to be rid of the source. It is possible to imagine reflex-like responses such as these being produced not by pain but by the ersatz state e, yet surviving the counterveiling evidence of painlessness in direct awareness. The state e simply excites those centers of the nervous system responsible for such reflex-like behavior. But this behavior alone is not enough to single out pain from certain feelings of cold, itchiness or other discomfort. For Smith to be undergoing a true ersatz pain, one satisfying a functional definition of pain, Smith's state must produce effects much more complex than reflexes. What might make it plausible, though false, that Smith were in pain would be his propositional attitudes. These might include his tendencies to ascribe a location to his state, to associate it with a shape and boundaries, to characterize it by type and severity, to compare it to other mental states, and to continuously monitor it for changes in all these respects. But it is not plausible that Smith could be producing all these nonqualitative aspects of propositional attitudes on the basis of an ersatz state at the very same time that he was directly aware of his painlessness and producing on that basis a parallel and contradictory set of propositional attitudes.

For example, one of the normal effects of having a genuine pain is being prompted to utter, with sincerity and conviction, sentences like, "I hurt." The details of how that takes place are mysterious, but whatever they are, e by hypothesis simulates the nonqualitative aspects. It is not enough for e merely to cause Smith to parrot the words; he must understand and mean them. And he must say them for the very same reasons--or at least for functionally equivalent reasons--for which he says them when prompted by real pain. How could it be possible for Smith to do that at the very same time he was prompted by awareness of the absence of pain to say, "I don't hurt"? Even if a hypothetical ersatz pain were to set off the kind of alarm in Smith real pain does, provoking him to wince and have other pain-related reflexes, it could hardly, for example, create in him any conviction or sincerity in his avowals of pain. He would by hypothesis attribute to his ersatz pain the same location g had, but when he would search that location he would find nothing. If his direct awareness of this absence of pain were to have its normal effects, then it would have to undermine any conviction Smith might have tended to have that he really were in pain, and moreover, to rob his avowals of some of their sincerity. But then, e would no longer simulate the causal role of g. Sincerity and conviction have their own nonqualitative effects--including tendencies to avow sincerity and conviction--and, because of

the impact of direct awareness, there would be no e that had all the effects of this kind that g has.

II. Assimilating Absent Qualia Cases to Spectrum Inversion

It may seem that we could alleviate these difficulties if e would produce its nonqualitative effects not independently of Smith's direct awareness of its painlessness but rather partly in virtue of it. This option is diagrammed in Fig. 8.

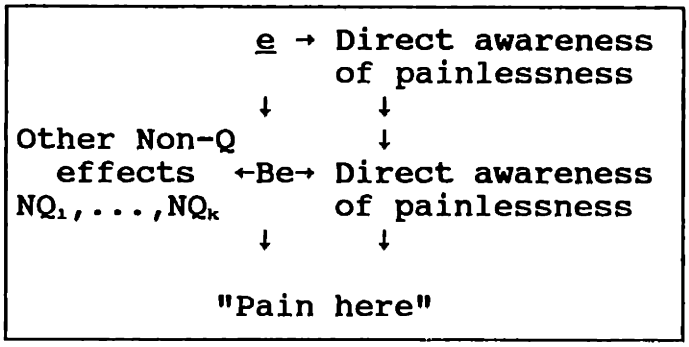

Fig. 8 Option 3: Qualia Inversion

Option 3 takes absent qualia cases to be like the standard cases of spectrum inversion but with two important differences. The first concerns what is switched. In the standard thought experiments about spectrum inversion, the presence of a simple phenomenal quality is made intrapersonally or interpersonally to assume the normal functional causal role of the presence of a second simple phenomenal quality. And the presence of the second quality

is made to assume the normal causal role of the presence of the first. Here, by contrast, it is the absence of one phenomenal quality which is made to assume the normal functional causal role of the presence of a second: painlessness is made to have the effects of painfulness. The second difference concerns how much is switched. Here, unlike what one finds in the standard thought experiments about spectrum inversion, one finds only what I will call a partial inversion. One phenomenal quality assumes the normal functional role of a second but the second keeps its normal functional role: both painlessness and painfulness assume the same functional role, the role normally had by painfulness.

As Shoemaker has shown, it takes considerable work to develop thought experiments about spectrum inversion that meet obvious counterexamples.[3] The published speculations about absent qualia do not even hint about how such counterexamples could be met if absent qualia cases were to be construed like cases of spectrum inversion, much less with these differences.

Obvious counterexamples appear very formidable. Consider one kind of difficulty with standard models of spectrum inversion. Pre-reflectively, we can imagine a color inversion of blue and red, say, because we can imagine

---

3. Sydney Shoemaker, "The Inverted Spectrum," Journal of Philosophy (1981).

blue images doing all the causal work of red images and red images doing all the causal work of blue images. On reflection, however, we might develop a nagging suspicion that some of the features of red images are not causally interchangeable with corresponding features of blue images. A mutual substitution, for example, might fail to preserve "betweenness" and "distance" relationships among color appearances: we might be prone to say certain things about *the color spectrum with the two colors out of their normal* places that we might not be prone to say of it in the normal case.

Shoemaker rightly points out that many philosophical points about spectrum inversion need rely only on the mere conceptual possibility that there are colors--though perhaps not red and blue--where <u>there are no such differences</u> between the normal and the inverted cases.[4] However, the situation is different in the case of Option 3. It really does make a difference whether we can have the mixing up of direct awareness and nonqualitative effects contemplated there or whether it is just an illusion that we can. For if we cannot mix the two things up, then this exhausts the options and no way remains to create an ersatz state <u>e</u> that is both functionally equivalent to <u>g</u> and can be

---

4. Sydney Shoemaker, "Functionalism and Qualia," <u>Philosophical Studies</u> (1975), and reprinted in his <u>Identity, Cause and Mind</u> (Cambridge: Cambridge University Press, 1984); in the latter, see p. 196.

distinguished from it through Be. And, in fact, we cannot. There are insuperable difficulties with Option 3 corresponding to each of the two ways it differs from standard models of spectrum inversion.

First, the fact that it is painfulness and painlessness that we are "inverting" with Option 3 means that there will be difficulties like the failures of "betweenness" and "distance" comtemplated above with red and blue. Painfulness and painlessness are not remotely similar enough to think that they could exchange function roles. Say, for example, that g, Smith's genuine pain, is a mild but sharp pain, located in the big toe of his left foot near the surface of the skin. That representation of his bodily condition presents Smith with a range of data about himself and produces its effects in him in virtue of the details in the data. There is in e no comparable collection of data. Smith's direct awareness of his painlessness has no comparable detail; and even if he were to focus his direct awareness on his left foot, no such detail or anything isomorphic to it would appear. There is thus no way that e, Be or the direct awareness associated with them could produce effects like those of g if these effects must emerge in virtue of the qualitative character of Smith's experience.

Second, there is a problem in the supposition that Option 3 is a case of what I called partial inversion, with

both g and e producing the same effects in virtue of different phenomemal character. For Smith's direct awareness, while he has a genuine pain like g, is of nonuniformity of feeling in his left foot; this direct awareness and any work that his imagination might put it to will produce nonqualitative effects in cognition and behavior. These effects would differ from those g and e would, by hypothesis, produce in virtue of the phenomenal feel of his left foot; since both painlessness and painfulness produce the normal effects of painfulness, it would functionally be for Smith as if there were uniform pain in his left foot. But this would make Smith's direct awareness of nonuniformity to his feelings there epiphenomenal, as I have argued an impossibility. Thus, there could not be an ersatz state like e as contemplated by Option 3.

These three options exhaust the conceivable ways that a hypothetical state like e might through Be produce the effects that they share with g and Bg. Yet none of them are real possibilities, and we must conclude that there could not exist in Smith such a state as e distinguishable from g.

III. The Location Problem

Now I want to look at a different set of problems. The difficulties I just discussed are difficulties that anybody

would encounter trying to make sense of the notion of absent qualia. The problems that I am about to review in this section, which in my summary of the argument in Chapter Seven I collectively referred to as the location problem, are special problems that arise when one tries to make sense of our having knowledge of and referring to our absent qualia. They are problems for the anti-functionalist whose argument requires that we could distinguish absent qualia states from genuine ones, since that would require that we could know of and refer to distinguishing features of our absent qualia states and I will argue we cannot.

What qualitative character is belief in Be supposed to refer to? How does it refer? I shall assume that if knowing Be and Bg is to distinguish the ersatz pain e from the genuine pain g, then n, the feature of e by which Be distinguishes it from g, must have at least the following two properties:

(1) e's character n is to be discovered in the concrete raw feel of some stretch of Smith's experience rather than in something Smith merely imagines feeling or merely conceptualizes; *and*

(2) Be refers to n in virtue of some means or other of directing Smith's attention to n.

If Bg and Be are to do the distinguishing they are required to do, then, in support of condition (1), they must do so by incorporating different concrete aspects of the experience

Smith undergoes while in the genuine and the ersatz states. What he merely imagines or conceptualizes feeling and does not really feel could not play the required role in distinguishing ersatz from genuine pain unless there were also a more direct way of doing the distinguishing, that of (so to speak) grasping and showing the distinguishing feeling directly. For referring devices that picked out only in virtue of what Smith imagines or conceptualizes could only place the distinguishing feeling in a type. Unless there were a more direct way of picking out the distinguishing feeling there would always be the question of whether the type did actually include the specific feeling.

In support of condition (2), Bg and Be can constitute bits of knowledge only if Smith is aware of concrete pieces of his experience as the specific raw feels he takes to be possessed by the states he believes to be pain. It is not enough for them merely to be lurking in experience. They must deliberately enter Smith's thinking in the special way appropriate to reference.

I shall argue that for these requirements of distinguishing to be satisfied, there would have to be qualitative aspects to Smith's mental processes that here, as before, would produce nonqualitative effects upsetting the functional isomorphism of any hypothetical ersatz pain e to the genuine pain g.

Difficulties with satisfying these two conditions

within the constraints of the functional causal role e by definition possesses are nicely illustrated by the candidate for e's character proposed by Conee. Think of e, as I have previously suggested, as a purely mechanical state of Smith's brain, which when prompted by pain's normal causes, spews out pain's normal effects with the maximum possible conformity to g's functional causal role. Belief in Be, according to Conee, is about the phenomenal character of "the seeking-and-finding-no-feeling sort of experience." This experience, he asserts, is included in the mental process of "[g]iving attention to a mental state one is in, attempting to introspect some qualitative character, and failing to find any"; assume that e somehow prompts this mental process. It is while in e that Smith, Conee writes, "could introspect and come to know" Be, making "direct reference" to the peculiar phenomenal character of the seeking-and-finding-no-feeling experience. Somehow what Smith makes direct reference to gets incorporated into Be's content.

Clearly, there are problems with Conee's exposition. The most glaring one is that there could not possibly be any experience of "seeking-and-not-finding" causally related to e or to belief in Be. The states g and e must share a functional causal role, and so too must the beliefs Be and Bg. But neither pair can share a role if e and belief in Be are to be causally linked to a "seeking-and-not-finding"

experience, since g and belief in Bg are not linked to such an experience or any mental processes, qualitative or nonqualitative, normally associated with such an experience.

It is not just that Conee has made a bad choice of words to describe the experience. Neither can there be any wider mental process of "giving attention to a mental state one is in, attempting to introspect some qualitative character, and failing to find any" for the experience to be included in. Smith's normal pain-state g is not causally related to any such mental process: when Smith is in the genuine state g, he does not "attempt to introspect" the qualitative character of g, as he might for some problematic state like e. Thus, the functionally equivalent e cannot be related to an "attempt to introspect" either. Nor, when Smith is in e, would he "give attention" to e in anything like the way he "gives attention" to g, since, by hypothesis, there is in e nothing for Smith to attend to, given e's lack of qualitative character. Thus, another failure of functional equivalence.

Finally, let us imagine, to the extent that it is possible to imagine this, that Smith in e were to go through mental processes of the very same kinds that he would normally go through while in real pain except for a single difference--that of there being while undergoing e no feeling of pain. I will call any ersatz pain that fits this description a virtual pain. The idea of what Smith would

"normally" go through in real pain is an informal one but it excludes from the domain of virtual pains any ersatz pains associated with epiphenomenality or irrationality (like those I entertained in the last section), since normal pains are not. Now, even if e were a virtual pain, would it be accurate in such a case to describe what happens to Smith in e as his (to use Conee's characterization) "failing to find" qualitative character? There is something that he would, by hypothesis, fail to find--namely, g or the qualitative character of g. But there is no obvious reason for supposing that he would inevitably fail to find "any" qualitative character, as Conee writes. One part of focussing one's attention introspectively upon a qualitative mental state like pain is focussing upon a phenomenal location. One need not focus upon it as a particular phenomenal location; it is enough to focus at that location. At the same phenomenal location where he experiences g's qualitative character, Smith in e would ordinarily find, if the only difference for him when in e were the absence of g's character, not, as Conee's characterization has it, a complete absence of qualitative character. Instead, there would be a presence of qualitative character, but of other sorts than g's. For example, there might be proprioceptive feelings connected with the sense of movement and the so-called position sense (the quality of experience by which a person with eyes closed knows where parts of the body are in

space) and perhaps also feelings of heat, of tenseness, of muscle fatigue, and so forth. We can conceive of cases of complete absence of qualitative character, as when one loses a limb and then does not even have the ghostly continuing feelings there that some amputees report but has, instead, no feeling at all. But these cases are not like the case of e.

It may seem that the problem is of Conee's own making, that the descriptions of the experience whose character belief in Be refers to and of the mental process it is included in are unnecessarily inflated beyond the functional causal role g and e are supposed to share. But that is not so. There is an obviousness and reasonableness about these descriptions. While in e, Smith has no pain. His knowledge of that would seem to involve a univerally quantified kind of knowledge. I suggest that the belief Be, which is causally associated, does as well. To see this, first consider English sentences of the form "I have a pain." Let us assume that they are elliptical and can be expanded into sentences like (A).

(A) I have a pain at location L.

Constants that occupy the referential position filled in (A) by "L" purport to refer to phenomenal locations in Smith; even if we are uncertain about the ontological status of phenomenal locations, it is still safe to say that Smith

represents to himself his pain status by beliefs of the form of (A). Assume that Smith has feelings of some sort or other at all the normal places in, say, the big toe of his right foot. There thus exist some true beliefs or other of the form of (A), where the noun phrase and the constant are replaced by expressions purporting to refer to qualia and locations in Smith's big toe on his right foot. Now, the sentence "I have no pain," by contrast, can be expanded into (A').

(A') ($\forall$ L) $\neg$ (I have a pain at L)

Knowledge of (A'), even restricted to some region of Smith's body, requires quantified knowledge about a set of locations. Such universally quantified knowledge cannot be obtained by acquaintance without knowledge by acquaintance of each of those locations.

Now, the case of Smith's hypothetical belief Be would need to be similar. Not only would Be need to be causally associated with knowledge like that Smith would express by (A') but the reference of Be to the qualitative character of not feeling pain at any phenomenal location entails Smith's having knowledge by acquaintance of at least some feature or other at each location. By contrast to this, Smith's knowledge of what he feels while in g, such as his knowledge of Bg, does not entail any universally quantified knowledge about a set of locations, nor is it causally associated in

the way Be is with any other universally quantified kind of knowledge. There is thus, on these present assumptions about how one might produce a state like e, a big difference between g and e and between knowledge of Bg and knowledge of Be. It is so big that it is difficult to see how belief in Be, on these assumptions, could refer to the sort of thing Be is supposed to and remain functionally equivalent to Bg.

Consider this objection: "You ignore the fact that g and e need only share nonqualitative causal relations. This leaves room among the qualitative causal relations they do not share for a 'big difference' in mental processes that gives e a knowable qualitative character while keeping it functionally equivalent to g. After all, searching phenomenal locations for a qualitative character is a qualitative mental process." The objector, however, ignores the fact that such allowably different qualitative causal relations as the objector insists on would normally themselves have nonqualitative effects and produce a failure of functional equivalence.

IV. Might There Be Ersatz Pains with Abnormal Causes and Effects?

Let me pursue the objector's suggestion further. Could we perhaps suspend some of the present assumptions about how one might produce a state like e? So far I have assumed that Smith in e would be pretty normal except for his having

this pain-like state, which lacks the feel of pain but is itself otherwise like a normal pain. What if we look at more abnormal cases? What if there were a case in which the qualitative mental processes associated with reference to e's qualitative character n, which I have been describing as search-like and thus outside e's causal role, did not have the causal relations to nonqualitative effects normally associated with reference to qualitative character? Is this possible? If it were, then the search-like mental processes Smith needs to refer to n might do their work outside e's causal role without upsetting e's isomorphism to g.

I shall now argue that this is not possible. First, let us review the paradigm of Smith's searching out and finding e which I have been assuming so far. In Fig. 9, I set out diagrams of how we might understand g as a normal case of pain and of how we might understand e as a case of virtual pain.

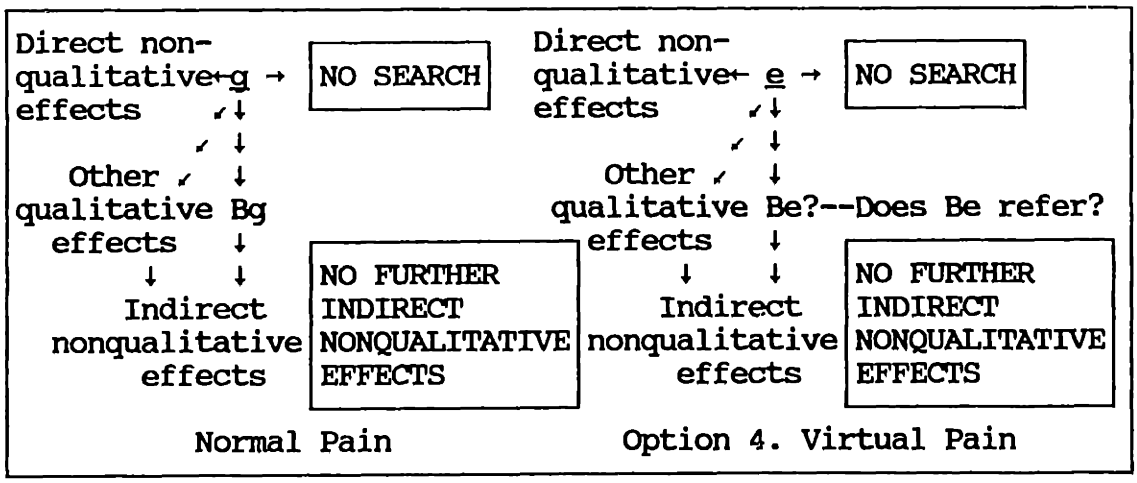

Fig. 9 Normal Pain and Virtual Pain

If Smith is asked, "Do you have pain in your right leg?", he will most likely try to answer the question by directing his attention to his leg. If he is in g, no search will normally be needed to answer the question. Assuming g is intense and distinct enough to make itself known easily, attending to his leg will be enough for Smith to realize he has pain in the big toe of his right foot. If he is in e, by contrast, his attention will not immediately be drawn to anything in introspection, even though he takes himself to be in pain. How then can Smith, in virtue of believing Be, point out to himself some raw feel in himself as the phenomenal character of the state he takes to be pain? Raw feels are for the most part at phenomenal locations; not only is there no obvious candidate in introspection for e's specific raw feel at any specific location but Smith does not even have access to any property of e that would distinguish e introspectively from most other mental states. Smith would have to undergo some further mental process of collecting evidence about e if he were ostend to himself some raw feel as e's specific raw feel; however, this would be incompatible with construing e as a virtual pain, one differing from some normal pain only in its lack of qualitative character.

Obviously, Smith cannot conduct a reportable search through phenomenal locations in his leg, either; otherwise, e would not be functionally equivalent to g. This leaves

several apparent alternatives. Two of them I set out in Fig. 10.

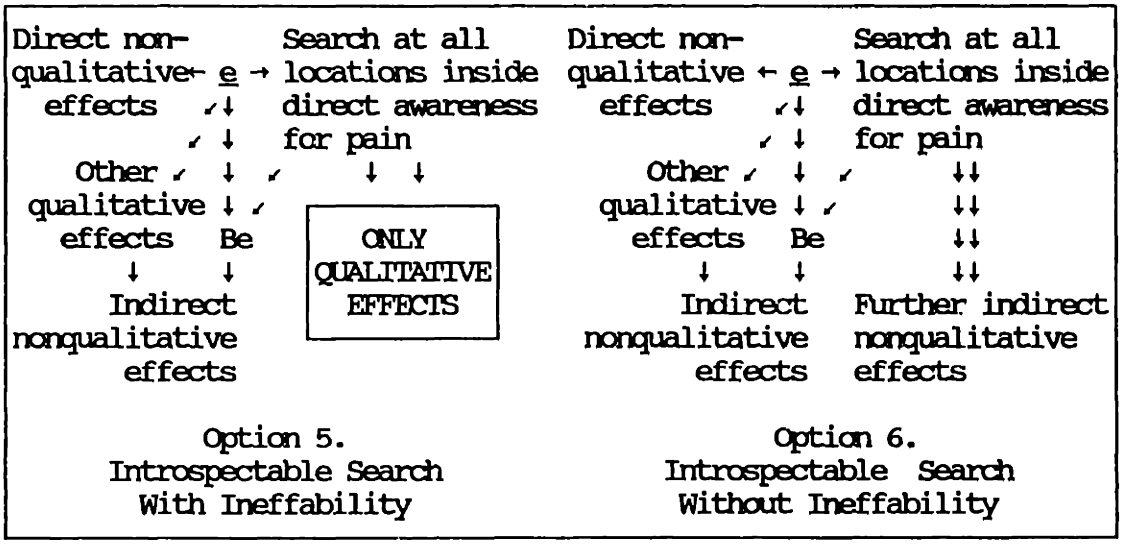

Fig. 10 Options 5 & 6: Introspectable Searches and Ineffability

Option 5 exploits the idea that e's functional equivalence to g could be maintained if the search that fixes Be's reference to e's character were ineffable, without nonqualitative effects. But this is not a real alternative. Earlier I argued against the related idea that Smith's direct awareness of the painlessness of his ersatz pains would not interfere with their functional causal role in relevant ways because his awareness could be epiphenomenal. The argument here against Option 5 is similar and just as straightforward. The introspective access which Smith would normally have to any search throughout phenomenal locations of his body for evidence for fixing his reference to e's character would be robust, the kind of thing Smith could think about out loud or to himself. For example, the

product of the search which Conee entertains Smith carrying out is, by hypothesis, a failure to find, which would appear to be reportable. While we do sometimes seem to check our direct awarenesses for their contents automatically, in ways that turn up results without our quite being able to say how, it is impossible to make sense of such a thing in this case. How can we understand Smith's running through his direct awareness in a search of the sort appropriate to his taking some specific raw feel to be e's raw feel without there being some effect among his intentional states? It would certainly beg the question against functionalism just to assume we can understand that.

Also untenable is the idea which I have labeled Option 6, that e's functional isomorphism could be maintained even if the hypothetical search for e's character produced non-isomorphic effects. This will not work because e, by hypothesis, produces the search; any non-isomorphic, nonqualitative effects would be indirect effects of e and would contradict the requirement that all of e's indirect effects be part of the functional role it shares with g. The functional definition determining e would thus not be the best one possible.

What if the reference-fixing search for e's character were to take place outside Smith's awareness altogether? In that case, there would be no ineffable qualitative effects from e, nor would there be extra psychological effects to

upset e's isomorphism to g. This is Option 7, diagrammed below in Fig 11.

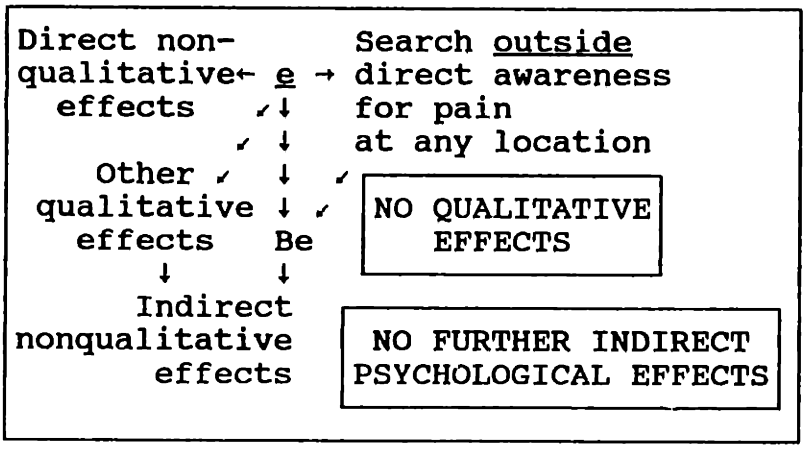

Fig. 11 Option 7: Search Outside Awareness

Suppose that this works as follows. Inside Smith's head there is a machine--say a souped-up cerebroscope. Unbeknownst to him, the machine looks for pain and feeds its negative results into the process that makes him believe that the state he thinks is pain feels like n.

But this will not work either. How could Smith know that something feels like this character--n--unless he had conscious access to the hypothetical search that n is part of? In the story under examination, Smith distinguishes e from g in virtue of an aspect of the overall experience he has while undergoing e, something he can point out to himself. There would be no way to point out something like the phenomenal character of a "seeking-and-not-finding" sort of experience unless Smith had before his mind such an experience, the experience of a search. It makes no sense

to suppose that he could run through a search, so to speak, cerebroscopically, outside awareness, and in virtue of that just know its phenomenal character. For to know that, by way of a singular proposition, is to point out something that depends for its existence on acquaintance with how it is an aspect of the wider course of experience, and a merely cerebroscopic search would not provide Smith with this wider acquaintance. It would be as if one could just know what the surface color appearance of an apple was like without any acquaintance with the apple's surface.

By way of concluding my counterargument to the First Objection, let me examine one kind of response to it. Throughout, it may seem that the problems I have cited turn on the particular examples I have used, and that these problems could be eliminated by producing different examples. So imagine that instead of Bg and Be we try to distinguish g and e by way of Ag and Ae.

(Ag) At the phenomenal location where I believe I am in pain I find this character in introspection (making direct reference to its character, partly that of feeling painful).

(Ae) At the phenomenal location where I believe I am in pain I find this character in introspection (making direct reference to its character, partly that of feeling painless).

Here, no search of the sort entertained for Be is required.

Smith does not have to look for phenomenal character to have some seeking-and-not-finding experience. It is enough simply to inspect one phenomenal location, the one where he thinks he is having the pain. Since the extra nonqualitative effects of a search no longer appear, it may seem that we have escaped the sorts of difficulties associated with them.

My reply is that even if the appeal to a search is extreme, it emphasizes a general problem associated with reference to any qualitative character that would be sufficient to distinguish an ersatz from a genuine state. In general, any reference to raw feels adequate for distinguishing an ersatz state by way of some effect in Smith's beliefs like Be or Ae would be fixed by a process of referring that would have different nonqualitative effects from the process of referring in the case of a counterpart genuine-state belief.

Say, to take another extreme case, that we tried to distinguish a pain in the big toe of Smith's left foot from an ersatz counterpart by two beliefs that each referred to the entire phenomenal feel of his left leg. It will not do to ostend the respective feels by Smith's attending to the feel of just a part of his left leg--as one might do in ostending an elephant by attending to part of <u>it</u>, say its left leg. For one could not distinguish the elephant from a weird hybrid that had elephant-like legs beneath the body of

a zebra just by ostending a leg. Similarly, Smith could not distinguish the feel of his entire left leg from the feel of a leg phenomenally the same but except for a pain by attending to the part of the phenomenal feel separate from the pain. Only by attending to <u>that aspect</u> of the leg in virtue of which there is a difference could Smith hope to <u>distinguish</u> the feel of his painful leg from the feel of a painless one. But the <u>means</u> of picking out the one raw feel will differ from the means of picking out the other when the feels to be distinguished are regions of feeling that differ only by the presence and absence of some qualitative property. The difference is reflected in the common-sense metaphors we associate with the two ways of picking out. In the case where the qualitative property is present, it is as if the attention is attracted by a magnet, fastened to some region of phenomenal space along the contours of the image or feeling that occupies it. In the case where the qualitative property is absent, it is as if attention requires effort, needing a <u>focussing</u> of attention in order to attend to one phenomenal region that does not substantially differ from its neighboring regions.

V. The Second Objection

I will now turn from the unsuccessful First Objection to what I labeled in the summary in section two as the

Second Objection. This too is an attack upon the functionalist's case against the possibility of absent qualia, but unlike the First Objection, it is successful. Although related, it is also different from the First Objection in important ways. It suffers none of the problems that defeat the First Objection, in fact exploiting the failure of the latter to make its case.

In this final section I review why it is that the First Objection fails. Shoemaker is also correct that it fails to make room for the possibility of ersatz states in sentient creatures like Smith but for the wrong reasons, and I compare his reasons with the right ones. Then I spell out the Second Objection. I argue that the First Objection fails because the side-effects that create difficulties in constructing ersatz states only arise in a creature with qualitative states. If we consider instead creatures that are qualia-free, as we do in the Second Objection, there are no opportunities for such side-effects to arise.

My case against the First Objection establishes that the functionalist is correct in claiming that ersatz pain is not possible in a sentient and sapient creature like Smith. Even given a weak sort of distinguishing, one that does not even require the capacity to compare, no ersatz pain in Smith could be distinguished from genuine pains. But it is plausible to think, as the functionalist insists, that Smith would need to be able at least <u>weakly</u> to distinguish any

ersatz pain he would have for him to be able to have them.

Still, even though those functionalists who argue like Shoemaker does are right about this, they are right for the wrong reasons. Shoemaker argues that any feature purported to be a reliable indicator distinguishing between genuine and ersatz qualitative states would never be able to do so. He contends that if some proposed functional definition detailing the functional causal role the states were supposed to share were to omit the feature, that would merely mean that the proposed functional definition was not the best one possible, not that the feature was a distinguishing feature. This leaves any hypothetical ersatz-state subject in the same epistemological relation to his or her ersatz state as the genuine-state subject is in to his or her genuine state. The two states would have the very same functional causal roles and there would thus be no reliable indicators to distinguish them. Thus he or she is in no position to know of being in an ersatz state or a genuine state. Since we always do know that, the argument goes, ersatz states in us must be impossible.[5]

Shoemaker's premise that a hypothetical difference between ersatz and genuine states can always just be "added" to any functional causal role they purportedly share, however, is misguided. Just as we can conceive of two

---

5. Shoemaker, "Absent Qualia Are Impossible," op. cit., pp. 589-590.

functionally isomorphic states that have different qualitative features--with one red and the other some spectral inverse like green, for example--we can conceive, prima facie, of two functionally isomorphic states, one having and the other lacking qualitative character. There is nothing in the concept of functional isomorphism that rules out that conceivable possibility. Moreover, we can conceive of the qualitative member of the pair as causing one set of appropriate effects, the qualia-free member causing appropriate but different effects, at least without initial contradiction. Shoemaker's premise that hypothetical differences can be "added" to any purportedly shared functional causal role to eliminate any appearance of differences is true only where the differences are indisputably nonqualitative. If, by contrast, they would be qualitative, it would be question-begging just to assume that they could be just "added". We have Conee's reasons for thinking otherwise.

It is not Shoemaker's question-begging reasoning about these matters but rather the set of difficulties emerging from the location-problem argument and the other arguments of this chapter which confirms Shoemaker's negative conclusion about absent qualia in a sentient and sapient creature like Smith. Ersatz states are not possible in Smith because, for one thing, he cannot locate them given the constraints of the causal roles they would have, and

thus he cannot refer to them or know of them. For Smith to undergo ersatz states, they must have the same relations to states of knowledge that their genuine counterparts do. Since they cannot, they cannot exist in Smith or any other creature capable of knowing its own qualitative states. It is for this reason, as we have seen, that there are no beliefs like the qualitative belief in Be and no nonpropositional mental states like e that are possible for Smith and creatures like him to have.

Nevertheless, although Shoemaker is correct, and Conee is wrong, about this relation among ersatzness, sentience and sapience, Shoemaker and Conee both seem to be unjustifiably confident about a further claim: that if ersatz states are not possible in creatures like Smith, they are not possible at all. But neither writer gives a direct argument for this principle.

This further claim is in fact refutable. It is because of that that the Second Objection is possible. The Second Objection makes its case for the possibility of ersatz states not on the basis of sentient creatures like Smith but on the basis of qualia-free creatures.

The Second Objection invites us, just as the First Objection does, to try to imagine a state that, like e, satisfies the best possible functional description of genuine pain while lacking qualitative character. The Second Objection, however, invites us to imagine such a

thing being realized not in a sentient and sapient creature like Smith, as the First Objection did, but in a nonsentient entity. What is required is some entity close enough to Smith in functional organization that, even though it might not be functionally isomorphic through and through, it has enough isomorphism (among pain's normal effects, for example) to realize at least one state functionally equivalent to pain. But let us suppose that it lacks qualitative character throughout--in all its states. Even though such an entity would lack all qualitative character, this presents no obstacle to its realizing what I defined as nonqualitative functional characterizations of genuine qualitative states like pain, descriptions formulated purely in nonqualitative terms. Nothing in my counterargument to the First Objection conflicts with the existence of <u>this</u> kind of ersatz state. For all the problems I cited arise from side-effects that could not appear in a creature without qualitative states.

Now, the existence of such an entity as this constitutes a counterexample to EP, the premise attacked unsuccessfully by the First Objection. For if <u>e</u> is possible it <u>is</u> possible even though, as K asserts and contrary to EP's consequent, the presence or absence of qualitative character makes a difference that distinguishes genuine pain from ersatz pain. It makes a very obvious difference even though, as I argued in Chapter Six, there is no general

epistemological principle supporting K. If the counterargument to the First Objection is correct, we know whenever we are having genuine pains that we are having them rather than ersatz pains. We know this in virtue of the presence of the qualitative character of the genuine pain, together with an a priori argument that ersatz pains are not possible in somebody who has the genuine states they would be ersatz counterparts to. In creatures like Smith and the rest of us who feel pain, ersatz pains do not conform to the functional template of causes and effects of genuine states. They create too many difficulties with nonqualitative side-effects. In entities that do not feel pain or have any other states with qualitative character, there are no such problematical side-effects. In them, ersatz pains are possible, even though in us it is always possible to distinguish being in a genuine state from being in an ersatz state. Thus, Shoemaker's EP is false, and the anti-skeptical argument against the possibility of absent qualia which depends upon EP is defeated.

Bibliography

Akins, Kathleen. "What Is It Like to Be Myopic and Boring?" In Bo Dalhbom, ed., Dennett and His Critics. Cambridge, Mass.: Blackwell, 1993.

Aristotle, De Anima.

Armstrong, D. M. "Is Introspective Knowledge Incorrigible?" The Philosophical Review 72 (1963).

Armstrong, D. M. A Materialist Theory of the Mind (London: Routledge & Kegan Paul, 1968).

Austin, David F. What's the Meaning of "This"? Ithaca: Cornell University Press, 1990.

Bach, Kent. "De re Belief and Methodological Solipsism." In Andrew Woodfield, ed., Thought and Object. Oxford: Clarendon Press, 1982.

Block, Ned. "Are Absent Qualia Impossible?" The Philosophical Review 89 (1980).

Block, Ned. "Troubles with Functionalism." In Ned Block, ed., Readings in the Philosophy of Psychology, vol. 1. Cambridge, Mass.: Harvard University Press, 1980.

Block, Ned. "What Is Functionalism?" In Ned Block, ed., Readings in the Philosophy of Psychology, vol. 1. Cambridge, Mass.: Harvard University Press, 1980.

Block, Ned. "What Narrow Content Is Not." In B. Loewer and G. Rey, eds., Meaning in Mind: Fodor and His Critics. Cambridge, Mass.: Blackwell, 1991.

Churchland, Paul. Matter and Consciousness. Cambridge, Mass.: M.I.T. Press, 1984.

Churchland, Paul. A Neurocomputational Perspective. Cambridge, Mass.: M.I.T. Press, 1989.

Churchland, Paul. "Reduction, Qualia, and the Direct Introspection of Brain States." Journal of Philosophy 82 (1985).

Conee, Earl. "The Possibility of Absent Qualia." The Philosophical Review 94 (1985).

Cottingham, John, Robert Stoothoff and Dugald Murdoch, eds. The Philosophical Writings of Descartes. Cambridge: Cambridge University Press, 1984 and 1985. 2 vols.

Davidoff, Jules. Cognition through Color. Cambridge, Mass.: M.I.T. Press, 1991.

Dennett, Daniel. Consciousness Explained. Boston: Little, Brown, 1991.

Dennett, Daniel. "Quining Qualia." In A. Marcel and E. Bisiach, eds., Consciousness in Contemporary Science. New York: Oxford University Press, 1988.

Dear, Steven P., James A. Simmons and Jonathan Fritz. "A Possible Neuronal Basis for Representation of Acoustic Scenes in Auditory Cortex of the Big Brown Bat." Nature 364 (1993).

Dicker, Georges. Descartes: an Analytical and Historical Introduction. New York: Oxford University Press, 1993.

Evans, Gareth. The Varieties of Reference. Oxford: Oxford University Press, 1982.

Feldman, Richard. "Fallibilism and Knowing That One Knows." The Philosophical Review 90 (1981).

Flanagan, Owen. Consciousness Reconsidered Cambridge, Mass.: M.I.T. Press, 1992.

Fodor, Jerry. "The Big Idea." Times Literary Supplement, July 3, 1992.

Fodor, Jerry. Psychosemantics. Cambridge: M.I.T. Press, 1987.

Foster, John. The Immaterial Self. London: Routledge, 1991.

Frankfurt, Harry G. Demons, Dreamers, and Madmen. Indianapolis, 1970.

Geach, Peter. God and the Soul. New York: Schocken, 1969.

Goldman, Alvin. "Discrimination and Perceptual Knowledge." Journal of Philosophy 73 (1976).

Goldman, Alvin. Epistemology and Cognition. Cambridge, Mass.: Harvard, 1986.

Goldman, Alvin. "What Is Justified Belief?" In George Pappas, ed., Justification and Knowledge. Dordrecht: Reidel, 1979.

Handford, Martin. Where's Waldo? New York: Little, Brown, 1987.

Hare, R. M. "Pain and Evil." Proceedings of the Aristotelian Society, Supplementary Volume 38 (1964).

Hardin, C. L. Color for Philosophers. Indianapolis: Hackett, 1988.

Hardin, C. L. "Reply to Levine." Philosophical Psychology 4 (1991).

Hill, Christopher. Review of Metaphysics of Consciousness, by William Seager. Canadian Journal of Philosophy, forthcoming.

Hill, Christopher. Sensations. New York: Cambridge University Press, 1991.

Hooker, Michael. "Descartes's Denial of Mind-Body Identity." In Michael Hooker, ed., Descartes: Critical and Interpretive Essays. Baltimore: Johns Hopkins University Press, 1978.

Hooker, Michael. "A Mistake Concerning Conception." In Stephen F. Barker and Tom L. Beauchamp, eds., Thomas Reid: Critical Interpretations. Philadelphia: Philosophical Monographs, 1976.

Horgan, Terence. "Jackson on Physical Information and Qualia." Philosophical Quarterly 34 (1984).

Hume, David. A Treatise of Human Nature.

Hume, David. Enquiry Concerning Human Understanding.

Jackson, Frank. "Epiphenomenal Qualia." Philosophical Quarterly 32 (1982).

Jackson, Frank. "What Mary Didn't Know." Journal of Philosophy 83 (1986).

Kaplan, David. "Afterthoughts." In J. Almog, J. Perry and H. Wettstein, eds., Themes from Kaplan. New York: Oxford University Press, 1989.

Kaplan, David. "Demonstratives." In J. Almog, J. Perry and H. Wettstein, eds., Themes from Kaplan. New York: Oxford University Press, 1989.

Kim, Jaegwon. "Phenomenal Properties, Psychophysical Laws, and the Identity Theory." Monist (1972).

Kim, Jaegwon, and Richard Brandt. "The Logic of the Identity Theory." Journal of Philosophy 64 (1967).

Kinsbourne, M., and E. K. Warrington. "Observations on Colour Agnosia." Journal of Neurology, Neurosurgery and Psychiatry 27 (1964).

Kripke, Saul. "Identity and Necessity." In Milton Munitz, ed., Identity and Individuation. New York: New York University Press, 1971.

Kripke, Saul. Naming and Necessity. Cambridge, Mass.: Harvard University Press, 1980.

Kripke, Saul. Wittgenstein on Rules and Private Language Cambridge, Mass.: Harvard University Press, 1982.

Lehrer, Keith. Knowledge. New York: Oxford University Press, 1974.

Levine, Joseph. "Materialism and Qualia: the Explanatory Gap." Pacific Philosophical Quarterly 64 (1983).

Levine, Joseph. "On Leaving Out What It's Like." In Martin Davies and Glyn Humphreys, eds., Consciousness: Psychological and Philosophical Essays. Oxford: Blackwell, 1993.

Lewis, David. "Mad Pain and Martian Pain." In David Lewis, Philosophical Papers, volume 1. New York: Oxford University Press, 1983.

Lewis, David. "What Experience Teaches." In William Lycan, ed., Mind and Cognition. Oxford: Basil Blackwell, 1990.

Loar, Brian. Mind and Meaning. Cambridge: Cambridge University Press, 1981.

Loar, Brian. "Phenomenal States." Philosophical Perspectives 4 (1990).

Locke, John, Essay.

Lucian. The Fly. In A. M. Harmon, tr., Lucian, volume 1. Cambridge, Mass.: Harvard University Press, 1913.

Lycan, William. Consciousness. Cambridge, Mass.: M.I.T. Press, 1987.

Lycan, William. "Kripke and the Materialists." Journal of Philosophy 71 (1974).

Madell, Geoffrey. Mind and Materialism. Edinburgh: Edinburgh University Press, 1988.

Madell, Geoffrey. "Neurophilosophy: A Principled Sceptic's Response." Inquiry 29 (1986).

Malcolm, Norman. "Wittgenstein's Philosophical Investigations." In V. C. Chappell, ed., The Philosophy of Mind. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Markie, Peter J. Descartes's Gambit. Ithaca: Cornell University Press, 1986.

McGinn, Colin. "Anomalous Monism and Kripke's Cartesian Intuitions." Analysis 37, no. 2 (1977).

McGinn, Colin. "Can We Solve the Mind-Body Problem?" Mind 98 (1989).

Nagel, Thomas. Mortal Questions. Cambridge: Cambridge University Press, 1979.

Nagel, Thomas. "Physicalism." In David M. Rosenthal, Materialism and the Mind-Body Problem. Indianapolis: Hackett Publishing Company, 1987.

Nagel, Thomas. The View from Nowhere. New York: Oxford University Press, 1986.

Nagel, Thomas. What Does It All Mean? New York: Oxford University Press, 1987.

Nagel, Thomas. "What Is It Like to Be a Bat?" Philosophical Review 83 (1974).

Nemirow, Lawrence. Review of Mortal Questions, by Thomas Nagel. Philosophical Review 89 (1980).

Nozick, Robert. Philosophical Explanations. Cambridge, Mass.: Harvard University Press, 1981.

Parfit, Derek. "The Puzzle of Reality." Times Literary Supplement, July 3, 1992.

Parfit, Derek. Reasons and Persons. Oxford: Oxford University Press, 1984.

Peacocke, Christopher. "No Resting Place: a Critical Notice of The View from Nowhere, by Thomas Nagel." The Philosophical Review 98 (1989).

Plato, Republic.

Pliny the Elder. Natural History. Cambridge, Mass.: Harvard University Press, 1942.

Putnam, Hilary. "The Meaning of 'Meaning.'" In K. Gunderson, ed., Language, Mind, and Knowledge. Minneapolis: Univ. of Minnesota Press, 1975.

Quine, W. V. O. Ontological Relativity and Other Essays. New York: Columbia University Press, 1968.

Quine, W. V. O. Word and Object. Cambridge, Mass.: M.I.T. Press, 1960.

Rawls, John. A Theory of Justice. Cambridge, Mass.: Harvard University Press, 1971.

Rey, Georges. "A Reason for Doubting the Existence of Consciousness." In Richard J. Davidson, Gary E. Schwartz and David Shapiro, eds., Consciousness and Self-Regulation, volume 3. New York: Plenum, 1983.

Rey, Georges. "Sensational Sentences." Unpublished typescript, July 1989.

Robinson, Howard. "The Anti-Materialist Strategy and the 'Knowledge Argument.'" In Howard Robinson, ed., Objections to Physicalism. Oxford: Oxford University Press, 1993.

Robinson, Howard. "Introduction." In Howard Robinson, ed., Objections to Physicalism. Oxford: Oxford University Press, 1993.

Robinson, Howard. Matter and Sense. Cambridge: Cambridge University Press, 1982.

Rorty, Richard. "Incorrigibility as the Mark of the Mental." Journal of Philosophy 67 (1970).

Ryle, Gilbert. The Concept of Mind. New York: Barnes and Noble, 1949.

Schiffer, Stephen. "The Basis of Reference." Erkenntnis 13 (1978).

Seager, William. Metaphysics of Consciousness. London: Routledge, 1991.

Shoemaker, Sydney. "Absent Qualia Are Impossible." The Philosophical Review 90 (1981).

Shoemaker, Sydney. "Functionalism and Qualia." Philosophical Studies 27 (1975).

Shoemaker, Sydney. Identity, Cause and Mind. Cambridge: Cambridge University Press, 1984.

Shoemaker, Sydney. "The Inverted Spectrum." Journal of Philosophy (1981).

Shoemaker, Sydney. "Qualia and Consciousness." Mind 100 (1991).

Sittig, O. "Störungen im Verhalten gegenüber Farben bei Aphasischen." Monatsschift für Psychiatrie und Neurologie 49 (1921).

Smart, J. J. C. Philosophy and Scientific Realism London: Routledge and Kegan Paul, 1963.

Smart, J. J. C. "Sensations and Brain Processes." In V. C. Chappell, ed., The Philosophy of Mind. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Strawson, P. F. Individuals. New York: Anchor, 1963.

Tye, Michael. "The Subjective Qualities of Experience." Mind 98 (1986).

Tylor, E. B. Primitive Culture, 4th ed. London: John Murray, 1903.

Van Gulick, Robert. "Understanding the Phenomenal Mind: Are We All Just Armadillos?" In Martin Davies and Glyn Humphreys, eds., Consciousness: Psychological and Philosophical Essays. Oxford: Blackwell, 1993.

White, Stephen. "Curse of the Qualia." Synthese 68 (1986).

White, Stephen. "Transcendentalism and Its Discontents." In Stephen White, The Unity of the Self. Cambridge, Mass.: M.I.T. Press, 1991.

Williams, Bernard. Descartes. New York: Penguin Books, 1978.

Wilson, Margaret. "Cartesian Dualism." In Michael Hooker, ed., Descartes: Critical and Interpretive Essays. Baltimore: Johns Hopkins University Press, 1978.

Wittgenstein, Ludwig. Philosophical Investigations. New York: Macmillan, 1968.

Wittgenstein, Ludwig. Philosophical Remarks. New York: Barnes & Noble, 1975.