

Imaginability, Conceivability, Possibility and the Mind-Body Problem

Author(s): Christopher S. Hill

Source: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 87, No. 1 (Jul., 1997), pp. 61-85

Published by: Springer

Stable URL: <https://www.jstor.org/stable/4320765>

Accessed: 31-01-2020 03:26 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



Springer is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*

IMAGINABILITY, CONCEIVABILITY, POSSIBILITY AND
THE MIND-BODY PROBLEM*

(Received 26 July 1995)

In the early seventies Kripke unveiled a line of thought that was designed to resuscitate dualistic philosophies of mind.¹ This line of thought was greeted with a chorus of objections, some of which have helped considerably to bring the relevant issues into sharper focus. Still, despite the number and philosophical value of these replies, the line of thought has continued to attract sympathetic attention. Indeed, it figures prominently in three of the most original and provocative contributions to the contemporary literature on the mind-body problem – W.D. Hart's *Engines of the Soul*, Stephen Yablo's "The Real Distinction Between Mind and Body" (as amplified in its companion piece "Is Conceivability a Guide to Possibility?"), and David Chalmers's *The Conscious Mind*.²

In addition to being provocative, Kripke's line of thought was characterized by broad relevance. In one version it challenged token-materialism, and in another, type-materialism. Both versions continue to attract admirers. (Token-materialism asserts that each concrete psychological particular, such as the pain I experienced at 10:00 last night, is identical with a concrete physical particular, such as the firing of my C-fibers at 10:00 last night. Type-materialism comes in more than one variety. The strongest version claims that every type of psychological state is identical with some type of physical state. A much weaker version, which is substantially more plausible, claims that every type of *sensory* state is identical with some type of physical state.)

In this paper I will try to show that those who continue to be favorably impressed by Kripke's line of thought have underestimated the resources that are available to materialists. Specifically, I will maintain that the most widely accepted part of that line of thought, the argument against type materialism, admits of an answer that is fully satisfactory.

Two last preliminary comments. First, I will be concerned only to defend type-materialism about qualitative states, that is, type-materialism about states that are like *being a pain* in that they are forms of sentience. Second, although it is of course my intention to carry the discussion forward, I will not be aiming at radical originality. Thus, I will be relying heavily on ideas that are found in earlier papers by Thomas Nagel and myself. My aims are to clarify these ideas, to develop and supplement them in certain ways, and to show that it is possible to embrace them without incurring any questionable philosophical commitments.

I

I will begin with a brief review of the portion of Kripke's line of thought that counts directly against type-materialism. (I will follow Kripke in focussing on pain and the physical state which in fact accompanies pain in the human brain. (As is usual, I will refer to this brain state as "C-fiber stimulation.") But the argument should be seen as implicitly general: according to Kripke, what the argument claims about pain and its accompanying brain state could also be claimed, with equal legitimacy, about any other pair consisting of a type of sensation and its corresponding brain state.)

Kripke's Argument

First premise: It appears to be possible for there to be instances of the property *being a pain* that are not instances of the property *being a case of C-fiber stimulation*. For we can easily imagine, and easily conceive of, a disembodied person who is experiencing pain.

Second premise: It appears to be possible for there to be instances of *being a case of C-fiber stimulation* that are not instances of *being a pain*. For we can easily imagine, and easily conceive of, a zombie whose C-fibers are undergoing a high degree of electro-chemical activity.

Third premise: Where X and Y are any two properties, if it seems to be the case that X and Y are *separable*, in the sense that it seems to be possible for there to be instances of X that are not instances of Y, then, unless this appearance of separability can be explained away, it really is the case that X and Y are separable.

Fourth premise: Where X and Y are any two properties, if X and Y are separable, then it is not the case that X and Y are identical.

Fifth premise: In general, if X is a commonsense natural kind and Y is a scientific kind that can plausibly be identified with X, then there is a tendency for it to appear to us that X is separable from Y. However, it is possible in most cases to explain away the apparent separability of X and Y by attributing it to a tendency to confuse X with a different property – a property that normally guides us in recognizing instances of X, but that is only contingently connected with X. (To elaborate: In most cases, if X is a commonsense natural kind, then there exists a property Z such that (a) instances of Z are normally instances of X, (b) this connection between Z and X is contingent, and (c) we normally recognize instances of X by identifying them as instances of Z (i.e., Z is the property that guides us in “picking out” instances of X). When X is associated with a property Z that meets this condition, it is possible to explain away X’s apparent separability from an associated scientific kind Y by saying that the appearance of separability is due in part to a tendency to misconstrue possible situations that contain instances of Z as situations that contain instances of X, and in part also to a tendency to misconstrue possible situations that lack instances of Z as situations that lack instances of X. To be more specific, when it appears to a subject that it is possible for there to be an instance of X that is not an instance of the associated scientific kind Y, it is possible to explain this appearance away by saying that the subject has (i) imagined or conceived of a possible situation that contains an instance of Z that is not an instance of Y and (ii) misconstrued this situation as one in which there is an instance of X that is not an instance of Y. Equally, when it appears to a subject that it is possible for there to be an instance of Y that is not an instance of X, it is possible to explain this appearance away by saying that the subject has (i) imagined or conceived of a possible situation that contains an instance of Y that is not an instance of Z and (ii) misconstrued this situation as one in which there is an instance of Y that is not an instance of X.)

Sixth premise: The apparent separability of *being a pain* and *being a case of C-fiber stimulation* cannot be explained away in the way indicated in the fifth premise. (Thus, there is no property Z such that (a) instances of Z are normally instances of *being a pain*, (b) this

connection between Z and *being a pain* is contingent, and (c) we normally recognize instances of *being a pain* by identifying them as instances of Z. *Being a pain* is *itself* the property that guides us in recognizing instances of *being a pain*. (Kripke puts the point in the following way: "Pain, on the other hand, is not picked out by one of its accidental properties; rather it is picked out by the property of being pain itself, by its immediate phenomenological quality."³))

Seventh premise: The paradigm described in the fifth premise is the *only* model for explaining appearances of separability away.

Lemma: By the fifth, sixth, and seventh premises, it is not possible to explain away the apparent separability of *being a pain* and *being a case of C-fiber stimulation*.

Conclusion: By the first premise, the second premise, the third premise, the fourth premise, and the lemma, *being a pain* is not identical to *being a case of C-fiber stimulation*.

In addition to sketching this argument, Kripke takes pains to illustrate the paradigm of explanation that is described in the fifth premise. Kripke begins this part of his discussion by pointing out that, despite the fact that the commonsense natural kind heat is known to be identical with the scientific kind molecular kinetic energy, it seems possible for heat to exist without being accompanied by molecular motion. Kripke then maintains that we are normally guided in recognizing heat by the property *being the external phenomenon that causes the sensation S*, where "S" names the sensation that heat normally produces in human observers in the actual world. Next, he says that when it seems to us that heat can exist without being accompanied by molecular motion, this is because we are misconstruing a situation that we are imagining or conceiving of. In particular, he says, when we have an impression of possibility of the sort in question, we have it because (a) we are imagining or conceiving of a situation in which *being the external phenomenon that causes the sensation S* is exemplified but there is no molecular motion, and (b) we are misconstruing this situation as one in which heat is present but molecular motion is not. That is to say, as Kripke sees it, we are confusing a situation in which the property by which we normally identify heat is exemplified with a situation in which heat itself is present.

Before concluding this review of Kripke's line of thought, we should recall that Kripke nowhere provides any defense of the assumption that I have listed as the seventh premise, the assumption that the explanatory paradigm illustrated by the heat example is the *only* paradigm for explaining appearances of possibility away.⁴ As we will see, this premise is false. Indeed, it is irredeemably false, in the sense that there is no true claim, or set of true claims, that is capable of taking its place in the argument.

II

I will now sketch an explanation of the apparent separability of pain and C-fiber stimulation that does not conform to the paradigm we have just been considering. Instead of exploiting the distinction between a property and its mode of presentation (that is, the distinction between a property X and the property that guides us in recognizing instances of X), this alternative explanation exploits a distinction between two types of imagination.

The explanatory strategy I have in mind is set forth in a little-noticed footnote in Thomas Nagel's wonderful "What Is It Like to Be a Bat?"⁵ The relevant part of the footnote runs as follows:

A theory that explained how the mind-brain relation was necessary would still leave us with Kripke's problem of explaining why it nevertheless appears contingent. That difficulty seems to me surmountable, in the following way. We may imagine something by representing it to ourselves either perceptually, sympathetically, or symbolically. I shall not try to say how symbolic imagination works, but part of what happens in the other two cases is this. To imagine something perceptually, we put ourselves in a conscious state resembling the state we would be in if we perceived it. To imagine something sympathetically, we put ourselves in a conscious state resembling the thing itself. (This method can only be used to imagine mental events and states – our own or another's.) When we try to imagine a mental state occurring without its associated brain state, we first sympathetically imagine the occurrence of the mental state: that is, we put ourselves in a state that resembles it mentally. At the same time, we attempt to perceptually imagine the non-occurrence of the associated physical state, by putting ourselves into another state unconnected with the first: one resembling that which we would be in if we perceived the non-occurrence of the physical state. Where the imagination of physical features is perceptual and the imagination of mental features is sympathetic, it appears to us that we can imagine any experience occurring without its associated brain state, and vice versa. The relation between them will appear contingent even if it is necessary, because of the independence of the disparate types of imagination.⁶

According to Nagel, then, there are three different types of imagination; the symbolic, the sympathetic, and the perceptual. We use the sympathetic imagination in imagining mental states and the perceptual imagination in imagining brain processes. It is also part of Nagel's story that we can use the perceptual imagination to form images of the *absence* of brain processes, that is, images that represent situations in which no brain processes exist. Nagel uses this story as the foundation of an account of how we are able to imagine disembodiment. According to Nagel's account, when we imagine a mental state that is not accompanied by a brain process, what we do is to splice together a sympathetic image of a situation that contains a mental state and a perceptual image of a situation in which no brain process is present. The fact that we can do this without incoherence has no tendency to show that it is objectively possible for the imagined mental state to exist without being accompanied by a brain process. Rather our ability to do it is due to the fact that there are two types of imagination that operate independently of one another.

The key element of this explanation is a claim that can be expressed as follows: If P is a property of which one can be introspectively or perceptually aware, then, when one imagines an instance of P, what one does is to put oneself into a state which is similar to the state one is in when one is experientially aware of an instance of P. This claim is intuitively plausible, and it also seems to be supported by the beautiful experimental results about images and the imagination that psychologists obtained in the seventies and eighties. Thus, there are many results, such as the classic experiments by Roger Shepherd and his colleagues concerning the processes by which we determine congruence and incongruence relations between geometric forms, which strongly suggest that visual images of objects and situations have many of the same properties as the corresponding perceptual states.⁷ Moreover, Nagel's claim receives additional support from the results which indicate that imagination can facilitate perceptual tasks. Ronald A. Finke summarizes these results in the following passage:

[T]he process of forming a visual image can serve a perceptual anticipatory function: it can prepare a person to receive information about imagined objects. Mental imagery may therefore enhance the perception of an object by causing the selective priming of mechanisms in the visual system.⁸

As Finke points out, this priming hypothesis provides strong support for the view that the process of forming a mental image involves many of the neural mechanisms that underlie visual perception.⁹

This brings us to the question of whether Nagel's explanation is genuinely compatible with materialism. At first sight, at least, the answer appears to be "yes." Consider, for example, the doctrine that pain is the very same thing as an electrochemical process in one's C-fibers. If this doctrine is true, one might ask, then how is it possible for experiences that count as imaginative presentations of pains to be quite different in character from experiences that count as imaginative presentations of cases of C-fiber stimulation? Nagel can respond by saying that the experiences are different because they are produced by different psychological mechanisms: experiences that count as imaginative presentations of pains are produced by mechanisms that serve the sympathetic imagination, and experiences that count as imaginative presentations of C-fiber stimulation are produced by mechanisms that serve the perceptual imagination. But is it possible for psychological mechanisms to produce radically different experiences if those experiences are in fact presentations of the same property? Well, yes; this happens all of the time in the case of perceptual presentations. Compare a visual presentation of the surface of a piece of sandpaper with a tactual presentation of the same surface. Any two such presentations will be quite different in point of intrinsic character; but still, the properties that are presented by the former will overlap with the properties that are presented by the latter.

This answer is plausible; but it seems that there is an aspect of the worry about the compatibility of Nagel's explanation with materialism that it fails to address. Let us trade our hypothesis about pain in for the slightly more concrete hypothesis that pain is the very same thing as electrochemical activity in a network of C-fibers that is essentially G-shaped, where G is a certain geometric structure. Now it can seem that if this hypothesis is true, then any presentation of pain, whether a genuine case of awareness or an imaginative construction, should have features which indicate that what is presented is activity in a network that essentially exemplifies G. In other words, it can seem that there should be something G-ish about both real and imagined pains, something that would present a foothold

for the concept of G. If G-ishness is part of the very essence of pain, shouldn't we be able to detect at least a hint of G-ishness by inspecting a presentation of pain?

I think that this intuition is quite natural, but I also think that it is due to a confusion. If presentations of pain do not present a foothold for the concept of G, this is because the mechanisms that control the applicability of the concept are exclusively perceptual, in the sense that they are exclusively responsive to features of perceptual presentations. Thus, like all other concepts with spatial significance, the concept of G owes its existence to our need to have concepts that make it possible for us to classify extramental entities that are given to us perceptually; and, as a result, the presentations that are capable of triggering the concept of G are limited to veridical presentations of instances of G and to other mental states that are similar to such presentations. Presentations of pain, whether genuine states of awareness or imaginative constructions, are quite different from veridical presentations of instances of G. Hence, when we examine such presentations with a view to detecting a hint of G-ishness, it is inevitable that we will come up empty handed.

It appears, then, that Nagel's account of how we are able to imagine disembodiment has a couple of important virtues. Unfortunately, it is also true that it has a flaw. Nagel's account presupposes that it is possible for us to perceive brain processes, and this presupposition is highly questionable. To be sure, we are able to use the naked eye to perceive whole brains and various parts of brains. Further, by focussing microscopes on preparations of dead tissue, we are able to perceive certain aspects of the structure of individual brain cells. But neither of these things count as perceiving electrochemical activity in living neurons. Accordingly, it seems reasonable to say that brain processes lie on the theoretical side of the fuzzy line that divides theoretical entities from observable entities. Our access to brain processes is mediated by theories. We cannot be said to perceive them.

If it is wrong to say that one can perceive a brain process, then it must also be wrong to say that one imagines a brain process by putting oneself into a state which resembles the state one is in when one perceives a brain process. So Nagel's account of what is involved

in imagining a brain process is mistaken. We must replace it with a more realistic picture.

Reflection suggests that there are two ways of imagining a brain process. First, it is possible to do so by putting oneself into a state which resembles the state one is in when one *indirectly* perceives a brain process, that is, by putting oneself into a state which resembles the state one is in when the following conditions are satisfied: (a) one is perceiving a piece of apparatus whose function it is to detect the presence of various kinds of brain process; (b) the apparatus indicates that a brain process of a certain type T is currently occurring; and (c) one is aware that the apparatus indicates that a process of type T is occurring. Second, it is possible to imagine a brain process by putting oneself into a state which resembles the state one is in when one is perceiving a *model* of a brain process and one is perceiving it *as* a model, that is, by putting oneself into a state which resembles the state one is in when the following conditions are satisfied: (d) one is perceiving a series of events that are more accessible to the senses than brain processes; (e) there is a structural isomorphism between the series of events and the segments of a certain brain process; (f) one is aware of this isomorphism; and (g) in perceiving the series, one makes use of one's awareness of the isomorphism, that is, one thinks of the series as a representation of the brain process in question.

Assuming that this account is more or less correct, we can explain what is involved in imagining the *absence* of a brain process by saying that one does this by putting oneself in an imaginative state that fails to satisfy conditions (a)-(c) and that also fails to satisfy conditions (d)-(g). (If it should turn out that this account is inadequate, we can improve it by adding that requirement that the state must include the *thought* that the situation one is imagining is devoid of brain processes.)

At this point, we should take note of the fact that the explanatory paradigm that is developed above can be applied not only to explain intuitions about disembodiment, but also to explain intuitions about the possibility of zombies. Just as intuitions of the first kind can be explained as due to one's sympathetically imagining the presence of pain while "perceptually" imagining the absence of C-fiber stimulation, so also intuitions of the second kind can be explained as

due to one's sympathetically imagining the absence of pain while "perceptually" imagining the presence of C-fiber stimulation. (I put "perceptually" in scare quotes to remind the reader of the need for qualifications of the sort described two paragraphs back – that is, qualifications based on conditions like (a)–(c) and (d)–(g).) It is clear, I think, that the possibility of forming imaginative representations of the latter sort is no less compatible with type-materialism than the possibility of forming imaginative representations of the former sort.

Before concluding these Nagelian reflections, we must consider one further issue. Although I have tried to show that Nagel's footnote offers us an adequate explanation of certain of our intuitions about disembodiment and zombiehood, I have not yet indicated how it might be true that our Nagelian account explains these intuitions *away*. One explains a set of intuitions by describing the mechanisms that produce them. To explain the members of the set *away* one must in addition provide evidence which calls the reliability of the relevant mechanisms into question, i.e., one must produce evidence which makes it reasonable to doubt that the intuitions produced by these mechanisms are quite likely to be true.

To see that there are grounds for doubting the reliability of intuitions about the possibility of disembodiment and zombiehood, we should first take note of the fact that other intuitions about separability can plausibly be claimed to be due to mechanisms of the same sort as the mechanisms I have claimed to be responsible for these "Cartesian" intuitions. Here is a description of one of the "Cartesian" mechanisms: it produces intuitions to the effect that pain can occur without being accompanied by C-fiber stimulation by splicing together images of the presence of pain and images of the absence of C-fiber stimulation. Now consider intuitions to the effect that heat can occur without being accompanied by molecular motion. It is plausible, I think, to attribute such intuitions to a mechanism that operates by splicing together images of the presence of heat and images of the absence of molecular motion. Or consider intuitions to the effect that water can exist without being accompanied by H_2O . It is plausible to attribute such intuitions to a mechanism that operates by splicing together images of the presence of water and images of the absence of H_2O .

Summarizing, we can say that the mechanisms that are responsible for intuitions about the possibility of disembodiment and zombiehood are members of a family *F* of mechanisms that can be described as follows: where *M* is any psychological mechanism, *M* is a member of *F* if *M* operates by splicing together images of the presence (or absence) of a commonsense phenomenon (i.e., a phenomenon to which we have access by a commonsense faculty of awareness, such as introspection or visual perception) and images of the absence (or presence) of a theoretical phenomenon (i.e., a phenomenon to which we have access only via theory-construction and laboratory apparatus). Just as it is plausible to say that intuitions about the possibility of disembodiment and zombiehood are due to the Nagelian mechanisms that I have described above, so also, I wish to claim, it is plausible to say that our other intuitions about the separability of commonsense phenomena and theoretical phenomena are due, or at least tend to be due, to other members of *F*.^{10,11}

We are in a position to assess the reliability of many of the members of *F* without begging any questions concerning the mind-body problem, and when we make such assessments, we find that the results are quite negative. Intuitions to the effect that heat is separable from molecular motion are false, as are intuitions to the effect that water is separable from H_2O , as are intuitions to the effect that light is separable from electromagnetic radiation. And so on. Since it is plausible to say that the Nagelian mechanisms that are responsible for intuitions about disembodiment and zombiehood are fundamentally akin to the mechanisms that are responsible for other intuitions about the separability of commonsense and theoretical kinds, and since we know that the latter mechanisms are fundamentally unreliable, the type-materialist can claim to be in possession of an argument which makes it rational to doubt the reliability of Nagelian mechanisms.

To be sure, there is a significant difference between Nagelian mechanisms and the other members of *F* – Nagelian mechanisms make use of the sympathetic imagination as well as the perceptual imagination, but presumably this is not true, for example, of the mechanisms that are responsible for intuitions about heat and molecular motion. (Perhaps we could describe the non-Nagelian members of *F* as mechanisms that splice together images provided by the perceptual imagination and images provided by the “perceptual”

imagination.) It would be implausible to maintain, however, that this difference makes the Nagelian mechanisms more reliable than the other members of F.

III

We have thus far considered only intuitions about separability that are due to the imagination. Now we must take note of the fact that there are also intuitions about separability that are due to the faculty that is responsible for our ability to *conceive* of possible situations (hereafter called the *faculty of conception*). These two sources are quite different: the representations constructed by the imagination are largely qualitative in character, whereas the representations constructed by the faculty of conception are largely conceptual and propositional. (In saying that the representations constructed by the imagination are largely qualitative, I mean that, like sensory states, they have defining characteristics that are qualitative and that, again like sensory states, their ability to represent things is largely a function of these qualitative characteristics.) Because of this difference, our explanatory task is not yet complete. We have succeeded in explaining away intuitions concerning the separability of sensory states and brain states that are due to the imagination, but what about the corresponding intuitions that are due to the faculty of conception? Can they be explained away as well? If so, how?

To prevent misunderstanding, I should acknowledge that the imagination and the faculty of conception are importantly connected, in the sense that each tends to draw upon the resources of the other in constructing representations of possibilities. Suppose I imagine my daughter reading a book about ancient Mayan civilization. My representation of this situation will have a central qualitative component, depicting a young woman in a chair with a book in her hands, but it seems inevitable that it will also have an important conceptual component. How else could it invoke the highly abstract characteristic *being a book about the ancient Mayans*? Or suppose that while reading *The Warden* I form a conceptual representation of one of the scenes – say, a representation of a conversation between Mr. Harding and Archdeacon Grantly. Here my representation will consist largely of internal counterparts of the sentences that Trollope uses

in his description of the conversation. But it seems likely (though not inevitable) that my representation will have a significant qualitative dimension as well. My representation will very likely include a qualitative state which is like the perceptual state I would be in if I were observe two gentlemen who actually possess the physical features that Trollope attributes to his characters.

Thus, I do not wish to deny that the imagination and the faculty of conception are closely aligned. Rather, my point is this: However closely aligned these two powers may be in practice, it is in principle possible to use the faculty of conception to construct representations that are largely or entirely without a qualitative dimension. Because of this in-principle-independence of the faculty of conception, it is inappropriate to claim to have a basis for explaining a set of intuitions that are due to this faculty simply in virtue of having a basis for explaining the corresponding intuitions that are due to the imagination.

It appears, then, that it really is necessary to supplement the explanation in section II with a second explanation, a separate explanation which seeks to account for the intuitions of separability that are due to the faculty of conception. As a first step, I will introduce a quasi-technical notion – the notion of a *sensory concept*. Sensory concepts have two features which collectively distinguish them from all other concepts. First, trivially, if *X* is a sensory concept, then *X* denotes a type of sensory state. And second, if *X* is a sensory concept, and *S* is the sensory state that is denoted by *X*, then, in an important range of contexts in which one is justified in ascribing *X* to an instance of *S* (specifically, the contexts in which one is led by introspection to ascribe *X* to one of one's *own* sensory states), the sense experience that provides one's justification for the ascription is identical with the instance of *S* to which *X* is applied.

It follows from this characterization that the concept of pain is a sensory concept. This follows because, in the first place, the concept of pain denotes a sensory state, and in the second place, it is constitutive of the concept of pain that when one is justified in ascribing the concept to one of one's own sensory states, one is so justified in virtue of the fact that one is currently in pain.

It also follows that sensory concepts are importantly different from neuroscientific concepts – even if, as type-materialism

maintains, sensory concepts coincide with certain neuroscientific concepts in point of denotation. Assuming that type-materialism is correct, the concept of pain denotes the same state as the concept of C-fiber stimulation. But it is nevertheless true that these concepts are quite different in virtue of having different justification conditions. As we just noticed, ascriptions of the concept of pain to one's own states are paradigmatically justified by experiences that count as instances of pain. But the justification conditions for paradigmatic ascriptions of the concept of C-fiber stimulation have nothing to do with pain, even when the ascriptions are self-ascriptions. Rather, the experiences that are paradigmatically relevant to justifying ascriptions of the concept of C-fiber stimulation are of the following two kinds: first, experiences of the sort that confirm the neuroscientific theories in which the concept of C-fiber stimulation is embedded, and second, experiences of the sort that one has when one is observing an apparatus which indicates that there is an instance of C-fiber stimulation in one's current environment. (Experiences of the second sort might include, for example, the experiences that are associated with observing images generated by PET-scan devices.)

A qualification. It is somewhat tendentious to say, as I have said above, that ascriptions of concepts like the concept of pain to one's own states are justified by the occurrence of instances of the states to which the concepts refer. A number of philosophers prefer to say that ascriptions of the concept of pain to one's own states are not justified by sensory evidence of any kind. Thus, some hold that such ascriptions are self-justifying, and others think that they are *beyond* justification – that they neither admit of justification nor require it. Fortunately, my present purposes can be achieved even if the presuppositions of the foregoing characterization of sensory concepts is rejected. All that is strictly necessary for present purposes is that the reader grant two things; first, that concepts like the concept of pain may legitimately be said to have denotations (and in particular, that they may legitimately be said to denote sensory states); and second, that the justification conditions associated with these concepts have peculiarities that distinguish them from the justification conditions that are associated with neuroscientific concepts. If these two assumptions are granted, then everything else that I wish to say will go through.

We are now in a position to address the main question of the present section: Is it possible to explain the intuitions of separability regarding pain and C-fiber stimulation that derive from the faculty of conception in a way that is compatible with type-materialism? I will try to show that the answer should be “yes.”

The explanation I have in mind can be obtained by pruning, qualifying, and supplementing a construction that I presented in an earlier paper.¹² It consists of the following four propositions:

- (1) There are no substantive a priori ties between sensory concepts and neuroscientific concepts. That is to say, apart from various principles that are logically true (e.g., “For all *x*, if *x* is in pain, then either *x*’s C-fibers are stimulated or *x* in pain”), there are no principles involving both sensory concepts and neuroscientific concepts that can be known a priori to be true.
- (2) If there are no substantive a priori ties between two concepts, then it is possible to conjoin either of the concepts with the negation of the other without producing an inconsistency. That is to say, it is possible to use the concepts to conceive coherently of situations (i.e., to construct internally coherent descriptions that purport to represent situations) in which there are particulars that fall under one of the concepts but do not fall under the other concept.
- (3) If it is within a subject’s power to conceive coherently of its being the case that *p*, then, unless the subject has a good a posteriori reason to think that the proposition that *p* could turn out to be inconsistent with a set of propositions that have been shown independently to be necessary, the subject will come to believe that it is possible – or at any rate, come to believe that it is *prima facie* possible – for it to be the case that *p*.
- (4) If type-materialism is true at all, then, since it is in effect a set of identity statements involving rigid designators, it is necessarily true. However, due to the fact that type-materialism is not part of our commonsense metaphysics,

we do not *normally* regard ourselves as having a good a posteriori reason to think that propositions about disembodiment and the existence of zombies are incompatible with necessary truths that have been independently established. (Only the enlightened regard themselves as having such a reason!¹³)

All four of these claims are plausible. Thus, consider (1). Among other things, (1) asserts that there are no substantive a priori ties between the concept of pain and the concept of C-fiber stimulation. This is surely correct; it is in principle possible to master either of these concepts fully without having mastered the other. Moreover, as with (1), so also with (2)–(4) – reflection on them serves only to endear them to us further.

It is reasonable to view (1)–(4) as providing an explanation of the sort we are seeking. Thus, one of the components of (1)–(4) (viz., (3)) is a lawlike generalization, and two of the other components (viz., (1) and (4)) can be seen as descriptions of initial conditions. Moreover, (1)–(4) collectively imply the proposition that the faculty of conception is disposed to generate intuitions to the effect that sensory states and brain states are separable. Because of these features, (1)–(4) count as a deductive-nomological explanation of the intuitions that we have been concerned to understand.

Having found an explanation of these intuitions, we must ask whether a type-materialist can use this explanation as a basis for explaining the intuitions *away*. Now in order to do this, a type-materialist must provide reason to doubt that the mechanisms cited in the explanation are fully reliable, i.e., reason to doubt that the intuitions produced by the mechanisms in question are quite likely to be true. I will now try to show that this can easily be accomplished.

In effect, the explanation based on (1)–(4) claims that there is a class H of psychological mechanisms whose members work as follows: where M is any member of H, M takes two concepts as inputs, and then, provided that it is possible to conjoin each of the concepts with the negation of the other without generating an inconsistency, and provided also that there is no available a posteriori reason to think that the two concepts are necessarily coextensive, M delivers an intuition of possibility as an output. In other words, the explanation claims that there is a class H of mechanisms such

that, where *M* is any member of *H*, and *A* and *B* are any two properties, *M* produces an intuition to the effect that *A* is separable from *B*, provided only, first, that the concept of *A* is not connected by any analytic ties to the concept of *B*, and second, that there is no available a posteriori reason to think that the former concept is necessarily coextensive with the latter concept. Now it is clear that many mechanisms that belong to *H* are highly unreliable. Thus, for example, (2) and (3) imply that there is a member *M* of *H* such that *M* is capable of producing intuitions to the effect that heat is separable from molecular motion. This member of *H* is extremely unreliable; for as we know, all such intuitions are erroneous – it is impossible to have a situation in which heat is present that is not also a situation in which molecular motion is present.

To amplify: Strictly speaking, of course, (2) and (3) do not *by themselves* imply that there is a member of *H* that is capable of producing intuitions to the effect that heat is separable from molecular motion. Rather, (2) and (3) yield this implication when they are combined with two additional claims: first, the claim that there are no analytic links between the concept of heat and the concept of molecular motion; and second, the claim that we are capable of lacking a posteriori reasons for thinking that these two concepts are necessarily coextensive – that is, the claim that human beings can fail to be in possession of the standard scientific reasons for affirming the necessary coextensiveness of the concepts (and also, fail to be in possession of any other reasons for affirming this identity).¹⁴ I take it that neither of these claims is controversial.

We have observed that the mechanisms that are responsible for conceivability-based intuitions about disembodiment and zombiehood belong to a certain category of mechanisms, *H*, and we have also observed that a number of the mechanisms that belong to *H* are highly unreliable. In view of these observations, it is reasonable for us to have doubts about the reliability of the mechanisms that are responsible for conceivability-based intuitions about disembodiment and zombiehood. Moreover, this conclusion can be strengthened. For these “Cartesian” mechanisms belong to a sub-category of *H* whose members are *uniformly* unreliable – specifically, the sub-category which consists of mechanisms that are responsible for conceivability-based intuitions to the effect that commonsense kinds

are separable from the theoretical kinds that are correlated with them. (H includes *all* mechanisms which take pairs of concepts as inputs and deliver intuitions of separability as outputs. The sub-category includes only those mechanisms that take certain restricted pairs of concepts as inputs, specifically, pairs whose first component is a concept of a commonsense kind and whose second component is a concept of the theoretical kind that is correlated with the given commonsense kind.) We are committed to holding that all of the non-“Cartesian” members of this sub-category are extremely unreliable: that is to say, we are committed to holding that if M is a mechanism that produces conceivability-based intuitions to the effect that a *physical* commonsense kind is separable from its correlated theoretical kind, then that mechanism is highly unreliable. It is clear, I think, that because of these commitments, we are also committed to holding that there is a strong inductive argument for the unreliability of the “Cartesian” members of the sub-category.

IV

So far so good; but we must now take note of a difficulty. Proposition (1) asserts that identity statements linking sensory concepts and neuroscientific concepts cannot be known a priori to be true. Type-materialists must embrace this assertion – both because it is clearly correct and because it is integral to the foregoing explanation of intuitions of separability. But now we must ask: “Are type-materialists in a position to explain the a posteriority of the identity statements in question? Can they explain why mastery of the relevant concepts is not a sufficient condition of coming to see that the statements are true?” There is a persuasive line of thought which indicates that the answer is “no.”¹⁵

The argument I have in mind runs as follows: “Compare the statement that pain is identical with C-fiber stimulation with any other identity statement concerning a commonsense kind and a scientific kind. Compare it, say, with the statement that heat is identical with the kinetic energy of molecules. Both of these statements are a posteriori. However, the latter statement differs from the former in that its a posteriority seems unproblematic – indeed, inevitable. When we are aware of heat from a commonsense perspective, we

are not directly aware of heat itself. We are not in touch with its intrinsic nature. Rather, we are aware of it as a phenomenon that has a certain characteristic effect on human observers; we are aware of it via a certain familiar sensation. In view of this fact about heat, it is no surprise that we cannot see *a priori* that heat involves molecular motion. On the other hand, when we turn to consider pain, we find that the situation is quite different. There is a commonsense perspective (namely, the perspective afforded by introspection) which enables us to be directly aware of pain. When we are aware of pain from this perspective, we are aware of it as in itself. We are presented with the intrinsic nature of pain. Accordingly, if it is true that pain is the very same thing as C-fiber stimulation, should we not be aware of it *as* C-fiber stimulation? Further, should it not be possible for us to see, simply in virtue of having mastered the use of the concept of pain and the concept of C-fiber stimulation, that these two concepts are coreferential? Suppose that Jones has fully mastered the use of both concepts. In order to have attained full mastery of the concept of pain, Jones must at some point have been introspectively acquainted with pain. Since introspective acquaintance with pain is always direct acquaintance, Jones must be fully apprised of the intrinsic nature of pain. Now suppose it is true that the concept of pain and the concept of C-fiber stimulation are coreferential. Given that Jones is fully apprised of the intrinsic nature of pain, and given also that the concept of C-fiber stimulation picks out its referent by describing its intrinsic nature, should it not be possible for Jones to appreciate this coreferentiality? Moreover, should it not be possible for him to appreciate it purely on the basis of reflection?"

This line of thought brings two facts to our attention. First, it shows that anyone who thinks that "Pain is identical with C-fiber stimulation" is true should find it a posteriority problematic, and should therefore feel an obligation to explain this a posteriority. And second, it shows that the task of providing an explanation is non-trivial. In particular, it shows that it will not do to invoke the explanatory paradigm that works in other, superficially similar cases, such as the case of heat and molecular motion.

Fortunately, there is a way out. The question before us now is closely related to one that we considered in section II – viz., the question "If pain is the very same state as the firing of a G-shaped

network of C-fibers, then why don't our experiences of pain provide a foothold for the concept of G?" We answered the earlier question by reminding ourselves of the fact that the concept of pain serves different purposes than the concept of G. Whereas the concept of pain is designed to make it possible for us to classify particulars on the basis of information that is available via introspection, the concept of G is designed to make it possible for us to classify particulars on the basis of information that reaches us via the senses. Because of this difference, we observed, it is inevitable that the concept of pain should be responsive to different sensory presentations than the concept of G. That is to say, it is inevitable that the experiences that are causally responsible for applications of the former concept should be different from the experiences that are causally responsible for applications of the latter concept. As I see it, the present difficulty admits of a similar solution.

Setting third person ascriptions of pain aside, we can, I think, make the following observation: When someone applies the concept of pain to a particular, the experience that causes him or her to make the application is identical with the particular to which the concept is being applied, and therefore, in the case of a non-erroneous application, the particular in question is a pain. On the other hand, like the concept of G, the concept of C-fiber stimulation is designed to classify particulars on the basis of information that reaches us via the senses. Because of this, the experiences that guide us in applying the concept of C-fiber stimulation are experiences whose occurrence is restricted to contexts in which we are using the senses to observe certain external phenomena. To be a bit more specific, the experiences that are germane to the concept of C-fiber stimulation are of two kinds. First, there are the experiences that motivate and confirm the neuroscientific theories in which the concept of C-fiber stimulation is embedded. And second, there are the experiences that lead us to apply the concept in particular contexts. (This second group includes, for example, experiences of the sort that occur when we are using the senses to observe images produced by PET scanners and MRI devices.)

We may conclude, I think, that there is a materialistically acceptable explanation of the fact that the identity of pain and C-fiber stimulation can only be known a posteriori. In full dress, this expla-

nation runs as follows: “The concept of pain and the concept of C-fiber stimulation serve classificatory purposes of quite different kinds. As we saw, the former concept is designed to enable us to classify phenomena on the basis of information that comes to us via introspection, and the latter concept is designed to enable us to classify particulars on the basis of information that comes to us via the senses. Because of this difference, the experiences that control our use of the former concept are inevitably quite different from the experiences that control our use of the latter concept. But now, when the use of one concept is controlled by experiences that are altogether disjoint from the experiences that control the use of a second concept, it is in general impossible for us to determine, simply in virtue of having mastered the use of the concepts, whether they are coreferential. It should come as no surprise, therefore, that it is possible to master the use of the concept of pain and the concept of C-fiber stimulation without appreciating their coreferentiality. It is inevitable that their identity be knowable only *a posteriori*.”

I close the present section by pointing out that there is a second way of expressing the difference between the purposes that are served by the concept of pain and the purposes that are served by the concept of C-fiber stimulation. Instead of expressing this difference by appealing to the distinction between information that comes to us via introspection and information that comes to us via the senses, we can express it by invoking the closely related distinction between being **IN** a state and **OBSERVING** a state. Thus, we can say that the concept of pain is designed to enable us to classify states that we are aware of in virtue of being **IN** them, and that the concept of C-fiber stimulation is designed to enable us to classify states that we are aware of in virtue of **OBSERVING** them, where “observing” is understood broadly, in such a way as to be applicable to states of awareness that are mediated by perceptions of such apparatus as PET scanners and MRI devices.

V

There is an important issue that has not yet been addressed. I have maintained that the mechanisms that are responsible for a number of our modal intuitions are unreliable. This might be taken to show

that the line of thought in the present paper presupposes a general scepticism about our ability to obtain knowledge of modal facts. And, by the same token, it might be taken to show that the line of thought in question suffers from a disabling flaw. No argument that calls the general run of our claims to have modal knowledge into question can be fully satisfactory, for we feel sure that we possess a fairly large amount of such knowledge.

Fortunately, there are two features of the argument that prevent it from leading to a radical modal scepticism. First, my claims about unreliability have been restricted to intuitions that have implications concerning a posteriori questions about matters of fact. (Hereafter I will call such intuitions *a posteriori modal intuitions*.) Thus, consider the intuition that heat is separable from molecular motion. When it is combined with the true thesis that properties are necessarily identical if they are identical at all, this intuition leads to a negative answer to the question of whether heat and molecular motion are identical – a question that is a paradigm of a posteriority. Accordingly, the intuition is an a posteriori modal intuition. Second, I have not assumed that we are *incapable* of forming a posteriori modal intuitions that are correct, but only that our a posteriori modal intuitions tend to be incorrect when we form them without being fully apprised of the relevant empirical facts. Thus, for example, in my account of conceivability-based intuitions concerning the separability of heat and molecular motion, I claimed that the mechanisms that are responsible for such intuitions operate only when we lack the scientific information which indicates that the concepts of heat and molecular motion are necessarily coextensive.

In sum, while it is true that the lines of thought in earlier sections call a class of modal intuitions into question, this class is highly restricted, consisting as it does only of a posteriori modal intuitions that are formed independently of information concerning the relevant empirical matters. Accordingly, the foregoing lines of thought have no tendency to promote a *general* scepticism about the possibility of obtaining modal knowledge. Moreover, it seems to me to be *good* to be sceptical about the intuitions that belong to the given restricted class; for as I see it, it follows from their description that such intuitions could not possibly be worthy of our trust! ¹⁶

NOTES

* This is an expanded version of a section of a paper that was read at the 1994 meetings of the Pacific Division of the American Philosophical Association. I thank W.D. Hart and Stephen Yablo for extremely useful comments on that earlier paper. In writing the present paper I have benefitted substantially from criticism and encouragement from David Chalmers, Ivan Fox, and Stephen Yablo.

¹ See Saul A. Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980). See also Kripke's "Identity and Necessity" in Stephen P. Schwartz (ed.), *Naming, Necessity, and Natural Kinds* (Ithaca, NY: Cornell University Press, 1977), 66–101.

² W.D. Hart, *Engines of the Soul* (Cambridge: Cambridge University Press, 1988); Stephen Yablo, "The Real Distinction Between Mind and Body," *Canadian Journal of Philosophy*, Supplementary Volume 16 (1990), 149–201; Stephen Yablo, "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* 53 (1993), 1–42; David Chalmers, *The Conscious Mind: In Search of a Theory of Consciousness* (Oxford: Oxford University Press, forthcoming).

³ Kripke, *Naming and Necessity*, p. 152.

⁴ Kripke says that the model illustrated by the heat example is "the only model [he] can think of." See Kripke, "Identity and Necessity," p. 101.

⁵ Thomas Nagel, "What Is It Like to Be a Bat?" *The Philosophical Review* 83 (1974), 435–450.

⁶ Quoted from Nagel, *ibid.*, footnote 11. Oddly, the ideas in this footnote have not received any attention from the contemporary defenders of Kripke's argument, nor, as far as I know, from any of Kripke's other commentators.

⁷ See, e.g. the papers collected in Roger N. Shepherd and Lynn Cooper (eds.), *Mental Images and Their Transformations* (Cambridge, MA: MIT Press, 1982).

Of course, the interpretation of Shepherd's results is a matter of controversy. Not everyone would agree with the assessment offered above. For a quite different view, see Zenon Pylyshyn, "The Imagery Debate: Analog Media Versus Tacit Knowledge," in Ned. Block (ed.), *Imagery* (Cambridge, MA: MIT Press, 1981), 151–205.

⁸ Ronald A. Finke, "Mental Imagery and the Visual System," *The Scientific American* 1986, 88–95. The quoted passage is on p. 92.

⁹ *Ibid.*, p. 88.

¹⁰ In effect, I am here proposing a new explanation of such intuitions as the intuition that heat is separable from molecular motion – that is, an explanation that is quite different than Kripke's explanation. According to Kripke, it seems to us that heat is separable from molecular motion because we take ourselves to be imagining a situation in which heat is present but molecular motion is absent when we are actually imagining a situation in which the *sensation* of heat is present but molecular motion is absent (or because we take ourselves to be imagining a situation in which molecular motion is present but heat is absent when we are actually imagining a situation in which molecular motion is present but the *sensation* of heat is absent). As I see it, this explanation is fundamentally misguided; for as I see it, in non-pathological circumstances introspection gives us pretty accurate access to the contents of our own states of imagination. Accordingly, I prefer to explain the intuition about the separability of heat and molecular motion, and all

related intuitions, such as the intuition that water is separable from H_2O , by saying that they are due to image-splicing mechanisms of the sort described in the text.

¹¹ As indicated in the previous note, I here mean to be proposing an explanation of imagination-based intuitions of separability. There is one respect in which the present sketch of the explanation is incomplete. As I see it, the “splicing” mechanisms operate to produce an intuition to the effect that it is possible that p only when a subject is not already in possession of a “defeater” for that intuition – that is, only when a subject is not already in possession of reasons for believing that it is necessary that not- p . Nothing is said about the inhibiting effects of such defeaters in the text.

I am abstracting from considerations having to do with defeaters in the present section because I want there to be one place in the paper in which the main idea is presented without being accompanied by a forest of qualifications. However, when we consider conceivability-based intuitions of separability in section III, I will be at pains to acknowledge the role of defeaters in inhibiting intuitions of separability. See especially the penultimate and ante-penultimate paragraphs of Section III.

¹² See Christopher S. Hill, “Why Cartesian Intuitions Are Compatible With the Identity Thesis,” *Philosophy and Phenomenological Research* 42 (1981), 254–265. There is a related discussion in my *Sensations* (Cambridge: Cambridge University Press, 1991), pp. 90–95 and 96–98.

Like Nagel’s footnote, my discussions of Kripke’s argument appear not to have been answered by any of the contemporary defenders of the argument.

¹³ Several arguments that have been known to lead to enlightenment are collected in the author’s *Sensations*, chapters 2–4.

¹⁴ Actually, there are *two* ways in which one can fail to be in possession of the standard a posteriori reasons for affirming the necessary coextensiveness of the concept of heat and the concept of molecular motion. First, as noted in the text, one can be ignorant of the scientific account of the nature of heat – that is, of the empirical theory which implies that heat is identical to molecular motion. Second, one can be in possession of this theory, but not be fully aware that the theory gives one grounds for asserting that heat is *necessarily* identical to molecular motion. That is to say, it is possible to be in possession of all of the relevant scientific facts, but to fail for one reason or another to appreciate fully and non-confusedly the metaphysical implications of those facts. (As I see it, there is a fairly strong tendency to fail to appreciate the metaphysical implications in question – a tendency that can only be fully eradicated by considering the Marcus/Kripke argument for the necessity of identity. This is why twentieth century philosophers have sometimes experienced intuitions to the effect that heat is separable from molecular motion.)

¹⁵ The line of thought presented here is in some ways similar to the “property dualism argument” that is presented in Stephen White, “Curse of the Qualia,” *Synthese* 68 (1986), 333–368. (See in particular pp. 351–353.) By the same token, the answer to this line of thought that I give in the remainder of section IV can be taken as a reply to White’s argument. (Also, I believe that readers of section IV will be able to see that it provides type-materialists with a basis for answering Frank Jackson’s knowledge argument. See Jackson’s “Epiphenomenal Qualia,” *Philosophical Quarterly* 32 (1982), 127–136, and his “What Mary Didn’t Know,” *Journal of Philosophy* 83 (1986), 291–295.)

¹⁶ The question broached in this final section deserves a great deal more attention than I am able to give it in a paper that is primarily concerned with other matters. I hope to return to it on another occasion.

Department of Philosophy
University of Arkansas
Fayetteville, AR 72701
USA