sciendo

Petrus Hispanus Lectures 2003
The Harder Problem of Consciousness

Ned Block

New York University

Disputatio No. 8

November 2003

DOI: 10.2478/disp-2003-0007

ISSN: 0873-626X

Disputatio 15, November 2003

Petrus Hispanus Lectures 2003

The harder problem of consciousness[1]

Ned Block
New York University

I. The hard problem

T. H. Huxley famously said 'How it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of Djin when Aladdin rubbed his lamp.'[2] We do not see how to explain a state of consciousness in terms of its neurological basis. This is the Hard Problem of Consciousness.[3]

The aim of this paper is to present another problem of consciousness. The Harder Problem as I will call it is more epistemological than the

---

[1] This is a longer version of a paper by the same name that appeared in *The Journal of Philosophy,* XCIX, 8, August 2002, 391-425.

[2] T. H. Huxley, *Lessons in Elementary Physiology*. London: Macmillan, 1986, p.193; See Güven Güzeldere, 1997 'The Many Faces of Consciousness: A Field Guide' in Ned Block, Owen Flanagan and Güven Güzeldere, *The Nature of Consciousness: Philosophical Debates*, Cambridge: MIT Press, 1997, 1-67, footnote 6.

[3] See Thomas Nagel ('What is it like to be a bat?' *Philosophical Review* 83: 435-450, 1974). Joe Levine introduced the 'explanatory gap' terminology (Joe Levine, 'Materialism and qualia: the explanatory gap,' *Pacific Philosophical Quarterly* 64, 1983: 354-361) to be used later. David Chalmers and Galen Strawson distinguished between the hard problem and various 'easy problems' of how consciousness functions (David Chalmers, *The Conscious Mind*. New York: Oxford University Press, 1996, pp xxii-xxiii. Galen Strawson, *Mental Reality*. Cambridge: MIT Press, 1994, pp. 93-96).

Hard Problem. A second difference: the Hard Problem could arise for someone who has no conception of another person, whereas the Harder Problem is tied closely to the problem of other minds. Finally, the Harder Problem reveals an epistemic tension or at least discomfort in our ordinary conception of consciousness that is not suggested by the Hard Problem, and so in one respect it is harder. Perhaps the Harder Problem includes the Hard Problem and is best thought of as an epistemic add-on to it. Or perhaps they are in some other way facets of a single problem. Then my point is that this single problem breaks into two parts, one of which is more epistemic, involves other minds, and involves an epistemic discomfort.

II. Preliminaries

I believe that the major ontological disputes about the nature of consciousness rest on an opposition between two perspectives:

- Deflationism about consciousness, in which a priori or at least armchair analyses of consciousness (or at least armchair sufficient conditions) are given in non-phenomenal terms, most prominently in terms of representation, thought or function.
- Phenomenal realism, which consists in the denial of deflationism plus the claim that consciousness is something real. Phenomenal realism is metaphysical realism about consciousness and thus allows the possibility that there may be facts about the distribution of consciousness that are not accessible to us even though the relevant functional, cognitive and representational facts are accessible. Phenomenal realism is based on one's first person grasp of consciousness. An opponent might prefer to call phenomenal realism 'inflationism,' but I reject the suggestion of something bloated.

In its most straightforward version, deflationism is a thesis of a priori conceptual analysis, most prominently analysis of mental terms in functional terms. As David Lewis, a well known deflationist noted[4], this view is the heir of logical behaviorism. Phenomenal realism rejects these armchair philosophical reductive analyses. But phenomenal realists have no brief against *scientific* reduction of consciousness. Of course,

[4] David Lewis, 'An Argument for the Identity Theory,' *Journal of Philosophy* 63, 1966: 17-25.

there is no sharp line here, and since the distinction is epistemic, one and the same metaphysical thesis could be held both as a philosophical reductionist and as a scientific reductionist thesis.[5]

I apologize for all the 'isms' (deflationism, phenomenal realism and one more to come), but they are unavoidable since the point of this paper is that there is a tension between two of them. The tension is between phenomenal realism ('inflationism') and (scientific) naturalism, the epistemological perspective according to which the default view is that consciousness has a scientific nature — where this is taken to include the idea that conscious similarities have scientific natures. (A view on a given subject is the default if it is the only one for which background considerations give rational ground for tentative belief.) This paper argues for a conditional in which specifications of phenomenal realism and scientific naturalism (and a few other relatively uncontroversial items — including, notably, a rejection of a skeptical perspective) appear on the left hand side. On the right hand side we have a specification of the epistemic tension which I mentioned. Deflationists who accept the argument may opt for modus tollens, giving them a reason to reject phenomenal realism. Phenomenal realist naturalists may want to weaken their commitment to naturalism or to phenomenal realism. To put the point without explicit 'isms': Many of us are committed to the idea that consciousness is both real and can be assumed to have a scientific nature, but it turns out that these commitments do not fit together comfortably.

Modern phenomenal realism has often been strongly naturalistic (e.g. Levine, Loar, McGinn, Peacocke, Perry, Shoemaker, Searle and myself). Dennett has often accused phenomenal realists of closet dual-

---

[5] Deflationism with respect to truth is the view that the utility of the concept of truth can be explained disquotationally and that there can be no scientific reduction of truth. (Paul Horwich, *Truth,* Blackwell: Oxford, 1990. Second edition 1998, Oxford University Press: Oxford; Hartry Field, 'Deflationist Views of Meaning and Content,' *Mind* 103, 1994: 249-285.) Deflationism with respect to consciousness in its most influential form is, confusingly, a kind of reductionism — albeit armchair reductionism rather than substantive scientific reductionism — and thus the terminology I am following can be misleading. I may have introduced this confusing terminology (in my 1992 reply to Dennett and Kinsbourne, reprinted in Block, Flanagan and Güzeldere, op. cit., p. 177; and also in my review of Dennett in *Journal of Philosophy*, pp. 181-93, 1993).

ism. Rey has argued that the concept of consciousness is incoherent.[6] The upshot of this paper is that there is a grain of truth in these accusations.

Before I go on, I must make a terminological comment. Imagine two persons both of whom are in pain, but only one of whom is introspecting his pain state and is in that sense conscious of it. One could say that only one of the two has a *conscious* pain. This is *not* the sense of 'conscious' used here. In the sense of 'conscious' used here, just in virtue of *having* pain, *both* have conscious states. To avoid verbal disputes, we could call the sense of 'consciousness' used here *phenomenality*. Pains are intrinsically phenomenal and *in that sense* are intrinsically conscious. In that sense — but not in some other senses — there cannot be an unconscious pain.

The plan of the paper is this: first I will briefly characterize the Hard Problem, mainly in order to distinguish it from the Harder Problem. I will argue that the Hard Problem can be dissolved only to reappear in a somewhat different form, but that in this different form we can see a glimmer of hope for how a solution might one day be found. I will then move on to the Harder Problem, its significance and a comparison between the Hard and Harder Problems. I will conclude with some reflections on what options there are for the naturalistic phenomenal realist.

III. Mind-body identity and the apparent dissolution of the hard problem

The Hard Problem is one of explaining why the neural basis of a phenomenal quality is the neural basis of *that* phenomenal quality rather than another phenomenal quality or no phenomenal quality at all. In

---

[6] Georges Rey, 'A reason for doubting the existence of consciousness.' In *Consciousness and Self-Regulation*, vol 3. R. Davidson, G. Schwartz, D. Shapiro (eds.) Plenum, 1983, 1-39. In previous publications ('On a Confusion about a Function of Consciousness,' *The Behavioral and Brain Sciences* 18, 2, 1995, 227-247), I have argued that Rey's alleged incoherence derives from his failure to distinguish between phenomenal consciousness and other forms of consciousness (what I call access consciousness and reflexive consciousness). The incoherence that is the subject of this paper, by contrast, is an incoherence in phenomenal consciousness itself.

other terms, there is an explanatory gap between the neural basis of a phenomenal quality and the phenomenal quality itself. Suppose (to replace the neurologically ridiculous example of c-fibers that is often used by philosophers with a view proposed as a theory of visual experience by Crick and Koch[7]) that cortico-thalamic oscillation (of a certain sort) is the neural basis of an experience with phenomenal quality Q. Now there is a simple (over-simple) physicalist dissolution to the Hard Problem that is based on mind-body identity: Phenomenal quality Q = cortico thalamic oscillation (of a certain sort). Here's a statement of the solution:

> The Hard Problem is illusory. One might as well ask why H_2O is the chemical basis of water rather than gasoline or nothing at all. Just as water *is* its chemical basis, so Q just is its neural basis (cortico-thalamic oscillation), and that shows the original question is wrongheaded

I think there is something right about this answer but it is nonetheless unsatisfactory. What is right about it is that if Q = cortico-thalamic oscillation, that identity itself, like all genuine identities, is inexplicable.[8] What is wrong about it is that we are in a completely different epistemic position with respect to such a mind-body identity claim than we are with respect to 'water = H_2O.' The claim that Q is identical to cortico-thalamic oscillation is just as puzzling — maybe more puzzling — than the claim that the physical basis of Q is cortico-thalamic oscillation. We have no idea how it could be that one property could be identical both to Q and cortico-thalamic oscillation. How could one property be both subjective and objective? Although no one can explain

---

7 Francis Crick and Christof Koch, 'Towards a neurobiological theory of consciousness.' *Seminars in the Neurosciences* 2, 1990: 263-275.

8 We can reasonably wonder how it is that Mark Twain and Samuel Clemens married women with the same name, lived in the same city, etc. But we cannot reasonably wonder how it is that Mark Twain *is* Samuel Clemens. Imagine two groups of historians in the distant future finding a correlation between events in the life of Clemens and Twain. The identity explains such correlations, but it cannot itself be questioned. This point is made in Ned Block, 'Reductionism,' *Encyclopedia of Bioethics*, Macmillan, 1978, 1419-1424. See also Ned Block and Robert Stalnaker, 'Conceptual Analysis and the Explanatory Gap,' *The Philosophical Review,* January, 1999; and David Papineau, 'Consciousness, Physicalism and the Antipathetic Fallacy,' *Australasian Journal of Philosophy*, 1993. For a statement of a contrary view, see Chalmers, op. cit.

an identity, we can remove puzzlement by explaining how an identity can be true, most obviously, how it is that the two concepts involved can pick out the same thing. This is what we need in the case of subjective/objective identities such as the putative identity that Q = cortico-thalamic oscillation.

Joe Levine[9] argues that there are two kinds of identities, those like 'water = H_2O' which do not admit of explanation and those like 'the sensation of orange = cortico-thalamic oscillation' which are 'gappy identities' which do allow explanation. He argues that the 'left hand' mode of presentation of the latter is more substantive than those of the former. The idea is supposed to be that descriptive modes of presentation are 'pointers we aim at our internal states with very little substantive conception of what sort of thing we are pointing at — demonstrative arrows shot blindly that refer to whatever they hit.' By contrast, according to Levine, phenomenal modes of presentation really do give us a substantive idea of what they refer to, not a 'whatever they hit' idea. However, even if we accept this distinction, it will not serve to explain the 'gappiness' of mind-body identities. Consider that the mode of presentation of a sensation of a color can be the same as that of the color itself. Consider the identity 'Orange = yellowish red.' Both modes of presentation involved in this identity can be as substantive as those in the putatively 'gappy' identity just mentioned, yet this one is not 'gappy' even if some others are. To get an identity in which only one side is substantive, and is so a better analogy to the mind-body case, consider an assertion of 'orange = yellowish red' in which the left hand concept is phenomenal but the right hand concept is discursive.

IV. How to approach the hard problem

The standard arguments against physicalism (most recently by Jackson, Kripke and Chalmers) make it difficult to understand how mind-body identity could be true, so explaining how it could be true requires undermining those arguments. I will not attempt such a large task here, especially since the role of the discussion of the Hard Problem in this paper is mainly to contrast it with the Harder Problem to come. So I will limit my efforts in this direction to a brief discussion of Jackson's

[9] *Purple Haze*, Oxford: OUP, 2001.

famous 'knowledge' argument. I discuss this argument not because I think it is the most challenging argument against mind-body identity but rather because it motivates an apparatus which gives us some insight into what makes the Hard Problem hard. Jackson imagined a neuroscientist of the distant future (Mary) who is raised in a black and white room and who knows everything physical and functional that there is to know about color and the experience of it. But when she steps outside the room for the first time, she learns what it is like to see red. Jackson argued that since the physical and functional facts do not encompass the new fact that Mary learns, dualism is true.

The key to what is wrong with Jackson's argument (and to removing one kind of puzzlement about how a subjective property could be identical to an objective property) is the concept/property distinction.[10] Any account of this distinction as it applies to phenomenal concepts is bound to be controversial. I will use one such account without defending it, but nothing in the rest of the paper will be based on this account.

The expressions 'this sudden involuntary muscle contraction' and 'this [experience] thing in my leg' are two expressions that pick out the cramp I am now having in my leg. (These are versions of examples from Loar, op. cit.) In 'this [experience] thing in my leg,' attention to an experience of the cramp functions so as to pick out the referent, the cramp. (That is the meaning of the bracket notation. The 'this' in 'this [experience] thing in my leg' refers to the thing in my leg, not the experience.) The first way of thinking about the cramp is an objective concept of the cramp. The second is a subjective concept of the same thing — subjective in that there is a phenomenal mode of access to the thing picked out. Just as we can have both objective and subjective concepts of a cramp, we can also have objective and subjective concepts of a cramp *feeling*. Assuming physicalism, we could have an objective neurological concept of a cramp feeling, e.g. 'the phased locked 40 Hz oscillation that is occurring now.' And we could have a subjective concept of the same thing, 'this [experience] feeling.' Importantly, the same experience type could be part of — though function differently — in *both* subjective concepts, the subjective concept of the cramp and the

---

[10] The articles by Paul Churchland, Brian Loar, William Lycan and Robert van Gulick in Block, Flanagan and Güzeldere, op. cit. all take something like this line; as does Scott Sturgeon, 'The Epistemic View of Subjectivity,' *Journal of Philosophy,* XCI, 5, 1994; and Perry, op. cit.

subjective concept of the cramp feeling. Further, we could have both a subjective and objective concept of a single color. And we could have both a subjective and an objective concept of the experience of that color, and the same experience or mental image could function — albeit differently — in the two subjective concepts, one of the color, the other of the experience of the color.

Deflationists will not like this apparatus, but they should be interested in the upshot since it may be of use to them in rejecting the phenomenal realism in the antecedent of the conditional that this paper argues for.

Concepts in the sense used here are mental representations. For our purposes, we may as well suppose a system of representation that includes both quasi-linguistic elements as well as phenomenal elements such as experiences or mental images. Stretching terminology, we could call it a language of thought.[11]

In these terms, then, we can remove one type of puzzlement that is connected with the Hard Problem as follows: there is no problem about how a subjective property can be identical to an objective property. Subjectivity and objectivity are better seen as properties of *concepts* rather than properties of *properties*. The claim that an objective property is identical to a subjective property would be more revealingly expressed as the claim that an objective concept of a property picks out the same property as a subjective concept of that property. So we can substitute a dualism of concepts for a dualism of properties.

The same distinction helps us to solve the Mary problem. In the room, Mary knew about the subjective experience of red via the objective concept *cortico-thalamic oscillation*. On leaving the room, she acquires a subjective concept *this [mental image] phenomenal property* of the same subjective experience. In learning what it is like to see red, she does not learn a new fact. She knew about that fact in the room under an objective concept and she learns a new concept of that very fact. One

[11] Note that my account of subjective concepts allows for subjective concepts of many more colors or pitches than we can recognize, and thus my account differs from accounts of phenomenal concepts as *recognitional* concepts such as that of Loar, op. cit. On my view, one can have a phenomenal concept without being able to reidentify the same experience again. (See Sean Kelly, 'Demonstrative Concepts and Experience,' *Philosophical Review* 110, 3, 2001: 397-420, for arguments that experience outruns recognition.)

can acquire new knowledge about old facts by acquiring new concepts of those facts. New knowledge acquired in this way does not show that there are any facts beyond the physical facts. Of course it does require that there are concepts that are not physicalistic concepts, but that is not a form of dualism. (For purposes of this paper, we can think of physicalistic concepts as concepts couched in the vocabulary of physics. A physicalist can allow non-physicalistic vocabulary, e.g. the vocabulary of economics. Of course, physicalists say that everything is physical, including vocabulary. But the vocabulary of economics can be physical in that sense without being physicalististic in the sense of couched in the vocabulary of physics.)

Where are we? The Hard Problem in one form was: how can an objective property be identical to a subjective property? We now have a dissolution of one aspect of the problem, appealing to the fact that objectivity and subjectivity are best seen as properties of concepts. But that is no help in getting a sense of what *sorts* of objective concepts and subjective concepts could pick out the same property, and so it brings us no closer to actually getting such concepts. As Nagel (op. cit.) noted, we have no idea how there could be causal chains from an objective concept and a subjective concept leading back to the same phenomenon in the world. We are in something like the position of pre-Einsteinians who had no way of understanding how a concept of mass and a concept of energy could pick out the same thing.

V. Preliminaries before introducing the harder problem

Naturalism: Naturalism is the view that it is a default that consciousness has a scientific nature (and that similarities in consciousness have scientific natures). I will assume that the relevant sciences include physics, chemistry, biology, computational theory, and parts of psychology that don't explicitly involve consciousness. (The point of the last condition is to avoid the trivialization of naturalism that would result if we allowed the scientific nature of consciousness to be… consciousness.) I will lump these sciences together under the heading 'physical,' thinking of naturalism as the view that it is a default that consciousness is physical (and that similarities in consciousness are physical). So naturalism = default physicalism, and is thus a partly epistemic thesis. Naturalism in my sense recognizes that although the indirect evidence

for physicalism is impressive, there is little direct evidence for it. My naturalist is not a 'die-hard' naturalist, but rather one who takes physicalism as a default, a default that can be challenged. My rationale for defining 'naturalism' in this way is that this version of the doctrine is plausible, widely held, and leads to the epistemic tension that I am expositing. Some other doctrines that could be called 'naturalism' don't, but this one does. I think that my naturalism is close to what John Perry calls 'antecedent physicalism.' (See his *Knowledge, Possibility and Consciousness,* MIT Press: Cambridge, 2001.)

Functionalism: Functionalism and physicalism are usually considered competing theories of mind. However, for the purposes of this paper, the phenomenal realism/deflationism distinction is more important, and this distinction cross-cuts the distinction between functionalism and physicalism. In the terms used earlier, one type of functionalism is deflationist, the other phenomenal realist. The latter is Psychofunctionalism, the identification of phenomenality with a role property specified in terms of a psychological or neuropsychological theory.[12] At the beginning of the paper, I pointed to the somewhat vague distinction between philosophical and scientific reduction. Deflationist functionalism is a philosophical reductionist view whereas phenomenal realist Psychofunctionalism is a scientific reductionist view.

I will be making use of the notion of a superficial functional isomorph, a creature that is isomorphic to us with respect to those causal relations among mental states, inputs and outputs that are specified by common sense, or if you like, 'folk psychology.' Those who are skeptical about these notions should note that the point of the paper is that a nexus of standard views leads to a tension. This conceptual apparatus may be part of what should be rejected. Those who would like to see more on functionalism should consult any of the standard reference works such as *the Routledge Encyclopedia of Philosophy*. Or see http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/functionalism.html.

As I mentioned at the outset, this paper argues for a conditional. On the left side of the conditional are phenomenal realism and naturalism (plus conceptual apparatus of the sort just mentioned). My current point is that I am including Psychofunctionalism in the class of phenomenal

[12] Ned Block 'Troubles with Functionalism.' *Minnesota Studies in the Philosophy of Science* (C.W. Savage, ed.), Vol. IX, 1978, 261-325.

realist naturalist theories. Thus one kind of functionalism — the deflationist variety — is excluded by the antecedent of my conditional, and another — the phenomenal realist variety — is in the class of open options.

Anti-skeptical perspective: In what follows, I will be adopting a point of view that sets skepticism aside. '*Undoubtedly*, humans are conscious and rocks and laptops are not.' (Further, *bats* are undoubtedly conscious.) Of course, the anti-skeptical point of view I will be adopting is the one appropriate to a naturalist phenomenal realist. Notably, from the naturalist phenomenal realist perspective, the concept of a functional isomorph of us with no consciousness is not incoherent and the claim of bare possibility of such a zombie — so long as it is not alleged to be us — is not a form of skepticism.

Multiple realization/multiple constitution: Putnam, Fodor and Block and Fodor argued that if functionalism about the mind is true, physicalism is false.[13] The line of argument assumes that functional organizations are often — maybe even always — multiply realizable. The state of adding 2 cannot be identical to an electronic state if a non-electronic device (e.g. a brain) can add 2.

This 'multiple realizability' argument has become controversial lately[14], for reasons that I cannot go into here.[15] The argument I will be

---

[13] Hilary Putnam, 'Psychological Predicates,' later titled 'The nature of mental states.' In (Capitan & Merrill, eds.) *Art, Mind, and Religion*. Pittsburgh University Press, 1967. J. A. Fodor, 'Materialism,' Ch. 3 of *Psychological Explanation*. Random House: New York, 1968: 90-120. Ned Block & Jerry Fodor, 'What Psychological States are Not,' *Philosophical Review* 81, 1972: 159-81.

[14] The most important criticism is given in a paper by Jaegwon Kim. (Jaegwon Kim, 'Multiple realization and the metaphysics of reduction,' *Philosophy and Phenomenological Research* 52: 1-26, 1992. See also *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. MIT Press: Cambridge, 1998.) I believe that Kim's argument does not apply to phenomenality, as Kim himself hints. I will briefly summarize Kim's argument in this footnote and the reason why it does not apply to phenomenality later in Section VII. In 'Multiple realization and the metaphysics of reduction.' *Philosophy and Phenomenological Research* 52:1-26, 1992, Kim says, correctly I think, that Putnam (op. cit.) and Block and Fodor (op. cit) and Fodor (op. cit.) reject without adequate justification the option of identifying a multiply realizable special science property with the heterogeneous disjunctive property whose disjuncts are its physical realizers. (P is the disjunctive property whose disjuncts are F and $G \equiv P = \lambda x$ (Fx or

Gx). 'λx Fx' is read as the property of being an x such that Fx, i.e. F-ness.) Kim says that the nomic covariance of the special science property with the disjunction of physical realizers shows that the special science property is just as non-nomic as the heterogeneous physical disjunction. The upshot, he says, is that there are no special sciences. My 'Anti-reductionism Slaps Back,' *Mind, Causation, World, Philosophical Perspectives* 11, 1997, 107-133 replies by arguing that whether a property is nomic is *relative* to a *level* of science. Both the multiply realizable special science property and the disjunction of physical realizers are nomic relative to the special science level and both are non-nomic relative to the physical level. (A sketch of a different challenge is given in Section VII.)

Philip Kitcher and Elliott Sober have persuasively argued that certain biological kinds (e.g. fitness) are both multiply realizable and causal-explanatory. See Kitcher, '1953 and All That: A Tale of Two Sciences,' *Philosophical Review* XCIII, 1984; Sober, *The Nature of Selection: Evolutionary Theory in Philosophical Focus*, Cambridge: MIT, 1984. See also Alex Rosenberg, 'On Multiple Realization in the Special Sciences,' *Journal of Philosophy* XCVIII, 7, 2001. See also the wonderful example in Brian Keeley's 'Shocking Lessons from Electric Fish: The Theory and Practice of Multiple Realization,' *Philosophy of Science* 67, 2000.

[15] Kim accepts the standard argument that functionalism shows physicalism is false; though I do not think he would like that way of putting it. His stance is that of a deflationist functionalist about the mental. What makes human mental state M and Martian M both M is something functional, not something physical. However, he endorses structure-restricted physical identities: Martian M is one physical state, human M is another, and in that sense he is a physicalist. Since he is a physicalist — in that sense — and also a functionalist, he would not find the verbal formula that functionalism shows physicalism is false congenial.

Incidentally, the issue of multiple realization/reduction discussed here is quite different from the explanatory issue also discussed by Putnam and Fodor concerning whether macro phenomena always have micro explanations that subsume the macro explanations. See Elliot Sober, 'The Multiple Realizability Argument Against Reductionism,' *Philosophy of Science* 66, 542-564, on this issue.

William Bechtel and Jennifer Mundale, 'Multiple Realizability Revisited: Linking Cognitive and Neural States' *Philosophy of Science* 66, 1999, 175-207 argue that mental states of actual animals and people are not multiply realized. (In my terminology, they mean multiply constituted.) They note that when we are offered real examples of multiple realization, a closer analysis reveals small functional differences. The putative multiple realizers are at best *approximately* realizers of the same functional state. E.g. though language is normally based in the left hemisphere, people without a left hemisphere can learn language pretty well; but there are differences in their abilities to master difficult syntactic constructions. But the key issue — one that Bechtel and Mundale ignore and

giving is a version of the traditional multiple realizability argument (albeit an epistemic version), so I had better say a bit about what a realization is. One of the many notions of realization that would do for our purposes is the following. A functional state is a kind of second order property, a property which consists in having certain first order properties that have certain causes and effects.[16] For example, dormitivity in one sense of the term is the property a pill has of having some (first order) property that causes sleep. Provocativity is the property of having some (first order) property or other that makes bulls angry. We can speak of the first order property of being a barbiturate as being one realizer of dormitivity, or of red as being one realizer of provocativity.[17]

---

which undermines their point — is whether the *functional resemblances are explained by unitary properties at the realizer level*. For example, perhaps two adders that work in different ways always differ slightly, e.g. in the speed of adding. The question is whether the shared functional properties can be explained in terms of shared unitary properties at e.g. the microphysical level. In the case of adders, the answer is no.

16 The restriction to first order properties is unnecessary. See my 'Can the Mind Change the World,' in *Meaning and Method: Essays in Honor of Hilary Putnam*, edited by G. Boolos. Cambridge University Press: Cambridge, 1990.

17 An alternative notion of realization appeals to the notions of supervenience and explanation. The realized property supervenes on the realizer and the realizer explains the presence of the realized property. Possessing the realizer is one way in which a thing can possess the realized property. See Ernest Lepore, and Barry Loewer, 'Mind Matters,' *Journal of Philosophy* 93, 1987: 630-642, and Lenny Clapp, 'Disjunctive Properties: Multiple Realizations,' *Journal of Philosophy* XCVIII, 3, 2001.

Dormitivity in the sense mentioned is a second order property, the property of having some property that causes sleep. But one could also define dormitivity as a first order property, the property of causing sleep. That is, on this different definition, F is dormitive just in case F causes sleep. But if we want to ascribe dormitivity to *pills*, we will have to use the second order sense. What it is for a pill to be dormitive is for it, the pill, to have some property or other that causes sleep. Similarly, if we want a notion of functional property that applies to properties, the first order variant will do. But if we want to ascribe those properties to people, we need second order properties. What it is for a person to have pain, according to the functionalist, is for the person to have some property or other that has certain causal relations to other properties and to inputs and outputs.

If we understand realization, we can define constitution in terms of it. Suppose that mental state M has a functional role that is realized by neural state N. Then N constitutes M — relative to M playing the M-role. The point of the last condition is that ersatz M — a state functionally like M but missing something essential to M as phenomenality is to pain — would also have the M-role, but N would not constitute ersatz M merely in virtue of constituting M. So the M-role can be multiply realized even if mental state M is not multiply constituted.

There is an obvious obscurity in what counts as *multiple* realization (or constitution). We can agree that neural property X is distinct from neural property Y and that both realize a single functional property without agreeing on whether X and Y are variants of a single property or two substantially different properties, so we will not agree on whether there is genuinely multiple realization. And even if we agree that X and Y are substantially different, we may still not agree on whether the functional property is multiply realized since we may not agree on whether there is a single disjunctive realization. These issues will be discussed further in Section VII.

VI. Introducing the harder problem

My strategy will be to start with the epistemic possibility of multiple realization and use it to argue for the epistemic possibility of multiple constitution of mentality. I will then argue that the epistemic possibility of multiple constitution of phenomenal properties is problematic. I will use a science fiction example of a creature who is functionally the same as us but physically different. Those who hate science fiction should note that the same issue arises — in more complicated forms — with respect to real creatures, such as the octopus, which differ from us both physically and functionally.

(1) We have no reason to believe that there is any deep physical property in common to all and only the possible realizations of our superficial functional organization. Moreover — and this goes beyond what is needed for (1) — but it does make (1) more vivid: we have no reason to believe that we cannot find or *make* a merely superficial isomorph of ourselves. By 'merely superficial isomorph,' I mean an isomorph with respect to folk psychology and whatever is logically or nomologically entailed by folk psychological isomorphism, but that's all. For example, the fact that pains cause us to

moan (in circumstances that we have some appreciation of but no one has ever precisely stated) is known to common sense, but the fact that just-noticeable differences in stimuli increase with increasing intensity of the stimuli (the Weber-Fechner Law) is not. So the merely superficial isomorph would be governed by the former but not necessarily the latter. The TV series *Star Trek: The Next Generation* (2/26/89) includes an episode ('The Measure of a Man') in which there is a trial in which it is decided whether a human-like android, Lt. Commander Data, may legally be turned off and taken apart by someone who does not know whether he can put the parts together again. (The technology which allowed the android to be built has been lost.)[18] Let us take Commander Data to be a merely superficial iso-

[18] Here is a brief synopsis by Timothy Lynch (tlynch@alumni.caltech.edu, quoted with permission), http://www.ugcs.caltech.edu/st-tng/episodes/135.html: "While at Starbase 173 for crew rotation, Picard runs into an old acquaintance, Captain Phillipa Louvois, who once prosecuted him in the Stargazer court-martial, but is now working for the JAG (Judge Advocate General) office in this sector. Also on hand is Commander Bruce Maddox, who once on board the Enterprise, announces his intention to dismantle Data. Maddox is an expert in cybernetics, and has worked for years to recreate the work of Dr. Soongh, and he believes examining Data will give him the boost he needs to create many more androids like Data. However, when Picard, wary of Maddox's vagueness [actually, Maddox appears to have no idea whether he can put Data back together], declines the offer, Maddox produces orders transferring Data to his command. After talking to Data, Picard goes to Phillipa to find a way to block the transfer. Unfortunately, the only option is for Data to resign from Starfleet. This he does, immediately, but is interrupted while packing by Dr. Maddox, who claims Data cannot resign. Data says that he must, to protect Soongh's dream. Maddox takes his complaint to Phillipa, and claims that since Data is property, he cannot resign. As she starts looking into this possibility, Data is thrown a going-away party and wished well in whatever he chooses to do. However, Phillipa then tells Picard and Riker that, according to the Acts of Cumberland, Data is the property of Starfleet, and thus cannot resign, or even refuse to cooperate with Maddox. Further, if a hearing is held to challenge this ruling, since Phillipa has no staff, serving officers must serve as counsel, with Picard defending and Riker prosecuting. Riker does some research and presents a devastating case for the prosecution, turning Data off while talking about cutting Pinocchio's strings. Picard, taken aback, asks for a recess, and talks to Guinan. Guinan subtly gets Picard to realize that if Data, and his eventual successors, are held to be 'disposable people,' that's no better than slavery all over again. Picard, renewed, presents his defense. He asks Data why he values such things as his medals, a gift from Picard, and especially a holographic image of Tasha (surprising Phillipa with Data's statement that they were 'intimate'). He

morph of us (ignoring his superior reasoning and inferior emotions). Then (1) can be taken to be that we have no reason to believe that Commander Data is not nomologically or otherwise metaphysically possible. Note that I am not making so strong a claim as made in Block and Fodor (op. cit.) — that there is empirical reason to suppose that our functional organization is multiply realizable — but only that we have no reason to doubt it.

The strategy of the argument, you recall, is to move from the epistemic possibility of multiple realization to the epistemic possibility of multiple constitution. (1) is the epistemic possibility of multiple realization.

(2) Superficial functional equivalence to us is a defeasible reason for attributing consciousness. That is, superficial functional equivalence to us provides a reason for thinking a being is conscious, but that reason can be disarmed or unmasked, its evidential value cancelled.

(2) consists of two claims, that superficial functional equivalence to us is a reason for attributing consciousness and that that reason is defeasible. The first claim is obvious enough. I am not claiming that the warrant is a priori, just that there is warrant. I doubt that there will be disagreement with such a minimal claim.

What is controversial about (2) is that the reason is claimed to be defeasible. Certainly, deflationary functionalists will deny the defeasibility. Of course, even deflationary functionalists would allow that *evidence* for thinking something is functionally equivalent to us can be defeated. For example, that something emits English sounds is a reason to attribute consciousness, but if we find the sound is recorded, the epistemic value of the evidence is cancelled. However, (2) does not merely say that functional or behavioral *evidence* for consciousness can be defeated. (2) says that even if we *know* that something is functionally equivalent to us, there are things we can find out that cancel the reason we have to ascribe consciousness (without challenging our knowledge of

---

calls Maddox as a hostile witness, and demands from him the requirements for sentience. Finally, Picard points out that the possibility of thousands of Datas is becoming a race, and claims that 'Starfleet was founded to seek out new life — well there it sits!!' Phillipa rules in favor of Data, who refuses to undergo Maddox's procedure. Maddox cancels the transfer order, and Data comforts Riker, saying he will not easily forget how Riker injured himself (by prosecuting) to save Data."

the functional equivalence). A creature's consciousness can be unmasked without unmasking its functional equivalence to us.

Here is a case in which the epistemic value of functional isomorphism is cancelled: The case involves a *partial physical* overlap between the functional isomorph and humans. Suppose that there are real neurophysiological differences of kind — not just complexity — between our conscious processes and our unconscious — that is, non-phenomenal — processes. Non-phenomenal neural process include, for example, those that regulate body temperature, blood pressure, heart rate and sugar in the blood — brain processes that can operate in people in irreversible vegetative coma. Suppose (*but only temporarily* — this assumption will be dispensed with later) that we find out that *all* of the merely superficial isomorph's brain states are ones that — in us — are the neural bases *only of phenomenally unconscious states*. For example, the neural basis of the functional analog of pain in the merely superficial isomorph is the neural state that regulates the pituitary gland in us. This would not *prove* that the isomorph is not phenomenally conscious (for example, since the contexts of the neural realizers are different), but it would cancel or at least weaken the force of the reason for attributing consciousness provided by its functional isomorphism to us.

The role of this case is to motivate a further refining of our characterization of Commander Data and to justify (2) by exhibiting the epistemic role of a defeater.

Let us narrow down Commander Data's physical specification to rule out the cases just mentioned as defeaters for attribution of consciousness to him. Here is a first shot:

- Commander Data is a superficial isomorph of us.
- Commander Data is a merely superficial isomorph. So we have no reason to suppose there are any shared non-heterogeneously-disjunctive physical properties between our conscious states and Commander Data's functional analogs of them that could be the physical basis of any phenomenal overlap between them, since we have no reason to think that such shared properties are required by the superficial overlap. Further, one could imagine this discussion taking place at a stage of science where we could have rational ground for believing that there are no shared physical properties (or more generally scientific properties) that could be the physical basis of a phenomenal overlap. Note that no stipulation can rule out certain shared physical properties, e.g. the disjunctive prop-

erty of having the physical realizer of the functional role of one of our conscious states or Commander Data's analog of it.

- The physical realizers of Commander Data's functional analogs of conscious states do not overlap with any of our brain mechanisms in any properties that we do not also share with inorganic entities that are uncontroversially mindless, e.g. toasters. So we can share properties with Commander Data like having molecules. But none of the realizers of Commander Data's analogs of conscious states are the same as realizers of, for example, our states that regulate our blood sugar — since these are organic.
- Commander Data does not have any part which itself is a functional isomorph of us and whose activities are crucial to maintaining the functional organization of the whole.[19]

The point of the last two conditions is to specify that Commander Data has a realization that cannot be seen to defeat the attribution of consciousness to him either a priori or on the basis of a theory of *human* consciousness. (For example, the last condition rules out a 'homunculi-headed' realization.) It would help if I could think of all the realizations that have these kinds of significance. If you tell me about one I haven't thought of, I'll add a condition to rule it out.

Objection: we are entitled to reason from same effects to same causes. Since our phenomenal states play a role in causing our behavior, we can infer that the functionally identical behavioral states of Commander Data are produced in the same way, that is, phenomenally. To refuse to accept this inference, the objection continues, is to suppose that the presence or absence of phenomenality makes no causal difference.

Reply: Consider two computationally identical computers, one that works via electronic mechanisms, the other that works via hydraulic mechanisms. (Suppose that the fluid in one does the same job that the electricity does in the other.) We are not entitled to infer from the causal efficacy of the fluid in the hydraulic machine that the electrical machine also has fluid. One could not conclude that the presence or absence of the fluid makes no difference, just because there is a functional equivalent that has no fluid. One need not be an epiphenomenalist to take seriously the hypothesis that there are alternative realizations of

[19] Following Putnam, op. cit. This stipulation needs further refinement, which it would be digressive to try to provide here.

the functional roles of our phenomenal states that are phenomenally blank.

We might suppose just to get an example on the table that the physical basis of Commander Data's brain is to be found in etched silicon chips rather than the organic carbon basis of our brains.[20]

The reader could be forgiven for wondering at this point whether I have not assembled stipulations that close off the question of Commander Data's consciousness. Naturalism includes the doctrine that it is the default that a conscious overlap requires a physical basis, and it may seem that I have in effect stipulated that Commander Data does not have any physical commonality with us that could be the basis of any shared phenomenality. The objection ignores the option of a shared *disjunctive* basis and certain other shared bases to be discussed below.

(3) Fundamentally different physical realization from us *per se* is not a ground of rational belief in lack of consciousness. So the fact that Commander Data's control mechanisms are fundamentally different is not a ground of rational belief that he has no phenomenal states. Note that I don't say that finding out that Commander Data has a silicon-based brain isn't a *reason* for regarding him as lacking consciousness. Rather I say that the reason falls below the epistemic level of a ground for rational belief.

(4) We have no conception of a ground of rational belief to the effect that a realization of our superficial functional organization that is physically fundamentally different along the lines I have specified for Commander Data is or is not conscious. To use a term suggested by Martine Nida-Rümelin in commenting on this paper, Commander Data's consciousness is meta-inaccessible. Not only do we lack a ground of belief, but we lack a conception of any ground of belief. This meta-inaccessibility is a premise rather than a lemma or a conclusion because the line of thought I've been presenting leads up to it without anything that I am happy to think of as an argument for it. My hope is that this way of leading up to it will allow the reader to see it as obvious.

---

[20] See Sydney Shoemaker, 'The Inverted Spectrum,' *Journal of Philosophy* 79, 7, 1982: 357-81. Shoemaker makes assumptions that would dictate that Commander Data overlaps with us in the most general phenomenal property, *having phenomenality* — in virtue of his functional likeness to us. But in virtue of his lack of physical overlap to us, there are no shared phenomenal states other than phenomenality itself. So on Shoemaker's view, phenomenality is a functional state, but more specific phenomenal states have a partly physical nature.

We can see the rationale for meta-inaccessibility by considering John Searle's Chinese Room argument. Searle famously argued that even if we are computational creatures, we are not either sentient or sapient merely in virtue of that computational organization. In reply to his critics[21], he says repeatedly that a machine that shares our computational organization and is therefore behaviorally and functionally equivalent to us — and therefore passes the Turing Test — need not be an intentional system (or a conscious being). What would make it an intentional system — and for Searle, intentionality is engendered by and requires consciousness — is not the functional organization but rather the way that functional organization is implemented in the biology of the organism. But, to take an example that Searle uses, how would we know whether something made out of beer cans is sentient or sapient? He says: 'It is *an empirical question* whether any given machine [that shares our superficial functional organization] has causal powers equivalent to the brain.' (p. 452) 'I think it is evident that all sorts of substances in the world, like water pipes and toilet paper, are going to lack those powers, but that is *an empirical claim* on my part. On my account it is a *testable empirical claim* whether in repairing a damaged brain,' we could duplicate these causal powers. (p. 453) 'I offer no a priori proof that a system of integrated circuit chips could not have intentionality. That is, as I say repeatedly, *an empirical question*. What I do argue is that in order to produce intentionality the system would have to duplicate the causal powers of the brain and that simply instantiating a formal program would not be sufficient for that' (p. 453; emphasis and bracketed clause added).

I do not deny that one day the question of whether a creature like Commander Data is phenomenally conscious may *become* a testable empirical question. But it is obvious that we do not *now* have any conception of how it could be tested. Searle has suggested (in conversation) that the question is an empirical one in that if I were the device, I would know from the first person point of view if I was conscious. But even if we accept such a counterfactual, we cannot take it as showing that the claim is testable or empirical in any ordinary sense of the term.

Though I am tweaking Searle's flamboyant way of putting the point, my naturalist phenomenal realist view is not that different from his. I agree that whether physically different realizations of human functional

---

[21] 'Author's Response,' *The Behavioral and Brain Sciences* 3/450-457, 1980.

organization are conscious is not an a priori matter and could be said to depend on whether their brains have 'equivalent causal powers' to ours — in the sense of having the power to be the physical basis of conscious states. (However, I don't agree with Searle's view that the neural bases of conscious states 'cause' the conscious states in any normal sense of 'cause.') I agree with him that consciousness is a matter of the biology of the organism, not (just) its information processing. The issue that I am raising here for naturalist phenomenal realism threatens my view as much as his.

I am not denying that we might some day come to have the conception we now do not have. (So I am not claiming — as McGinn does — that this knowledge can be known now to be beyond our ken.[22]) I am merely saying that at this point, we have no idea of evidence that would ground rational belief, even a hypothetical or speculative conception. Of course those who meet Commander Data will reasonably be sure that he is conscious. But finding out that he is not *human* cancels that ground of rational belief.

Perhaps we will discover the nature of human consciousness and find that it applies to other creatures. E.g. the nature of human consciousness may involve certain kinds of oscillatory processes that can apply to silicon creatures as well. But the problem I am raising will arise in connection with realizations of our functional organization that lack those oscillatory processes. The root of the epistemic problem is that the example of a conscious creature on which the science of consciousness is inevitably based is us (where 'us' can be construed to include non-human creatures which are neurologically similar to humans). But how can science based on us generalize to creatures that don't share our physical properties? It would seem that a form of physicalism that could embrace other creatures would have to be based on them at least in part in the first place, but that cannot be done unless we already know whether they are conscious.

I have left a number of aspects of the story unspecified. What was the aim of Commander Data's designer? What is to be included in the 'common sense' facts about the mind that determine the grain of the functional isomorphism?

I keep using the phrase 'ground of rational belief.' What does it mean? I take this to be an epistemic level that is stronger than 'reason

---

[22] Colin McGinn, *The Problem of Consciousness.* Oxford: Blackwell, 1991.

for believing' and weaker than 'rational certainty.' I take it that a ground of rational belief that p allows knowledge that p but mere reason for believing p does not.

VII. Disjunctivism and the epistemic problem

I now move to the conditional that I advertised earlier. Let us start by supposing, but only temporarily, that physicalism requires a deep (non-superficial) unitary (non-heterogeneously-disjunctive) scientific (physical) property shared by all and only conscious beings. This version of physicalism seems at first glance to be incompatible with Commander Data's being conscious, and the corresponding version of naturalism (which says that physicalism is the default) seems at first glance to be epistemically incompatible with phenomenal realism. That is, naturalism says the default is that Commander Data is not conscious but phenomenal realism says that the issue is open in the sense of no rational ground for belief either way. This is a first pass at saying what the Harder Problem is.

If this strong kind of physicalism really is incompatible with Commander Data's being conscious, we might wonder whether the reasons we have for believing physicalism will support this weight. I will pursue a weaker version of physicalism (and corresponding version of naturalism) that does not rule out consciousness having a physical basis that is disjunctive according to the standards of physics. However, as we will see, the stronger version of physicalism is *not* actually incompatible with Commander Data's being conscious, and the difference between the stronger and weaker versions makes no important difference with respect to our epistemic situation concerning Commander Data's consciousness.

Disjunctivism is a form of physicalism that allows that consciousness is a physical state that is disjunctive by the standards of physics. As applied to the current issue, Disjunctivism allows that if Commander Data is conscious, the shared phenomenality is constituted by the property of having Commander Data's electronic realization of our shared functional state or our electro-chemical realization.

In footnote 14, I mentioned Kim's critique of the multiple realizability argument against physicalism. He argues that if mental property M is nomically equivalent to a heterogeneous disjunction N, we should

regard M as non-nomic and non-'real' because N is. He argues that if human thought can be realized by very different physical mechanisms from, say, Martian or robot thought, then the real sciences of thought will be the sciences of the separate realizations of it. To call them all 'thought' is simply to apply a superficial verbal concept to all of them, but the laws of human thought will be different from the laws of Martian thought. The real kinds are not at the level of the application of verbal concepts.[23]

Even those who are sympathetic to this picture of thought must make an exception for consciousness (in the sense, as always in this paper, of phenomenality). We can be happy with the view that there is a science of human thought and another science of machine thought, but no science of thought per se. But we should not be happy with the idea that there is a science of human phenomenality, another of machine phenomenality, etc. For since the overlap of these phenomenalities, *phenomenality*, is something real and not merely nominal as in the case of thought, it must have a scientific basis. If a phenomenal property is nomically coextensive with a heterogeneous neural disjunction, it would not be at all obvious that we should conclude that the phenomenal property is non-nomic and non-'real' because the disjunction is. The phenomenal realist naturalist point of view would be more friendly to the opposite, that the disjunction is nomic and 'real' because the phenomenal property is.

The real problem with Disjunctivism is that whether it is true or not, we could have no good reason to believe it. To see this, we shall have to have a brief incursion into the epistemology of reductive theoretical identity.

The epistemology of theoretical identity

Why do we think that water = H_2O, temperature = mean molecular kinetic energy and freezing = lattice formation?[24] The answer begins

---

[23] See my replies to Kim, 'Anti-reductionism Slaps Back,' *Mind, Causation, World, Philosophical Perspectives* 11, 1997,107-133; and 'Do Causal Powers Drain Away,' *Philosophy and Phenomenological Research* LXVII, 1, July 2003, with a response by Kim.

[24] The temperature identity is oversimplified, applying in this form only to gases. Paul Churchland raises doubts about whether there is a more abstract

with the fact that water, temperature, freezing and other magnitudes form a family of causally inter-related 'macro' properties. This family corresponds to a family of 'micro' properties: H_2O, mean molecular kinetic energy, formation of a lattice of H_2O molecules. And the causal relations among the macro properties can be explained if we suppose the following relations between the families: that water = H_2O, temperature = mean molecular kinetic energy and freezing = lattice formation. For example, as water is cooled, it contracts until about 4 degrees (F) above freezing, at which point it expands. Why? Why does ice float on water? Here is a sketch of the explanations: The oxygen atom in the H_2O molecule has two pairs of unmated electrons, which attract the hydrogen atoms on other H_2O molecules. Temperature = mean molecular kinetic energy. When the temperature (viz., kinetic energy) is high, the kinetic energy of the molecules is high enough to break these hydrogen bonds, but as the kinetic energy of the molecules decreases, each oxygen atom tends to attract two hydrogen atoms on the ends of two other H_2O molecules. When this process is complete, the result is a lattice in which each oxygen atom is attached to four hydrogen atoms. Ice is this lattice and freezing is the formation of such a lattice. Because of the geometry of the bonds, the lattice has an open, less dense structure than amorphously structured H_2O (viz., liquid water) — which is why ice (solid water) floats on liquid water. The lattice forms slowly, beginning about 4 degrees above freezing. (The exact temperature can be calculated on the basis of the numerical values of the kinetic energies needed to break or prevent the bonds.) The formation of large open lattice elements is what accounts for the expansion of water on the way to freezing. (Water contracts in the earlier cooling because decreasing kinetic energy allows more bonding, and until the bonding reaches a stage in which there are full lattice elements, the effect of the increased bonding is make the water more densely packed.)

Suppose we reject the assumption that temperature is identical to mean molecular kinetic energy in favor of the assumption that temperature is merely correlated with mean molecular kinetic energy? And

---

identity in *Matter and Consciousness,* MIT Press: Cambridge, 1984. I think those doubts are deflected in Simon Blackburn's 'Losing Your Mind: Physics, Identity and Folk Burglar Prevention,' Chapter 13 of *Essays in Quasi-Realism*, Blackwell: Oxford, 1993.

suppose we reject the claim that freezing is lattice-formation in favor of a correlation thesis. And likewise for water/H_2O. Then we would have an explanation for how something that is *correlated* with decreasing temperature causes something that is *correlated* with frozen water to float on something *correlated* with liquid water, which is not all that we want. Further, if we assume identities, we can explain why certain macro properties are spatio-temporally coincident with certain micro-properties. The reason to think that the identities are true is that assuming them gives us explanations that we would not otherwise have and does not deprive us of explanations that we already have or raise explanatory puzzles that would not otherwise arise. The idea is not that our reason for thinking these identities are true is that it would be nice if they were true. Rather, it is that assuming that they are true yields the *most explanatory overall picture*. In other words, the epistemology of theoretical identity is just a special case of inference to the best explanation.

Some suppose that substance identities such as 'water = H_2O' are on a different footing from 'property' identities, and that substance identities can be established on purely spatiotemporal grounds. (Jaegwon Kim gave a paper at Columbia in December, 1999 making this suggestion, and Tim Maudlin argued that all theoretical identities are established on spatiotemporal grounds when I gave this paper at Rutgers.) But deciding that water and H_2O *are spatio-temporally coincident* is part of the same package as having decided that they are one and the same. For example, the air above a glass of water buzzes with bits of water in constant exchange with water in the atmosphere, a fact that we can acknowledge only if we are willing to suppose that those H_2O molecules *are* bits of water. The claim that water is H_2O and that water and H_2O are spatio-temporally coincident stand or fall together as parts of one explanatory package. And once we conclude that the substance liquid water = amorphous H_2O and that the substance frozen water = lattice-structured H_2O, we would be hard pressed to deny that freezing = lattice formation, since the difference between liquid and frozen water is that the former has an amorphous structure and the latter a lattice structure. Substance identities and property identities often form a single explanatory package.

Back to disjunctivism

With the epistemology of identity in place, we can now ask whether there could be an argument from inference to the best explanation to the conclusion that consciousness is a heterogeneous physical disjunction, the disjunction of our realization of the consciousness role and Commander Data's corresponding realization. Of course without a prior decision as to whether Commander Data's states are actually conscious, there could be no such argument. Putting this point aside, let us suppose, temporarily, that Commander Data is conscious. Even so, the prospects for an argument from inference to the best explanation to the identity of a phenomenal property with a disjunctive physical property are dubious. We can see this in two ways. First, let us attend to our explanatory practice. We have an important though vague notion of 'fundamentally different' that governs our willingness to regard some differences in realization as variants of the same basic type and others as fundamentally different. When we regard two realizations as fundamentally different, we prefer two non-disjunctive identities to one disjunctive identity. Here is an example: Molten glass hardens into an amorphous solid-like substance. (If there are absolutely no impurities, fast continuous cooling of water can make it harden without lattice formation in a similar manner.) We could give a disjunctive explanation of solid-like formation that included both freezing and this kind of continuous hardening. And if we preferred that disjunctive explanation to two distinct explanations, we would regard the hardening of glass as a kind of freezing and glass as a solid. But we do not take the disjunctive explanation seriously and so we regard glass as (strictly speaking) a supercooled liquid rather than a solid. And we do not regard amorphous hardening as freezing. We prefer two non-disjunctive identities, freezing = lattice-formation and hardening = formation of an amorphous super-cooled liquid to one disjunctive identity. Of course, the two processes (freezing and hardening) are functionally different in all sorts of fine-grained ways. But the functional roles of Commander Data's functional analogs of our conscious states are also functionally different from ours in all sorts of fine-grained ways. Commander Data is functionally equivalent to us in those functional roles known to common sense and anything else nomologically or logically required by that equivalence, but everything else can be presumed to be different. Since

we can stipulate that our physical realizations of our conscious states are fundamentally different from Data's, whatever exactly fundamental difference turns out to be, the methodology that applies to the hardening/freezing case can reasonably be applied to the case at hand.

Of course, there are cases in which we accept disjunctive identities, e.g. jade is nephrite or jadeite. But *jade* is a merely nominal category, which makes disjunctive identities acceptable even if not explanatory.

A second factor is that the disjunctive identity, if accepted, would rule out questions that the phenomenal realist naturalist does not want to rule out. The question of why it is that water is correlated with H_2O or why it is that heat is correlated with molecular kinetic energy are bad questions, and they are ruled out by the identity claims that water = H_2O and heat = molecular kinetic energy. Nor can the identities themselves be questioned. (See footnote 8.) If we were to accept that consciousness is a disjunction of the physical basis of our conscious states and Commander Data's realization of the functionally equivalent states, we would be committing ourselves to the idea that there is no answer to the question of why we overlap phenomenally with Data in one respect rather than in another respect or no respect at all. For the phenomenal realist, it is hard to imagine a ground for rational belief that these questions have no answers. One can imagine finding no other account remotely plausible, but why should the phenomenal realist accept a physicalist view that dictates that these questions are illegitimate rather than opt for a non-physicalist view that holds out some hope for an answer. (Remember that physicalism is only a default view.) Even if we should come to believe that dualism is unacceptable as well, our reason for accepting Disjunctive physicalism would not seem to get up to the level of a ground for rational belief.

Objection: You say identities cannot be explained, but then you also say that we can have no reason to accept a disjunctive physicalistic identity because it is not explanatory.

Reply: Identities cannot be explained, but they can contribute to explanations of other things. My point about the epistemology of identity is that it is only because of the explanatory power of identities that we accept them and the disjunctive identity countenanced by Disjunctivism does not pass muster.

Disjunctivism is one way of making naturalism compatible with Commander Data being conscious, but there are others. One is the

view that consciousness is as a matter of empirical fact identical to the superficial functional organization that we share with Commander Data. We might call this view Superficialism (with apologies to Georges Rey who has used this term for a somewhat different doctrine). Recall that the phenomenal realist/deflationist distinction is an epistemic one, so any ontological view could in principle be held as having either epistemic status. Superficialism is the *phenomenal realist* claim that consciousness is identical to the superficial functional organization that we share with Commander Data — as distinct from the deflationist version of this claim mentioned earlier.

Note that Superficialism says consciousness is a role property, not a property that *fills* or realizes that role. A role property is a kind of dispositional property. Now there is no problem about dispositions being caused: spraying my bicycle lock with liquid nitrogen causes it to become fragile. So if pain is a superficial functional state, we can perhaps make use of that identification to explain the occurrence of pain in neural terms. Whether dispositions are causes — as would be required by this identity — is a more difficult issue that I shall bypass. (Does a disposition to say ouch cause one to say ouch?)

The difficulty I want to raise is that even if identifying pain with a superficial functional role does license explanations of the *superficial* causes and effects of being in pain, the identification cannot in the same way license explanations of the *non-superficial* causes and effects of being in pain. Suppose, for example, that psychologists discover that pain raises the perceived pitch of sounds. Even if we take the thesis that pain is a disposition to say ouch to help us to explain why pain causes saying ouch, it will not explain the change in pitch. The epistemic difficulty I am pointing to is that there is no good reason why the causal relations *known to common sense* ought to be explained differently from the ones not known to common sense. So the identification raises an explanatory puzzle that would not otherwise arise, and that puts an epistemic roadblock in the way of the identification. This is perhaps not a conclusive difficulty with the proposal, but it does put the burden of proof on the advocate of the identification to come up with explanatory advantages so weighty as to rule out the explanatory disadvantage just mentioned.[25]

[25] I am grateful to David Chalmers for pressing me for a better treatment of this issue.

Of course, this objection will not apply to the phenomenal realist identification of consciousness with its *total* functional role as opposed to its superficial functional role. Since the physiology of Commander Data's states differs from ours, their total functional roles will differ as well. So this would be a chauvinist proposal that would beg the question against Commander Data's consciousness.

Martine Nida-Rümelin objected that there are a vast number of properties, maybe infinitely many, that are entailed nomologically or logically by the superficial functional equivalence, and each of these is both shared with Data and is a candidate for the nature of consciousness. Certainly a full treatment would attempt to categorize these properties and assess their candidacy. Some — e.g., possessing complex inputs and outputs — can be eliminated because they are also shared with mindless computers. Of course, there may be others that are not so easily dismissed.

The upshot

I said earlier that it seemed at first glance that a form of physicalism that required that consciousness be constituted by a unitary physical property dictated that Commander Data is not conscious. We can now see that at second glance, this is not the case. Even if we preclude a disjunctive physical basis to the phenomenal overlap between us and Commander Data (assuming that there is such an overlap), still the physicalist could allow that Commander Data is conscious *on Superficialist grounds*. And even if we reject Superficialism, there are other potential meta-inaccessible physical bases of a phenomenal overlap between us and Commander Data.

The upshot is that physicalism in neither the stronger (unitary physical basis) nor weaker (physical basis that may or may not be unitary) versions mentioned above rules out Commander Data's being conscious. However, *the only epistemically viable naturalist or physicalist hypothesis* — the only naturalist or physicalist hypothesis we have a conception of a reason for accepting — is a deep unitary physical or otherwise scientific property in common to all and only conscious beings, a naturalistic basis that Commander Data does not share. So for the physicalist, Commander Data's consciousness is not *epistemically* viable.

Thus our knowledge of physicalism is *doubly* problematic: we have no conception of a ground of rational belief that Commander Data is or is not conscious, and we have no way of moving from a conclusion that Commander Data is conscious to any consequence for the truth of physicalism. And this holds despite the fact that physicalism is our default view. *Physicalism is the default and also inaccessible and meta-inaccessible*. The *practical* significance — if we ever make a robot that is functionally equivalent to us — is that the question of its consciousness and also of physicalism are inaccessible and meta-inaccessible. But even if we decide that the robot is conscious, we will have a choice between dualism and an epistemically non-viable version of physicalism (Disjunctivism or Superficialism). This is all part of the Harder Problem. A second part follows.

But first I will discuss the question of whether the epistemic tension itself is a good reason to conclude that Commander Data is not conscious. The short version of my answer is that while the epistemic tension is a bad consequence of our phenomenal realist view that it is an open question whether Commander Data is conscious, it is not the kind of bad consequence that justifies us in concluding that he is not conscious. I will justify this claim.

Objection: You say disjunctivism is epistemically defective, but isn't it also metaphysically defective? How could a unitary phenomenal property be identical to a physical property that is non-unitary?

Reply: There is no logical flaw in disjunctivism. If a unitary phenomenal property is identical to a non-unitary physical property, then one property is both unitary from the mental point of view and non-unitary from the physical point of view. We are willing to allow that unitary properties of economics, sociology and meteorology are non-unitary from the physical point of view. Why shouldn't we include mentality too?[26]

Of course, there are views that are worthy of being called 'naturalism' that dictate that disjunctivism is metaphysically defective. But they are not the 'naturalism' that I am talking about. The naturalist I am talking about, you will recall, is also a phenomenal realist. And being a phenomenal realist, this naturalist keeps the question open of whether creatures that are heterogeneous from a physical point of view nonethe-

[26] See my 'Anti-reductionism Slaps Back,' op. cit. for more on this topic.

less overlap phenomemenally. If you like, this is a naturalistic concession to phenomenal realism.

Objection: Silicon machinery of the sort we are familiar with is manifestly not conscious. The only reason we could have to suppose that Commander Data's brain supported consciousness would be to find some kind of physical similarity to the states that we know underlie human consciousness, and that possibility has been ruled out by stipulation. Moreover, we can explain away our tendency to think of Commander Data as conscious as natural but unjustified anthropomorphizing.

Reply: Naturalism and Phenomenal Realism do not dictate that Commander Data is not conscious or that the issue of his consciousness is not open. Recall that Disjunctivism and Superficialism are metaphysically (though not epistemically) viable. Further, naturalism gives us no evidence against or reason to doubt the truth of either Disjunctivism or Superficialism. Hence naturalism (and physicalism) give us no reason to doubt the consciousness of Commander Data. Imagine arguing at Commander Data's trial that he is a zombie (or that there is no matter of fact as to whether he is conscious) while conceding that his zombiehood is not even *probabilified* by naturalism unless we set aside Disjunctivism and Superficialism, options on which he may be conscious. And imagine conceding that we are setting these options aside not because we have any evidence against them or reason to think they are false but because we cannot conceive of any way in which they may be known. He could reasonably say (or to be neutral, produce the noise), 'Your lack of a conception of how to find out whether I am conscious is no argument that I am a zombie; I similarly lack a conception of how to find out whether you are conscious.' In any case, phenomenal realism is a form of metaphysical realism, so the phenomenal realist cannot suppose that our ignorance, even necessary ignorance, is not a reason to suppose that Commander Data is not conscious or that there is no matter of fact as to whether he is.

Why should the phenomenal realist take the consciousness of anything other than humans seriously? One answer can be seen by considering what happens if one asks Commander Data whether red is closer to purple than blue is to yellow. Answering such questions requires, in us, a complex multi-dimensional phenomenal space — in part captured by the color solid — with phenomenal properties at many levels of abstractness (cf. Loar, op. cit.). Commander Data's functional equiva-

lence to us guarantees that he has an internal space that is functionally equivalent to our phenomenal space. But anyone who grasps our phenomenal space from the first person point of view has to take seriously the possibility that an isomorphic space in another being is grasped by him from a similar first person perspective. Talking of our 'functional equivalence' to Commander Data tends to mask the fact that we are like him in a complex structure or set of structures. If one thinks of the functional similarity as limited to saying 'Ouch' when you stick a pin in him, it is easy to miss the positive phenomenal realist rationale for regarding Commander Data's consciousness as an open question. Thus the phenomenal realist and the deflationist converge on not closing off the possibility that Commander Data is conscious.

To make the plausibility of Commander Data's consciousness vivid, I include in Figure 1 below stills from Commander Data's trial.

Figure 1: Stills from 'The Measure of a Man,' episode 35 of *Star Trek: The Next Generation*. Commander Data is on the left in the first, in which his hand is removed by his prosecutor after he has turned Data off, emphasizing his robotic nature, and in the middle on the right, in the dock.

Objection (made by many critics): Why should the mere epistemic possibility of a bad consequence of physicalism threaten physicalism? No one thinks that the mere epistemic possibility of an object that has mass traveling faster than light threatens relativity theory. If relativity is true, nothing can travel faster than light. Similarly, if physicalism is true, there is no conscious Commander Data.

Reply: Relativity theory gives us reason to believe that it is impossible for anything to travel faster than light. But physicalism does not

give us reason to believe that there can be no Commander Data or that it is impossible that Commander Data is conscious. Disjunctivism is not metaphysically suspect but only epistemically suspect: we have no conception of how we can know whether it is true or not. Our lack of knowledge is no argument against the consciousness of Commander Data.

Brian McLaughlin has argued (in a response at SOFIA, 2001) that I am mischaracterizing the epistemic role of functional similarity in our reasoning about other minds. The role of functional similarity is in providing evidence that others are like us in intrinsic physical respects, and that is the ground for our belief in other minds. In the case of Commander Data, that evidential force is cancelled when we find out what Commander Data's real constitution is. He notes that we are happy to ascribe consciousness to babies even though they are functionally very different from us because we have independent evidence that they share the relevant intrinsic physical properties with us. The same applies, though less forcefully, to other mammals, e.g. rabbits. He asks us to compare a human baby with a functionally equivalent robot baby. The robot baby's functional equivalence to the real baby gives us little reason to believe that the robot baby is conscious. Similarly, for the comparison between a real rabbit and a robot rabbit. Moving closer to home, consider a paralytic with Alzheimer's: little functional similarity to us, but we are nonetheless confident, on the basis of an inference from similarity in intrinsic physical properties, that the senile paralytic has sensory consciousness. The upshot, he says, is that material constitution and structure trumps function in our attribution of consciousness to others. And so, if we become convinced that Commander Data is unlike us in the relevant intrinsic physical respects, we should conclude that he is not conscious.

Reply: first, Commander Data shares with us *Disjunctivist* and *Superficialist material constitution and structure*, and so no conclusion can be drawn about the consciousness of Commander Data, even if McLaughlin is right about material constitution and structure trumping function. Nothing in McLaughlin's argument supplies a reason to believe that Disjunctivism or Superficialism are false. (Recall that I have argued that these views are *epistemically* defective, not that they are false.) He says that the relevant physical properties are 'intrinsic' but if that is supposed to preclude Disjunctivism or Superficialism, we are owed an

argument. Second, I do agree with McLaughlin that a substantial element of our belief in other consciousnesses depends on an inference to a common material basis. However, it would be a mistake to conclude that this inference provides the entire basis for our attribution of other consciousnesses. Our justification is an inference from like effects to like causes. Even if we find out that the causes of behavioral similarity are not alike in material constitution and structure, it remains open that the common cause is a similarity in *consciousness itself* and that consciousness itself has a disjunctive or superficial material basis or no material basis. (Recall that naturalism is committed to physicalism as a default, but a default can be overridden.)

Third, function is not so easy to disentangle from material constitution and structure, at least epistemically speaking. The opponent process theory of color vision originated in the 19th Century from common sense observations of color vision such as the fact that afterimages are of the complementary color to the stimulus and that there are colors that seem, e.g. both red and blue (purple) or red and yellow (orange) but no color that seems both red and green or both blue and yellow. The basic two stage picture of how color vision works (stage 1: three receptor types; stage 2: two opponent channels) was discovered before the relevant physiology on the basis of behavioral data. To the extent that Commander Data behaves as we do, there is a rationale for supposing that the machinery of Commander Data's color vision shares an abstract structure with ours that goes beyond the color solid.

The first of the epistemic difficulties on the right hand side of our conditional is that *physicalism is the default, but also inaccessible and meta-inaccessible.* We are now ready to state the second epistemic difficulty. Let us introduce a notion of the 'subjective default' view which we have rational ground for believing on the basis of background information — but only ignoring escape hatches — such as Disjunctivism and Superficialism — which we have no evidence against but which are themselves inaccessible and meta-inaccessible. Then the second epistemic difficulty is that of holding *both that it is an open question whether Commander Data is conscious and that it is the subjective default view that he is not.* These two epistemic difficulties constitute the Harder Problem.

Before I go on to consider further objections, let me briefly contrast the point of this paper with Nagel's famous 'bat' paper (op. cit.). Nagel's emphasis was on the functional differences between us and bats,

creatures which share the mammalian physical basis of sensation. My example, however, is one of a functionally identical creature, the focus being on the upshot of physical differences between us and that creature.

The issue of the application of our phenomenal concepts to exotic creatures is often mentioned in the literature, but assimilated to the Hard Problem (the 'explanatory gap'). (I am guilty too. That was the background assumption of the discussion of 'universal psychology' in my 'Troubles with Functionalism,' op. cit.) For example, Levine (*Purple Haze*, op. cit.) notes that we lack a principled basis for attributing consciousness to creatures which are physically very different from us. He says 'I submit that we lack a principled basis precisely because we do not have an explanation for the presence of conscious experience even in ourselves' (p. 79). Later he says 'Consider again the problem of attributing qualia to other creatures, those that do not share our physical organization. I take it that there is a very real puzzle whether such creatures have qualia like ours or even any at all. How much of our physicofunctional architecture must be shared before we have similarity or identity of experience? This problem, I argued above, is a direct manifestation of the explanatory gap' (p.89).

It might be objected that naturalism says the concept of consciousness is a natural kind concept and phenomenal realism denies it, so the tension is not epistemic, but is simply a matter of contradictory claims. But this is oversimple. Naturalism entails that the concept of consciousness is a natural kind concept in one sense of the term, since one sense of the term is just that it is the default that there is a scientific nature. Phenomenal realism does not deny this. Phenomenal realism denies something importantly different, which could be put in terms of Putnam's famous 'twin earth' example. We find that twin-water has a fundamentally different material basis from water, and that shows twin-water is not water. But if we find that Martian phenomenality has a fundamentally different material basis from human phenomenality, that does not show Martian phenomenality is not phenomenality. According to phenomenal realism, if it feels like phenomenality, it is phenomenality, whatever its material basis or lack of it.

Those who apply the scientific world view to consciousness often appeal to analogies between consciousness and kinds that have been successfully reduced. As noted earlier in connection with the Hard

Problem, there is some mileage in analogies to the identity of water with H_2O, heat with molecular kinetic energy and so on. But the fact that consciousness is not straightforwardly a natural kind concept puts a crimp in these analogies.

VIII. More objections

One can divide objections into those that require clarification of the thesis and those that challenge the thesis as clarified. The objections considered so far are more in the former category while those below are more in the latter.

Objections from indeterminacy

Objection: The issue of whether Commander Data is conscious just a matter of vagueness or indeterminacy in the word 'conscious.' If we reject property dualism, then the issue of whether Commander Data is conscious depends on extrapolating a concept of consciousness grounded in our physical constitution to other physical constitutions. If those other physical constitutions are sufficiently different from ours as is stipulated for Commander Data, then the matter is indeterminate and so a decision has to be made. Similarly, in extending the concept 'wood' to an alien form of life, we might find that it resembles what we have already called 'wood' in certain ways but not others and a decision will have to be made. (Hartry Field and David Papineau have pressed such views in commenting on an earlier version of this paper.)

Reply: No phenomenal realist — physicalist or not — should accept the assumption that the decision whether to attribute consciousness to Commander Data is a decision about whether to extrapolate from our *non-disjunctive and non-superficial* physical constitution to his. For as I have emphasized, the physical basis of our conscious states may be of the sort supposed by Disjunctivism or Superficialism, in which case there will be a matter of fact about Commander Data's consciousness — from a physicalist point of view.

I don't want to give the impression that phenomenal realism is incompatible with indeterminacy about consciousness. For example, perhaps a fish is a borderline case of consciousness. Similarly, Commander Data might be a borderline case of consciousness and therefore

indeterminate. On the phenomenal realist view of consciousness, it is an open question whether Commander Data is (a) conscious, (b) not conscious, (c) a borderline case. But there is no reason to think that Commander Data *must* be a borderline case. From the phenomenal realist point of view, epistemic considerations *alone* do not show metaphysical indeterminacy.

There is another kind of indeterminacy, exemplified by a familiar example of the Eskimo word for the whale oil that they use in daily life. Does their category include a petroleum product that looks and functions similarly, but is fundamentally different at a chemical level? There may be no determinate answer. If the Eskimo term is a natural kind term, the chemical specification is important; if the Eskimo term is not a natural kind term, perhaps the chemical specification loses out to function. But, as Gareth Evans once commented (in conversation), it may be indeterminate whether the Eskimo term is a natural kind term or not. So there may be no determinate answer to the question of whether the Eskimos should say that the petroleum product is 'oil.' David Lewis takes a similar stance towards consciousness. He supposes that in ascribing consciousness to an alien, we rely on a set of criteria that determine the population of the alien. If the alien has no determinate population, it is indeterminate in consciousness.[27]

The indeterminacy in the application of the Eskimo word can be resolved in the petroleum company's favor by introducing a coined expression (as Evans noted). For example, if there is an issue as to whether 'oil' is determinately a natural kind term, we can get rid of any indeterminacy of this sort by introducing 'oily stuff,' stipulating that anything that has the appearance and utility of oil is oily stuff (Chalmers, op. cit.; Block and Stalnaker, op. cit.). But in the case of consciousness, no such stipulation will help. Suppose I coin 'consciousish,' stipulating that comparisons do not depend on any hidden

---

[27] David Lewis, 'Mad Pain and Martian Pain' in N.Block (ed.) *Readings in Philosophy of Psychology* Vol. 1, Harvard University Press: Cambridge, 1980. Actually, Lewis' view is even weirder than I mention in the text. On Lewis' view, a creature which is both physically (and therefore functionally) just like us and which is now undergoing a state physically and functionally like one of our pains does not have pain if it is determinately a member of an appropriately different population. See Sydney Shoemaker's convincing refutation of Lewis, 'Some Varieties of Functionalism,' *Philosophical Topics* 12, 1 (1981), 357-381.

scientific essence. 'Consciousish' is not a natural kind term in the relevant sense. We may now ask: 'How could we get scientific evidence of whether or not Commander Data's current sensation is the same as my current sensation in respect of consciousishness?' The stipulation does not help. Alternatively, we could decide that 'consciousish' is a natural kind term, so Data is not consciousish. But the original question would recur as: 'Does Commander Data's state of consciousishness feel the same as ours?' I do not see how any coined term that was adequate to the phenomenon — from a phenomenal realist point of view — would fare any differently.

Another type of indeterminacy is exemplified in the question whether H_2O made out of heavy hydrogen (that is, D_2O) is a kind of *water* or not? There is no determinate answer, for our practice does not determine every decision about how the boundaries of a natural kind should be drawn. To decide the question of whether D_2O is a kind of water, we could either decide that water is a wide natural kind in which case the answer is yes or we could decide that water is a narrow natural kind in which case the answer is no. The issue would be settled. Suppose we try this technique to settle the issue of whether Commander Data is conscious. We could decide to construe 'consciousness' widely in case he is; or we could decide to construe 'consciousness' narrowly, in which case… What? Even if we decide to construe 'consciousness' narrowly, we can still wonder if the phenomenon picked out by it *feels the same* as what Commander Data has when he is in a functionally identical state! One can stipulate that 'Tuesdaysconsciousness' designates consciousness that occurs on Tuesday, but it still is in order to ask whether Tuesdayconsciousness feels the same as, say Thursdayconsciousness. Stipulations need not stick when it comes to the phenomenal realist conception of consciousness; any *adequate* concept of consciousness or phenomenality generates the same issue.

Closure of epistemic properties

In a response to this paper (SOFIA, 2001), Martine Nida-Rümelin gave a formalization of the argument that involved a principle of closure of epistemic properties such as being open or being meta-inaccessible. (Brendan Neufeld made a similar point.) E.g. she supposes that part of the argument goes something like this: supposing physicalism requires a

deep unitary property in common to conscious creatures, if Data is conscious, then physicalism is false; Data's consciousness is meta-inaccessible; so the falsity of physicalism is meta-inaccessible.

One can easily see that the form of argument is fallacious. If Plum did it, then it is false that the butler did it. But if it is inaccessible whether Plum did it, it does not follow that it is inaccessible whether or not the butler did it. We might find evidence against the butler that has nothing to do with Plum. The application of the point to the argument that Nida-Rümelin attributes to me is that even if Data's consciousness is inaccessible, we might have some independent reason to believe physicalism is false. I explicitly noted (and did in the earlier version) that I think the standard arguments against physicalism don't work.

Here is a standard problem with closure. (See my discussion of the tacking paradox in 'Anti-Reductionism Slaps Back,' op. cit.) Consider a meta-inaccessible claim, I, and an accessible claim, A. The conjunction I & A is meta-inaccessible, but a consequence of it, A, is not. So meta-inaccessibility is not transmitted over entailment. Briefly and metaphorically: fallacies of the sort mentioned seem to arise with respect to an epistemic property that applies to a whole even if only one of its parts has that property. The whole can then entail a different part that does not have that epistemic property. I doubt that my argument has that form, but if someone can show that it does, that will undermine it.

Objections concerning empirical evidence

Suppose my brain is hooked up to Commander Data's and I have the experience of seeing through his eyes. Isn't that evidence that he has phenomenal consciousness? Reply: maybe it is evidence, but it does not get up to the level of a rational ground for believing. Perhaps if I share a brain in that way with a zombie, I can see through the zombie's eyes because whatever is missing in the zombie brain is made up for by mine.

Suppose we discover what we take to be laws of consciousness in humans and discover that they apply to Commander Data. That is, we find that the laws that govern human consciousness also govern the functional analog of consciousness in Commander Data. Doesn't that get up to the level of rational ground for believing that Commander Data is

conscious? (I am grateful to Barry Smith for getting me to take this objection more seriously.)

Reply: Since Commander Data's brain works via different principles from ours, it is *guaranteed that his states will not be governed by all of the same laws* as the functionally equivalent states in us. Two computers that are computationally equivalent but physically different are inevitably different in all sorts of physical features of their operation, for example, how long they take to compute various functions, and their failure characteristics — such as how they react to humidity or magnetic fields. The most that can be claimed is that the state that is the functional analog of human consciousness in Commander Data obeys *some* of the laws that our conscious states obey. The problem is: are the laws that Commander Data does *not* share with us laws of consciousness or laws of his *different physical realizer*? Without an understanding of the scientific nature of consciousness, how are we supposed to know? A zombie might share some laws of consciousness, but not enough or not the right ones for consciousness. So long as Commander Data does not share *all* the laws of our conscious states, there will be room for rational doubt as to whether the laws that he does share with us are decisive. Indeed, if we knew whether Commander Data was conscious or not, we could use that fact to help us in deciding which laws were laws of consciousness and which were laws of the realization. But as this point suggests, the issue of whether Commander Data is conscious is of a piece with the epistemic problem of whether a given law is a law of consciousness or a law of one of the realizers of its functional role.

An example will be useful to clarify this point. All human sensory systems obey a power function, an exponential function relating stimulus intensity to subjective intensity as judged by subjects' reports. That is, subjective intensity = stimulus intensity raised to a certain exponent, a different exponent for different modalities. For example, perceived brightness is proportional to energy output in the visible spectrum raised to a certain exponent. This applies even to outré parameters of subjective judgments such as how full the mouth feels as a function of volume of wadges of paper stuck in the mouth or labor pains as a function of size of contractions. Should we see the question of whether Commander Data's sensations follow the power law as a litmus test for whether Commander Data has conscious experiences? No doubt the power law taps some neural feature. Is that neural feature

essential or accidental to the nature of consciousness? Roger Shepard has argued in his unpublished William James Lectures that the power law form would be expected in any naturally evolved creature. But that leaves open the possibility of artificial creatures or evolutionary singularities (subject to unusual selection pressures) whose sensations (or 'sensations') do not obey the power law. The question whether this is a law of consciousness or a law of the human realization of consciousness that needn't be shared by a conscious Commander Data is of a piece with the question of whether creatures like Commander Data (who, let us suppose, do not obey the law) are conscious. We cannot settle one without the other, and the epistemic problem I am raising applies equally to both.

Skepticism and the problem of other minds

Recall that I am arguing for a conditional. On the left are naturalism, phenomenal realism and the denial of skepticism. There is a superficial resemblance between the Harder Problem and the problem of other minds. But the problem of other minds is a form of skepticism. The non-skeptic has no doubt that *humans* are (sometimes) conscious, but when we find out that Commander Data is *not human,* denying skepticism does not help.

What is it about being human that justifies rejecting skepticism? It is not part of my project here to attempt an answer, but I have to say *something* to avoid the suspicion that our rationale for regarding other humans as conscious or rocks as not conscious might apply equally to Commander Data.

Elliot Sober's 'Evolution and the Problem of Other Minds'[28] argues plausibly that our rationale for attributing mental states to other humans is a type of 'common cause' reasoning. But such common cause reasoning is vulnerable to evidence against a common cause, e.g. evidence for lack of genealogical relatedness or evidence for different scientific bases for the similarity of behavior that is exhibited. Thus the rationale for attributing mentality to humans does not fully apply to Commander Data.

---

[28] *Journal of Philosophy,* XCVII, 7, July 2000, pp. 365-386.

Stephen White raises the skeptical worry of how we know that creatures whose brains are like ours in terms of principles of operation but not in DNA are conscious.[29] But this worry may have a *scientific* answer that would be satisfying to the non-skeptic. We might arrive at a partial understanding of the mechanisms of human consciousness that is sufficient to assure us that a creature that shared those mechanisms with us is just as conscious as we are even if its DNA is different. For example, we might discover a way to genetically engineer a virus that replaced the DNA in the cells of living creatures. And we might find that when we do this for adult humans such as ourselves, there are no noticeable effects on our consciousness. Or we might come to have something of a grip on why cortico-thalamic oscillation of a certain sort is the neural basis of human consciousness and also satisfy ourselves that many changes in DNA in adults do not change cortico-thalamic oscillation. By contrast, the Harder Problem may remain even if we accept the dictates of non-skeptical science.

IX. Supervenience and mind-body identity

Much of the recent discussion of physicalism in the philosophy of mind has centered on supervenience of consciousness on the brain rather than on good old-fashioned mind-body identity. Chalmers (op. cit., p. xvii) recommends this orientation, saying 'I find that discussions framed in terms of identity generally throw more confusion than light onto the key issues, and often allow the central difficulties to be evaded. By contrast, supervenience seems to provide an ideal framework within which key issues can be addressed.'

But the Harder Problem depends on the puzzling nature of multiple physical constitution of consciousness, a problem that does not naturally arise from the perspective that Chalmers recommends. Supervenience prohibits any mental difference without a physical difference, but multiple constitution is a physical difference without a mental difference. Of course nothing prevents us from stating the issue in supervenience

---

[29] Stephen White, 'Curse of the Qualia,' *Synthese* 68, 1983: 333-368. Reprinted in Block, Flanagan & Güzeldere, op. cit. The DNA issue is also mentioned in the version of Shoemaker's 'The Inverted Spectrum' in Block, Flanagan & Güzeldere, op. cit., pp. 653- 654.

terms. In those terms, it is the problem of how a unitary phenomenal property can have a non-unitary (heterogeneously disjunctive) supervenience base. But there is no reason why this should be puzzling from the supervenience point of view. Heterogeneous supervenience bases of unitary properties — e.g. adding — are common. What makes it puzzling is the thought that a phenomenal overlap between physically different creatures ought to have a unitary physical basis. That puzzle can be appreciated from the point of view of old-fashioned mind-body identity — which says that a phenomenal overlap is a physical overlap. (No one would identify adding with a physical (e.g. microphysical) property — it is obviously functional.) But it isn't puzzling from the supervenience point of view.

X. The hard and the harder

Are the Hard and Harder Problems really different problems? The Hard Problem is: why is the scientific basis of a phenomenal property the scientific basis of that property rather than another or rather than a non-phenomenal property? The question behind the Harder Problem could be put so as to emphasize the similarity: why should physically different creatures overlap phenomenally in one way rather than another or not at all? This way of putting it makes it plausible that the Harder Problem includes or presupposes the Hard Problem. In any case, the Harder Problem includes an issue that is more narrowly epistemic than the Hard Problem The Hard Problem could arise for someone who has no conception of another person, whereas the Harder Problem is closely tied to the problem of other minds. Finally, the Harder Problem involves an epistemic discomfort not involved in the Hard Problem. My claim is that the 'Harder Problem' differs from the 'Hard Problem' in these ways independently of whether we choose to see them as distinct problems or as part of a single problem.

Is the Harder Problem harder than the Hard Problem? If the Harder Problem is the Hard Problem plus something else problematic, then it is trivially Harder. As indicated above, the Harder Problem has an epistemic dimension not found in the Hard Problem, so they are to that

extent incomparable, but the epistemic difficulty involved in the Harder Problem makes it harder in one way.

Both the Hard and Harder Problems depend on what we cannot *now* conceive. Even the epistemic difficulty may be temporary, unlike the epistemic difficulty of the concept of *the gold mountain that no one will ever have evidence of*. Perhaps we will come to understand the nature of human consciousness, and in so doing, develop an objective theory of consciousness that applies to all creatures, independently of physical constitution. That is, perhaps the concepts developed in a solution to the Hard Problem will one day solve the Harder Problem, though I think our relation to this question is the same as to the Harder Problem itself, namely we have no conception of how to find an answer.

XI. What to do?

Naturalism dictates that physicalism is the default, but also inaccessible and meta-inaccessible; and in the 'subjective' sense mentioned earlier, it is the default that Commander Data is not conscious, but at the same time phenomenal realists regard his consciousness as an open issue. This is the Harder Problem. Alternatively, we could see the problem this way: if Commander Data is conscious, then we have a choice of Superficialism, Disjunctivism and Dualism. The Naturalist will want to reject Dualism, but it is cold comfort to be told that the only alternatives are doctrines that are epistemically inaccessible. So this may lead us to want to say that Commander Data is not conscious. But we have no *evidence* that he is or is not conscious.

What to do? To begin, one could simply live with these difficulties. These are not paradoxical conclusions. Physicalism is the default and at the same time meta-inaccessible. It is the subjective default that androids like Commander Data are not conscious but it is an open question whether they are. Consciousness is a singularity — perhaps one of its singular properties is thrusting us into these epistemic discomforts.

Another option would be to reject or restrict the assumption of naturalism or of phenomenal realism. One way to slightly debase naturalism would be to take the problem itself as a reason to believe the Disjunctivist or Superficialist form of naturalism. Those who prefer

to weaken phenomenal realism can do so without adopting one of the deflationist views mentioned at the outset (functionalism, representationism and cognitivism). One way to restrict phenomenal realism is to adopt what Shoemaker (op. cit.) calls the 'Frege-Schlick' view, that comparisons of phenomenal character are only meaningful within the stages of a single person and not between individuals. Another proposal is slightly weaker than the Frege-Schlick view in allowing only interpersonal comparisons across naturalistically similar persons. That is, though comparisons of phenomenal character among subjects who share a physical (or other naturalistic) basis of that phenomenal character make sense, comparisons outside that class are non-factual. Or else a significant group of them are false. That is, Commander Data either has no consciousness or there is no matter of fact about his consciousness.

Naturalistic phenomenal realism is not an unproblematic position. We cannot completely comfortably suppose both that consciousness is real and that it has a scientific nature. This paper does not argue for one or another way out, but is only concerned with laying out the problem.[30]

Ned Block (ned.block@nyu.edu)
NYU Department of Philosophy
503A Silver Center, 100 Washington Square East
New York, NY 10003, USA

---

[30] I would like to thank David Barnett, Paul Boghossian, Tyler Burge, Alex Byrne, David Chalmers, Hartry Field, Jerry Fodor, Paul Horwich, Brian Loar, Tom Nagel, Georges Rey, Stephen Schiffer, Stephen White and the editors of this journal for comments on earlier drafts. I am also grateful to Alex Byrne and Jaegwon Kim for reactions when an ancestor of this paper was delivered at a Central APA meeting in 1998. My thanks to the Colloquium on Language and Mind at NYU at which an earlier version of this paper was discussed in 2000, and especially to Tom Nagel as the chief inquisitor. I am also grateful to the audience at the 2001 meeting of Sociedad Filosofica Ibero Americana (SOFIA) and especially to my respondents, Brian McLaughlin and Martine Nida-Rümelin. And I would also like to thank my graduate class at NYU, especially Declan Smithies, for their comments. In addition, I am grateful for discussion at a number of venues where earlier versions of this paper were delivered, beginning with the Society for Philosophy and Psychology meeting, June 1997.