# A POSTERIORI IDENTITIES AND THE REQUIREMENTS OF RATIONALITY

## I. Prologue

Imagine that a medical team and submarine have been miniaturized and injected into the brain of a conscious subject to correct an otherwise irreparable condition. As team leader your greatest fear is that the subject, who is unaware of his situation, will take aspirin in response to the extensive c-fiber firing that you are apprehensively watching develop. For, as you know, in the subject's compromised condition, aspirin would cause nerve impulses and thus electrical activity in the brain that would completely destroy normal cognition. What you don't know is that the brain you are trying to repair is your own--that the nerve impulses from your body (minus the brain) produce the inputs (via a wireless connection) to the very brain whose activities you are witnessing. As a result, and because of the stress of the situation, you reach for the bottle of aspirin. . .

## II. The Challenge to the Identity Theorist

Ned Block's paper is a model of clarity and care, and the impossibility of an adequately detailed reply is a matter of genuine regret. What I propose is to concentrate on the challenge that the property dualism argument poses for a certain form of the mind-body identity theory and on Block's positive response. The challenge is to the theorist who identifies pain with some physical phenomenon (call it c-fiber firing) and for whom the identification is a posteriori. It is to produce an adequate explanation of the fact that one could be perfectly rational in believing, for example, what one would naturally express by saying "I am in pain but my c-fibers are not firing." I shall argue that Block has provided neither an explanation nor grounds to suppose that one exists. I shall then summarize the points on which we disagree.

This challenge to the identity theorist is appropriate because if the identification is to be made a posteriori, there must be a coherently characterizable possibility--a possibility that makes rational disbelief of, or lack of belief in, that identity possible. The question, then, is what this entails. And the answer is that if identity is true, then the subject who doubts it must apparently have two different modes of presentation of the same token event which is both the pain and the c-fiber firing. The obvious analogy (and one that the identity theorist insists upon) is with the possibility of one's believing, for example, what one would express by saying 'Hesperus is inhabited and Phosphorus is not'. And what makes this possible is the existence of two modes of presentation of Venus such that one believes of Venus under one mode of presentation that it is inhabited and under the other that it is not. Thus Block is correct in assuming that the relevant conception of mode of presentation is one that serves to explain the differences in cognitive significance of coreferring expressions.

What, though, is a mode of presentation? If we follow the model of Phosphorus and Hesperus, we will distinguish cognitive modes of presentation

(CMoPs) and metaphysical modes of presentation (MMoPs). Let us imagine a rational subject S who believes of Venus that it is and is not inhabited because S associates the description 'the last heavenly body visible in the morning' with 'Phosphorus' and the description 'the first heavenly body visible in the evening' with 'Hesperus'. Such descriptions belong on the representational side of the line that divides representations and/or content from the world. Thus as used by S, 'Phosphorus' and 'Hesperus' refer to Venus in virtue of the two distinct cognitive modes of presentation which are the two descriptions.

But why do these two cognitive modes of presentation pick out Venus? They do so, of course, in virtue of two real (and, apparently, distinct) properties of the planet--the property of being the last heavenly body visible in the morning and the property of being the first heavenly body visible in the evening. These are the metaphysical modes of presentation corresponding to the cognitive modes--they belong on the nonrepresentational side of the line that divides representations and/or content from the world. And this notion of correspondence, it seems, will make for a very tight connection--indeed an a priori connection--between cognitive modes of presentation and their metaphysical counterparts. After all, it is in virtue of the property of being the last heavenly body visible in the morning that Venus is picked out for S by 'Phosphorus' and in virtue of its being the first heavenly body visible in the evening that it is picked out for S by 'Hesperus'--and not vice versa.

Furthermore, it seems that it is in virtue of the expression 'being the last heavenly body visible in the morning's expressing the property of being the last heavenly body visible in the morning that it has the meaning (in the sense of cognitive significance) that it has. Similarly for 'the first heavenly body visible in the evening'. (The proposal that this could be a matter of inferential role alone is not an option. I shall discuss this below.) If so, then the fact that the predicate expressions (embedded in the definite descriptions) express the properties they do explains why the sentence to which S assents--'Hesperus is inhabited and Phosphorus is not'--has as its content (in the relevant sense) a coherently characterizable possibility. It is simply the possibility that the properties of being the last heavenly body visible in the morning and being the first heavenly body visible in the evening were instantiated by different objects, only the second of which was inhabited. And the identity theorist who believes that the identity is a posteriori is committed to providing a coherent account of what the world would be like if the belief of the uninformed subject S were not mistaken.

The move from a difference in cognitive significance to the existence of two metaphysical modes of presentation presupposes that we individuate properties very thinly. But suppose it is objected that the property of being the last heavenly body visible in the morning might well be identical with the property of being the first heavenly body visible in the evening. This does, after all, seem possible. We can imagine that scientists determine that, given all the physically possible trajectories of "heavenly bodies," these two properties coincide as a matter of physical necessity. And we can suppose that they determine that because they have the same extension at all physically possible worlds they are the same property. Certainly nothing prevents our individuating properties in this

coarse-grained way for some explanatory purposes. But there are other explanatory purposes besides those of theoretical science, and doing justice to the rationality of the subject who believes of Venus that it is and is not inhabited is one such purpose.

We are, then, if we identify the property of being the last heavenly body visible in the morning and the first one visible in the evening, committed to recognizing other properties to provide a rationalizing explanation of S's belief. For example, we might recognize different aspects or second order properties of the property of being the last heavenly body visible in the morning (= the property of being the first heavenly body visible in the evening). Or we might recognize modal properties, such as the property a body has in virtue of there being a logically or conceptually (but not physically) possible world at which the last heavenly body visible in the morning is not the first heavenly body visible in the evening. Could we identify even the second order or modal properties? Quite possibly the answer is yes. But this is not to the point. For to do so in the absence of other distinct properties in virtue of which Venus figures twice in S's thought without his knowing it is to leave unrealized an explanatory project to which the identity theorist is committed.

## III. Block's Response

Block disputes this. According to Block we can have two distinct cognitive modes of presentation, each of which picks out the object in question in virtue of the same metaphysical mode of presentation. Block's argument relies on his Paderewski example.

> Our subject starts out under the false impression that there were two Paderewskis of the turn of the century, a Polish politician and a Polish composer. Later he has forgotten where he learned the two words and remembers nothing about one Paderewski that distinguishes him from the other. That is he remembers only that both were famous Polish figures of the turn of the twentieth century.

With regard to the example Block claims:

> . . . there are two CMoPs but only one MMop, the MMoP being, say, the property of being a famous turn of the twentieth century Pole named 'Paderewski'. Thus "Paderewski = Paderewski" could be informative to this subject, despite identical MMoPs for the two terms.
>
> But, far from being informative,
>
> (1) Paderewski = Paderewski

seems to express a proposition that the subject cannot even entertain. What the subject can entertain, of course, are quantified propositions ("There was exactly

one famous Polish Paderewski at the turn of the twentieth century,""There were none," "There were two," etc.). But the ability to entertain such quantified propositions is not the ability to entertain a proposition such as the one expressed by (1) that involves genuine singular reference. Certainly Block's subject can wonder whether there were two Paderewskis or one. And he can even wonder whether, for example, Paderewski ever wrote to Paderewski (in the sense that he can wonder whether one Paderewski ever wrote to the other). But he cannot wonder, for example, whether Paderewski ever wrote to Paderewski <u>or vice versa</u>, since he cannot distinguish the propositional contents of the two disjuncts. Thus we do not have a case in which there are two distinct cognitive modes of presentation of a single referent, and indeed there is no possibility of singular reference at all. And there is no such possibility precisely because there is no cognitive mode of presentation of the kind that would be required to pick out a particular object of reference. Thus we do not have the kind of case Block wants.

In light of these facts, why does Block insist that there are two CMoPs of Paderewski? Presumably he is not imagining that the subject produces two orthographically different terms for Paderewski by stipulation--say by introducing subscripts. For how would S suppose that '$\text{Paderewski}_1$' came to pick out one rather than "the other"? Block's reference to S's "separate files" suggests that what he has in mind in talking about different cognitive modes of presentation are not different things at the personal level, but at the level of subpersonal causal mechanisms. After all, even if S cannot distinguish between different tokens of 'Paderewski' in his own usage, it could have been the case that some were caused by one person and some by another. Does this give Block what he needs?

The answer is no. And the reason is that subpersonal causal chains, by themselves, don't provide rationalizing explanations of subjects' beliefs. Suppose that S thinks that he dimly remembers that Paderewski was a musician (overlooking momentarily his belief that there are two Paderewskis whom he has apparently no way of distinguishing in thought). Later he thinks he dimly remembers that Paderewski was a politician (again momentarily overlooking his belief in "the two Paderewskis"). And because he remembers his earlier thought that he would have expressed by saying "Paderewski was a musician" he concludes that Paderewski was both a musician and a politician. This is clearly irrational, and irrational by S's own lights. But of course, in fact, the causal chains that produce the two thoughts and the two tokens of 'Paderewski' originate in the same person. Clearly, though, this doesn't get S off the hook--the inference is still irrational, subpersonal facts notwithstanding. Even putting this point aside, however, it is difficult to see how the reference to different files could help Block. If there are two distinct files, there are two causal chains and the property of being the source of one is different from the property of being the source of the other. Hence there seems to be no example in which we have two CMoPs and only one MMoP.

We can now see the problem for Block's view. Complications aside, Block holds that our normal mode of access to our pains is via phenomenal concepts--

descriptions in connection with which an instance of the pain itself occurs. That is, the pain is given in such a way that we stand in a demonstrative relation to it. Thus our mode of access to the qualitative character of our pains is direct in the sense that the physical property that is identical with the feeling of pain is its own mode of presentation. But this is to confuse directness in Russell's sense--acquaintance (in which sense-data are given directly and <u>are</u> their own modes of presentation)--with ordinary demonstrative access. And as the demonstrative versions of Frege's problem show, even in such cases, ordinary objects are not their own modes of presentation. In Evans' example, one points out a window at a ship and says "That ship was built in Japan." And one points out another window and says, "That ship was not," without realizing that one has pointed to the same ship twice.[1] It seems clear, then, that even in demonstrative cases, ordinary objects are picked out in virtue of some but not others of their properties.

## IV. Epilogue

The example in the prologue is simply a demonstrative version of Frege's problem for pains and c-fiber firings. Thus Block's positive account in terms of demonstrative access neither provides the two CMoP's and their corresponding MMoP's necessary to rationalize the subject, nor does it absolve us of the responsibility of doing so. But now consider: For any physical property that we might try to identify with the qualitative character of pain, we can imagine such a demonstrative version of Frege's problem. Thus until we postulate a mentalistic property of (or suitably related to) the pain (= the c-fiber firing), our commitment to doing justice to the rationality of the subject who (intuitively speaking) fails to believe one of the relevant identities remains undischarged.

## V. The Basis of the Disagreement

Having reviewed in synoptic form the challenge that the property dualism argument poses for the physicalist identity theorist, and my reasons for thinking that Block's positive account fails, I shall conclude with ten basic points on which I think Block and I disagree.

(1) We need two CMoPs to deal with the various versions of Frege's problem. But why couldn't the object be given once under a mode of presentation and once "directly"? Consider, however, that the only model the physicalist identity theorist has of our being given something directly is the way in which something is given when we stand in a demonstrative relation to it. (Russell's notion of acquaintance, for example, is not available to the physicalist.) Even in such a case, however, the object is given under a cognitive mode of presentation, as is shown by the existence of the demonstrative versions of Frege's problem.

(2) The cognitive modes of presentation must be available to the subject at the personal level. That is, they must be consciously available and characterizable from the first person point of view. To suppose that CMoPs don't have to be

described from the subjective point of view and given at the personal level is to forget their role in providing a rationalizing explanation for the subject.

(3) We need two corresponding metaphysical modes of presentation. These are required for a number of explanatory purposes.

> (a) We must explain what features the object has in virtue of which it is picked out by the two distinct CMoPs.
>
> (b) We must explain what the difference in the meaning (cognitive significance) of the two CMoPs consists in. And it cannot be merely orthographic. We can have two orthographically distinct names that we use completely interchangeably precisely because we associate exactly the same cognitive modes of presentation with them. Nor can it be solely a matter of linguistic descriptive content, since we must explain how the words in the descriptions come to have their meanings, and we cannot have an infinite regress of descriptions. Similarly it must be more than a matter of inferential role if this is understood solely in terms of language-to-language connections. (To suppose otherwise would be to suppose that language was merely an uninterpreted formal calculus.) Furthermore, the explanation cannot be a matter of either subpersonal functional states or external causal chains that are unavailable to the subject, either alone or in connection with descriptive content. This means that differences in the cognitive modes of presentation must be more than merely a matter of orthography, syntax, or causation. And as the demonstrative versions of Frege's problem show, even in demonstrative cases we pick out objects in virtue of some of their properties and not others. Thus we have to explain how language is grounded in the world, subject to the constraints imposed by Frege's problem. This is why there must be thin properties and, in the case of predicates embedded in descriptions, properties that correspond to them a priori.
>
> (c) We must explain what coherent possibility allows a perfectly rational subject to doubt the identity in question. And thin properties which are related a priori to the subject's CmoPs and which could (without contradiction) have been instantianted by different objects provide the appropriate explanation.

(4) The distinction between, on the one hand, a rationalizing explanation of a subject and, on the other, a causal explanation of a subject of a subject who is simply assumed (without explanation) to be rational cannot be ignored. To ignore the requirement that we rationalize subjects is to take a <u>locally</u> <u>eliminativist</u> view of the intentional states (in this case the belief states) in question. This is because (minimal) rationality is required by the ascription of intentional states and the constraints of radical interpretation. So to ignore the constraint in a certain class of contexts is to forego the intelligibility of belief ascriptions in those

contexts, and so the idea that there are beliefs in those contexts. And local eliminativism is unstable. If we treat beliefs in this way whenever the alternative presents difficulties, we are not committed to rationality (in belief ascriptions) and so should simply count ourselves as eliminativists across the board where intentional states are concerned.

(5) Problems with the description theory of reference do not warrant abandoning the idea that CMoPs must be available to the subject at the personal level. What we need is to know--to put it somewhat metaphorically--how the object gets into the subject's thought (at the personal level). And there is nothing in the commitment to this being in virtue of two CMoP's at the personal level (in the relevant examples) that requires that they be descriptions. Even in demonstrative cases (at least some of which could not be a matter of descriptive access if demonstrative access is to ground language), the same arguments for the existence of two CMoPs available to the subject apply.

(6) Ordinary objects cannot be their own modes of presentation. Again contemporary "direct reference" views might tempt us to think otherwise. But the project of providing the referents of the singular terms of a language of a community, of saying what is expressed for the community by the predicates, and of providing the truth conditions of the sentences (e.g., as the output of a compositional theory) cannot be assumed to be the same as the project of explaining differences in the cognitive significance of the linguistic expressions of a speaker.

(7) We cannot, then, in general take current semantic theories as unproblematic in this context. Nor can we assume that semantics, apart from issues directly connected with qualia, is unproblematic for physicalism. To suppose that a theory of cognitive significance must be physicalistic and then to reject what would otherwise seem to be necessary constraints on such a theory would clearly beg the question.

(8) We cannot avoid the issue of narrow content. Theories of narrow content have been motivated both by issues of supervenience and issues of cognitive significance, but increasingly the emphasis has been on the latter.[2] Thus the discussion of issues of narrow content is not an optional extra in this context.

(9) We cannot avoid fine-grained properties by looking at a subset of the logically or conceptually possible worlds. Though, as we have seen, two properties might have the same extensions at all physically possible worlds (and so be deemed the same property), such coarse-grained properties are not in general adequate for the rationalizing project to which the identity theorist is committed.

(10) We cannot avoid questions of ontological commitment by adopting a question-begging criterion. If we are committed to the rationality of subjects and

to the existence of qualitative states like pain and intentional states like belief, then we are committed to the things we need to postulate to make coherent sense of these things. That the upshot is Cartesian is by no means a foregone conclusion--in part because of the apparent alternative of treating pain as an intentional state.[3] But even if this proves to be possible, my view certainly involves a distinction similar to that between verstehen and causal explanation. And this is not a consequence that I find unwelcome.

Stephen L. White
Department of Philosophy
Tufts University

**Footnotes**

1. Gareth Evans, The Varieties of Reference (Oxford: Clarendon Press, 1982), p. 84.
2. See my "Partial Character and the Language of Thought," Pacific Philosophical Quarterly 63 (1982), pp. 347-365, reprinted in The Unity of the Self (Cambridge: MIT Press/Bradford Books, 1991), pp. 27-49 and "Narrow Content," in The MIT Encyclopedia of the Cognitive Sciences, Robert Wilson and Frank Keil, eds., (Cambridge: MIT Press/Bradford Books, 1999), pp. 581-583.
3. See Michael Tye, Consciousness and Persons: Unity and Identity (Cambridge: MIT Press/Bradford Books, 2003), pp. 48-62.