Philosophy of Mind: Mind-Body Identity and Eliminative Materialism

Neil Gascoigne, Royal Holloway University of London

Abstract

A critical outline is given of Rorty's early, 'eliminativist' attempt to formulate a materialist version of the mind-body identity theory that does not fall foul of the 'irreducible properties objection' (the thought that if mental states are brain states then the latter must exhibit the same properties as the former). An explanation is offered of why Rorty continued to describe himself as a materialist/physicalist despite having come to reject any version of mind-body identity.

1. Introduction

A recent article on Richard Rorty's early work in the philosophy of mind asks 'Was Rorty an Eliminative Materialist?' (Ramsay 2020). Its author's conclusion is that Rorty's 'Disappearance' version of the identity theory[1] amounts to little more than 'good-old-fashioned reductionist/identity' and that Rorty was a rather "more conventional sort of materialist" than the modifier would lead us to believe (Ramsay 2020, p. 27). On the one hand this is old news: Rorty's version of eliminativism came under considerable pressure in the late 1960s and early 1970s from critics like Bernstein (1968), Cornman (1968), Savitt (1974), Bush (1974), Lycan and Pappas (1972), and Hiley (1978). On the other hand, Robert Brandom offers a rather more ardent evaluation, describing Rorty's eliminative materialism as "the first genuinely new response to the traditional mind-body problem that anyone had seen in a long time" (Brandom 2000, p. 157). To complicate matters further, in *Philosophy and the Mirror of Nature* Rorty upbraids his earlier self by concluding that both reductionist *and* eliminativist/disappearance versions of the identity theory are "merely awkward attempts to throw into current philosophical jargon our natural reaction to an encounter with the Antipodeans[2] […] they should both be abandoned, and with them the notion of 'mind-body identity'." (Rorty 1979, p. 119). However, he proceeds to commend to us "a materialism which is not an identity theory in *any* sense" (Rorty 1979, p. 119). While the earlier pieces aimed to "vindicate" (Rorty 2014, p. 169) materialism, then, he later cautions the materialist against "saying metaphysical things" (Rorty 1979, p. 120). And when after nearly two decades of engaging with Davidson's work he does so to elucidate a 'non-reductive physicalism' it is likewise on the understanding that such a view "will not permit [the

materialist] to gratify his metaphysical, reductionist needs […] [and] claim that he has finally grasped the 'essence' of the world or of human beings" (Rorty 1991, pp. 116—117).

These conflicting assessments of Rorty's contribution to the philosophy of mind invite us to reconsider the question not only of Rorty's eliminativism but of his apparently persisting adherence of some version of materialism (or physicalism: I'll take the terms as synonymous here). On the face of it we appear to be presented with two lines of interpretation. If 'materialism of a conventional sort' connotes anything it is presumably adherence to a metaphysical theory about the nature of reality. Since it seems odd to characterize a pragmatist as maintaining any such commitment, we might perhaps endorse Ramsay's judgement—echoing as it appears to do Rorty's self-evaluation—and conclude that the earlier work inclines towards a ('conventional') metaphysical materialism, and that that is in stark tension with his pragmatist leanings. Call this Option 1. The alternative proceeds on the assumption that the relationship between Rorty's pragmatism and his materialism at the time was rather more complex than Option 1 provides. Since it promises to be more accommodating of Brandom's encomium and to provide some insights both into Rorty's rejection of his earlier position and into his apparent adherence to some (albeit *non*-metaphysical) version of materialism/physicalism it is Option 2 that is promoted in this chapter.

Of significance here is Rorty's relationship with two figures: Wilfred Sellars and Willard Quine, who—along with Carnap, Nagel and others—introduced pragmatist themes into mainstream analytic philosophy and for whom physicalism is less an explicit metaphysical theory (at least sometimes) than the presupposition that philosophy should take its generalizing lead from the methods of the natural sciences. Although I'll touch on Sellars' contribution, Quine was the more influential thinker at the time of Rorty's original contribution to the philosophy of mind and I'll examine briefly his contribution to the eliminativist debate in section 3. Before proceeding to that it's important to have some sense of the lie of the land concerning the Mind-Body Identity Theory, so section 2 will outline how Rorty's 'disappearance' version of eliminativism functions as a response to the bête-noire of reductionist identity: the 'irreducible properties objection'. In section 4 we'll revisit the 'disappearance' theory in the light of Quine's account, and evaluate the charge that Rorty's attempt to show that statements like 'I'm in pain' can express truths marks it out as nothing more (or less) than 'good-old-fashioned reductionist/identity'. In accordance with Option 2, Section 5 will indicate what status we should accord the materialism of Rorty's early identity theory and how as a consequence we should understand his continuing adherence to some version of materialism.

2. Identity and Elimination

An eliminativist maintains that some of the terms we use don't—contrary, perhaps to our 'ontological intuitions'—pick out or refer to real things (where 'real' can be understood in a number of ways ranging from 'fundamental constituents of the Universe' to'posits of our best theories'). In a casual sense one can be an eliminativist about true love, the Yeti, or the Lost City of Atlantis; but the specific concern of the eliminative *materialist* is on the terms we understand ourselves to be using ('intuitively' as it were) to refer to *mental* states and features. The emphasis of the early eliminative materialists[3] was on episodes with 'phenomenal content', like the sensation of pain (for the masochist), the smell of newly-mown grass (for the gardening enthusiast), or the experience of seeing red (for those with anger-management issues): occasions where talk of the mental precipitates consideration of some *immaterial* ontology. 'Second Wave' eliminativists[4], however, tend to focus more on the propositional attitudes, the assumption being that the barriers to materialism presented by phenomenal states were "dissolving" (P.M. Churchland 1981, p. 67) under that first-wave onslaught. From this perspective the ambition is to show not so much that the existence of immaterial items has to be impugned but that commonsense psychology is "so fundamentally defective", so "radically false" a theory of human behaviour that it will be "displaced, rather than […] reduced, by complete neuroscience" (P.M. Churchland 1981, p. 67).

Rejecting a 'reduction' of the mental to the neurological in favour of the wholesale 'displacement' of the former by the later sloganizes the eliminativist's repudiation of 'good-old-fashioned reductionist/identity', a classic statement of which is Smart's:

> in so far as "after-image" or "ache" is a report of a process, it is a report of a process that *happens to be* a brain process. It follows that the thesis does not claim that sensation statements can be *translated* into statements about brain processes. Nor does it claim that the logic of a sensation statement is the same as that of a brain-process statement. All it claims is that in so far as a sensation is a report of something, that something is in fact a brain process. Sensations are nothing over and above brain processes. (Smart 1962, p. 56)

This marked an important break from two influential traditions. Positivists like Hempel and Carnap had argued that propositions containing problematic terms like those purporting to refer to mental items were to be "translated" into test sentences that elucidated their meaning through conditions of verification. Equally, ordinary language philosophers like Ryle had sought to demonstrate that the Cartesian *lapsus* was to miscategorise the logical space of mental talk by assimilating its mode of being to that of physical objects[5]. If what these have in common is the

desire to *eliminate* the referring use of mental terms in the interests of ontological parsimony, then Smart (and Place) aim to redeem that use through identity. An enthusiast for *any* version of mind-body identity that benefits materialism, then, will acknowledge the following desideratum:

(ID) Demonstrate that as a result of conceptual changes wrought by empirical inquiry we might discover that mental items are *in fact* identical with specific brain processes.

Despite the apparent virtues of the Reductionist account a problem was quickly identified. Smart's emphases underscore the fact that although the identity in question will have resulted from a scientific discovery and thus be contingent[6], it is nevertheless "strict" (Smart 1962, p. 57). But if identity is indeed 'strict' then it would need to conform to Leibniz's principle, according to which if two objects are identical then any property possessed by one is possessed by the other:

$$(x)(y)[(x = y) \supset (F)(Fx \equiv Fy)]$$

If my best friend (x), who likes beer (F), turns out to be Jones' killer (y) then 'Jones' killer likes beer' (or some such expression) is also true. The concern is that although a fondness for the amber nectar is not the preserve of the morally upstanding and is thus not inappropriately applied to a killer, the properties associated with sensations and brain processes seem peculiarly ill-suited to apply to one another. Indeed, extending the experientially evocative range of predicates we employ in sensation-talk to the objective domain of neurological posits appears to be the very paradigm of "a conceptual mistake" (Cornman 1962, p. 77).

In response to this concern, proponents of the 'Reductive' version of identity suggested that specifications of the "mental" side could be given in—in Ryle's (1953) phrase—'topic-neutral' terms. Here's Smart again:

> When a person says, "I see a yellowish-orange after-image," he is saying something like this: "*There is something going on which is like what is going on when… I really see an orange*". (Smart 1962, p. 61)

One problem here is that on the obvious understanding of 'strict' identity the criterion for ensuring that a property F ascribed to x is also ascribable to y is if the designating phrases used are full-blooded, synonymous translations. To take the above example, the F picked out using 'is fond of ale' can only be identified as the same property if the phrase 'likes beer' is identical in meaning (not 'close': that might denote a *different* property); and even if in this case it is it's evident that 'topic-neutral' accounts like Smart's fall far short of that requirement. Indeed, since there's a sense in which more or less *any* experience can be likened to another the only way to constrain the meaning of a possible topic-neutral translation seems to be in terms of the very phenomenal properties one is seeking to be neutral about (Cf. Cornman 1962, pp. 76—79; 1971, pp. 43—46).

For Rorty, what he refers to as the "irreducible-properties objection" (Rorty 2014, p. 147) is fatal for Reductive versions of the identity theory. The task, then, is to satisfy (ID) without falling victim to it. To that end he employs a rather different sort of thought-experiment, asking us to imagine a primitive tribe who think that illnesses are caused by demons, which can be seen in close proximity to the sick by witch-doctors who've ingested a certain psychoactive substance (Rorty 2014, pp. 110—111). The suggestion is that the field anthropologist would not conclude that demons are 'strictly' identical with hallucinations; rather, they'd argue that there are no such things as demons. But since we'd be hard pressed to *prove* that demons don't exist to a philosophically-sophisticated witch-doctor willing to take on board the findings of modern medical science, the best we could hope for is to convince them that populating one's ontology with demons not only does nothing to supplement the explanatory and predictive power of modern medicine but introduces distinct explanatory problems of its own. The aim, then, would be to get the population to a point where they'd acknowledge the following:

- *What we* (the people…) *used to call* 'demons' *are really* hallucinations.

Likewise, the 'eliminativist' version of mind-brain identity is presented as:

- *What people used to call* 'sensations' *are really* certain brain-processes.

For Rorty, eliminations of the form

- *What people used to call* Xs *are really* Ys

can "equally well be paraphrased as 'Elimination of the referring use of the expression in question [Xs]… from our language would leave our ability to describe and predict [using Y-talk] undiminished.'" (Rorty 2014, p. 112). Call this the *Explanatory Equivalence Thesis* (*EET*).

The immediate appeal of *EET* is that the conceptual mis-match between the two sides of the identity is removed. Consider another example:

- *What people used to call* 'mermaids' *are really* manatee

Mermaid-talk involves the ascription of all sorts of properties, but no one thinks that those are properties of manatee even if it is true that reports of mermaid-sightings can be correlated with the copresence of manatees (and empty rum bottles…). The properties of what people *used to call* 'mermaids' are the properties of things they were *really* talking about: manatees. Likewise, on this account, the properties of what people used to call 'the sensation of pain' are properties not of pain but of the thing they were actually referring to when they engaged in pain-talk in order to describe and predict: namely, the relevant brain-process.

The 'disappearance' theory seemingly avoids the irreducible properties objection because the problematic properties are eliminated along with the items they are supposed to be the properties of. But it's important to recall that although it is eliminativist it is *also* intended as a

response to (ID). To see the importance of this congruence consider the following two questions:

Q(1). Can sentences including sensation-terms be true (can they be used to 'report' states of the organism by the organism)?

Q(2). Do sensation-terms retain their referring use?

Treating these as notionally independent gives us the following:

Table 1		Q(1).	
		Yes	No
Q(2).	Yes	Reductive identity (Smart; Armstrong; Place)	
	No	'Disappearence' eliminativism (Rorty)	'Revisionary' eliminativism (Quine; Feyerabend; 'Second-wave' eliminativists)

The reductive identity theorist maintains sensation terms can be used to report truthfully on states of the organism because they refer to real things; namely to specific brain-states. The revisionary eliminativist, on the other hand, wants to say that sensation-talk does not correspond to any respectable item in our materialist ontology and therefore is not truth-evaluable. What makes Rorty's account *eliminativist*, then, is his stance on Q(2), the disappearance of the referring use of sensation terms. But what makes it an *identity* theory (Q(1)) is that although sensation-terms don't refer—don't pick out attributively *mental* items—we nevertheless use them (albeit unwittingly) to articulate truths about the very same things we talk about *when* we talk *explicitly* about (specific) brain processes.

This determination to retain a truth-telling function for sensation terms—albeit as disguised reports of brain-processess—was regarded at the time as a measure of the relative sophistication of Rorty's eliminativism vis-à-vis Quine's. That is to say, it was interpreted as an acknowledgement that the ontic status of sensations was underscored by the fact that "What we are aware of is not postulated, and only the postulated is eliminable" (Cornman 1968b, 61). But as Rorty notes, "all my new line amounts to [beyond Quine's] is the suggestion that the reporting role of sensation-discourse could be taken over by a neurological vocabulary" (Rorty 2014, p. 203. Fn. 13). In part this meant showing how the authority accorded to the avowal of sensation-

reports could be captured without ascribing to them the sort of haecceity one might associate with a *private* object of which one is directly *aware*. In 'Mind-Body Identity, Privacy, and Categories' (Rorty 2014, pp. 120—131), Rorty capitalises on various Wittgensteinian arguments to the effect that the putative privacy of sensations cannot account for the "epistemological peculiarity" (Rorty 2014, p. 120) characteristic of such talk. The later 'Incorrigibility as the Mark of the Mental' draws on Sellarsian considerations (Sellars 1997) to show how the authority vouchsafed to reporters of 'mental' terms might have emerged as a result of linguistic innovation: that with enough effort put in to instigating the right sort of language game 'posited' theoretical terms could indeed come to be used to make non-inferential reports.

Both lines aim to establish that if the 'peculiarity' can be metaphysically deflated—if a report of a good ole-timey physical brain state could be shown to explicate it satisfactorily—then there's no further impediment to asserting the identity in question and *vindicating* thereby materialism (cf. Rorty 2014, pp. 120, 169). That is to say, we can impugn the ontology of the immaterial without having to confront the starkly counter-intuitive situation the revisionary eliminativist offers, whereby we must conclude that the majority of our beliefs were and continue to be false or in some sense meaningless. There is of course a great deal more to be said about how Sellars' account of Jones' genius for 'inventing the mind' prefigures Rorty's understanding of how to naturalise ordinary language by showing how the norms associated with different sorts of language game can be recast in social terms; that is to say, can in general terms be seen as 'posited' *as opposed to* representations of a prelinguistic 'reality'. Indeed, that is the basis of Brandom's positive assessment of Rorty's contribution to the Mind-Body problem as we'll see in section 5. But since Rorty's self-evaluation highlights his relationship to Quine let's say a little about the latter's contribution to this debate.

3. Quine on Explication and Elimination

The eliminativist case for materialism turns on impugning the public use of terms which appear to mandate confidence in the existence of the ontologically obnoxious by attempting to make plausible the possibility of conceptual changes that would leave such terms exposed as non-referring. The most influential attempt to explore the metaphilosophical basis of this sort of approach is the concluding chapter of Quine's *Word and Object*. For Quine, the Oxford tradition of elucidating or therapeutically disposing of philosophical difficulties like the mind-body problem (and the related status of the 'mental') through conceptual analysis was doomed because founded on the assumption that the criterion of correctness of an analysis—of, say, the concept of mind—turned on the relation of synonymy. As Quine (1951), following White (1950) had sought to establish, the concept 'sameness of meaning' and the related distinction between

analytic and synthetic statements simply can't do the philosophical heavy lifting required since they are part of a closed circle of mutually supportive concepts[7]. Rather than spell the end of philosophy-as-analysis, however, for Quine this presages a return to a variety of pragmatism, with the natural sciences and philosophy united in the same task. In this brave new world philosophy's specific role is not to *analyse* but to *explicate*. Moreover,

> explication is elimination. We have, to begin with, an expression or form of expression that is somehow troublesome. It behaves partly like a term but not enough so, or it is vague in ways that bother us, or it puts kinks in a theory or encourages one or another confusion. But it also serves certain purposes that are not to be abandoned. Then we find a way of accomplishing those same purposes through other channels, using other and less troublesome forms of expression. The old perplexities are resolved. (Quine 2013, p. 240)

To take one of our earlier examples, while mermaid-talk clearly served 'certain purposes', mermaid-lore became increasingly difficult to square with our developing knowledge of the natural world. Rather than simply dismiss mermaid-reports as hallucinations, however, we can acknowledge that they served to report *some*—albeit less traditionally pulchritudinous—*thing*; namely, manatee. The pragmatic explication of the concept of the mermaid turns out to be an elimination of the referring use of 'mermaid': mermaid-talk 'disappears' from our toolset.

The question this prompts is that if explication-as-elimination is offered in lieu of analysis-as-elucidation, what criterion of success replaces synonymy? To approach this question, we must first note that for Quine explication is a sub-category of elimination. The *explication* of X is a sort of elimination-in-favour-of Y rather than a 'disappearance'-elimination. The criterion of an explicative elimination, then, is that there exists a "striking if partial parallelism of function" between the old and new forms of expression (Quine 2013, p. 241); or, in Quine's pithier formulation, determining whether something is an explication is "simply a matter of" determining "whether the ostensible objects of the defective noun played roles that still want playing by *some sort of object*" (Quine 2013, p. 241, emphasis added). By way of an example, consider the vast scale of conceptual change involved in the shift from a broadly hierological to a naturalistic worldview. On the one hand we might conclude that nothing corresponds in the new order to whatever function was played by the noun 'God' in the old. On the other hand, when Feuerbach insinuates in *The Essence of Christianity* (1890) that 'theology is anthropology' we might regard him as *explicating* the concept of theology and accommodating a role for the "old mysterious objects minus the mystery" (Quine 2013, p. 241).

From Quine's discussion we can distil a number of points:

I.   Criteria for distinguishing between varieties of elimination are *pragmatic*.

II. A disappearance-elimination of X-talk will lead to a denial of the existence of Xs.

III. Explication is reduction (albeit of a specific form).

How then does Quine apply this to mind-body identity? Here's his "defence of" (Quine 2013, p. 243) or "brief for" (Quine 2013, p. 244) physicalism:

> If there is a case for mental events and mental states, it must be just that the positing of them, like the positing of molecules, has some indirect systematic efficacy in the development of theory. But if a certain organization of theory is achieved by thus positing distinctive mental states and events behind physical behavior, surely as much organization could be achieved by positing merely certain correlative physiological states and events instead [...] the bodily states exist anyway; why add the others? Thus introspection may be seen as a witnessing to one's own bodily condition, as in introspecting an acid stomach, even though the introspector be vague on the medical details. (Quine 2013, pp. 243-244)

Acknowledging that this differs little from positions outlined by Carnap (1956) and Feigl (1958), he nevertheless links it explicitly to the discussion of elimination-explication. And so the question arises: Is this defence/brief an elimination or an explication/reduction? Well, if one thinks of reduction in terms of strict identity, involving the resolution of "the mental states into the independently recognized elements of physiological theory" then it is not "ambitiously reductive" (Quine 2013, p. 245). The physical states in question are "states of the undivided organism" (Quine 2013, p. 243). On the other hand, neither is it straightforwardly eliminative. Indeed, Quine seems to think that only the philosopher who regarded the behaviouristic stance outlined above as a *metaphysical* theory (Cf. Quine 2013, p. 244, fn. 5) as opposed to a "mildly conceived […] physicalism" (Quine 2013, p. 244) would think that there's a *real* distinction here at all. No metaphysical battle fought; no materialist-eliminativist victory to celebrate.

4. Minus the Mystery

According to Quine, the marker of ontological distinction is to be posited, where to be posited is to be *talked about*, and the talk in question relates to the customary desiderata of empirical inquiry. He can remain insouciant about the elimination-explication issue because from his *contemporary* standpoint (and given the supposition of a 'mildly conceived physicalism') there is no requirement to posit the existence of sensations. But what, then, are we to think about statements like 'I am in pain' uttered in the nineteenth century[8]? Do we vouchsafe that pain did have a referring use *then* because it was a posit of the best explanatory theory of the time and thus embrace ontological relativism? Or do we rather insist on our own standpoint and claim

that there can never have been genuine talk about sensations whilst acknowledging that that judgement might equally (come to) apply to the posits of the present? Returning to Rorty's response to Cornman, then, we can see that the posited/nonposited distinction is a red herring. Rather than desiring to redeem Quine's 'bodily witnessings' as the witnessings of *non*-posits, it is against the background of this problematic that we need to situate his attempt to be an eliminativist whilst keeping faith with (ID) and retaining a truth-telling role for sensation-talk.

With that in mind let's turn to a problem Rorty's version of eliminativism appears to present us with. A further example Quine gives in defence of his 'relaxed physicalism', relates to molecular theory. When we discover that tables are 'swarms of molecules' are we to conclude that tables don't exist or retain our attachment to tables? Do we have here an explication or an elimination simpliciter? Once again, Quine declares the option "unreal" (Quine 2013, p. 244). Now for Rorty there is likewise no difference in kind between:

- *What people used to call* 'tables' are clouds of molecules, and
- *What people used to call* 'sensations' are certain brain-processes.

For the 'disappearance' eliminativist the whole point is of course to impugn the existence of sensations. But if tables and sensations have in common that *both* are subject to in-principle elimination simpliciter why aren't we compelled to conclude that tables too don't exist?! To account for how the basic eliminativist formula might capture this difference Rorty offers a six-step schema (Rorty 2014, p. 116) by which linguistic practices might undergo a temporal shift to bring about the desuetude of the referring use of an observation term:

(1) X's are the subjects of both inferential and noninferential reports;

(2) empirical discoveries are made which enable us to subsume X-laws under Y-laws and to produce new X-laws by studying Y's;

(3) inferential reports of X's cease to be made;

(4) noninferential reports of X's are reinterpreted either (4a) as reports of Y's, or (4b) as reports of mental entities;

(5) noninferential reports of X's cease to be made (because their place is taken by noninferential reports either of Y's or of thoughts, hallucinatory images, etc.);

(6) we conclude that there simply are no such things as X's.

The first thing to note is that we can associate stage (2) with Quine's eliminative explication/reduction and (6) with elimination simpliciter. Whether or not the intervening steps take place, then, is entirely down to pragmatic constraints. Rorty's take on the 'unreality' of Quine's option here is that the transition beyond (2) does not take place because "explanations formulated in terms of" tables are so good, "on the ground which they were originally intended

to cover," that we "feel no temptation" to stop talking about them (Rorty 2014, p. 117). 'Table' *maintains* its (inferential and noninferential) referring function and we continue to have (had) true beliefs about *tables*. Turning to mermaids, there is a sufficient parallelism of function to motivate the eliminativist identification, but in this case nothing prevents the move to (6). As a consequence, the existence of mermaids is impugned, the term loses its referential use, *and* we never had true beliefs about mermaids.

What then of sensations? The initial complication is that Rorty offers as the explicit reason we *don't* move from (2) in this case the same as for tables. In that sense, then, the choice is 'unreal'. But if 'table' retains its referring function at stage (2) despite the *in principle* elimination outlined in (1)—(6), then why does 'sensation' lose *its* referring use? Note that in both cases we can articulate truths using the terms because they refer to something, so what's the difference? One way to think about this is to contrast mermaids and tables. The in-principle disappearance of 'table' would not leave what 'table' would as a consequence refer to at (4)—viz, clouds of molecules—belonging to different ontological category. Unlike tables, mermaids come with a rich mythological history that makes little sense from within the naturalistic framework of inquiry. Mermaid sightings were rare of course. But if they'd been as common as table-reports; if, in other words, manatees were much more widely distributed; the reporting function of 'mermaid' might well have been the dominant aspect of its use. In such circumstances, when the in-principle elimination is made the continuity of that reporting function (the 'striking if partial parallelism') could have been sufficient to warrant the continuing use of the term 'mermaid' to refer to manatee. The existence of mythological mermaids is impugned, but the beliefs folk had and have are true (albeit of manatee). For Rorty, then, the case of sensations is akin to the revised mermaid example: the core reporting function is taken up by talk about brain-states, thus guaranteeing the truthfulness of beliefs and the stability of stage (2); but the uses associated with immateriality are impugned and with them the referring use of such terms (as *mental*).

At the beginning of this section we noted that Quine presents us with an unsettling choice between

1. Relativizing truth to the X-positing scheme, with the consequence that the beliefs folk had about Xs are true-for-them, but not true-for-us; and,

2. Privileging our own scheme, so the beliefs folk had about Xs were false/unreal.

The '*What people used to call* Xs *are* Ys' schema presents a standpoint on X-practices from the perspective of Y-practices that allows that X-talk was truth-apt whilst avoiding the relativizing implications of 1. But what of the threat posed by 2: might it not turn out that Y-talk comes to be similarly impugned in the future? It's evident that for Rorty it can't. To see why we need to

remind ourselves that Quine's 'brief' for or "defence" of physicalism presupposes a 'mildly conceived' physicalism which "declares no unbridgeable differences in kind between the mental and the physical" (Quine 2013, p. 244). There is no 'unbridgeable difference' because the positing of mental items is answerable to the same explanatory requirements that operate for physical items. From this pragmatic perspective physicalism is the claim that the sorts of items we need to posit to predict and control are the ones that fit in with our overall scientific conception of nature. What then are we to make of Rorty's headlining attempt to vindicate materialism? Well, we vindicate materialism by showing that we don't require the positing of any other sorts of items in order to account for the pragmatically useful self-reports that folk are able to make on their current brain-states (albeit using terms used traditionally to connote *mental* items). But in order for this to be a vindication of materialism as opposed to a mere disparagement of *im*materialism the identity of function carried over from mental- to brain-state reports has to be specified in a very narrow way.

Rorty's eliminativist vindication of materialism thus turns on the intelligibility of EET. If one could account for behavior as least as well by referring to brain states as by referring to putatively mental items then reference to the latter might disappear from the language and one would conclude that—like demons and mermaids—there never were such things. The principle objection to this proposal[9] is that even if the reporting role of sensation-terms has an explanatory function the latter is not exhaustive of it. Going back to the eliminativist schema, anything tying-in sensation-talk (X-talk) with brain-process-talk (Y-talk) through the elaboration of psycho-physical biconditionals will simply export to the latter the non-explanatory predicates used to characterise the former. As with the traditional identity theory what we'd mean when we reported an event in brain-process talk would be what we meant when we used phenomenal terms ('intense', 'throbbing') to characterise our sensations. As a consequence, nothing would have been eliminated. Rorty's response to this is to argue that one would only assume that what is being said is ineliminable if one held that a first-person report is a report of a some*thing* (a non-posit) that "we directly experience" (Cornman 1968b, p. 17): a some*thing* that could not turn out at some future date to be a no*thing*—or, indeed, a *something else*—because what makes a sensation a *pain* as opposed to a *tickle* (or an *acid stomach*) is its intrinsic phenomenal feel. As it stands, however, this does not fully address the point. The proper target of the criticism is Rorty's adherence to *EET*. *Pace* Cornman's anti-Quinean point (section 2), this objection does not struggle with Wittgensteinian and Sellarsian expostulations against treating awareness as a tip-off to ontic propriety (and thus ineliminability); rather, it merely opposes the assumption that sensation-talk is an inchoate scientific theory that *could* be replaced by one with equal or greater

explanatory power and asks how we can get ourselves into a position to assert the explanatory equivalence of X- and Y- talk.

We get a better sense of this problem by looking at another criticism. Consider the following[10]:

1. 'There are Xs' is true just in case 'X' is a referring term.
2. 'Sensations' refers to brain-processes.
3. Brain-processes exist.
4. 'There are sensations' is true.
5. 'There are sensations' is true just in case sensations exist.
6. Sensations exist.

Recall (Table 1) that 'disappearance' eliminativism endorses (4) but rejects (6). But (6) is implied by (4) and (5), and (5) seems to follow from (4) and (1) and the following:

(R) 'X' is a referring term just in case Xs exist.

Accordingly, one either denies (2) and embraces 'revisionary' eliminativism or accepts (6) and with it the reductive version of materialism. Now, the thrust of 'disappearance' eliminativism is that since it makes sense to argue that the referents of sensation-talk are brain-states we redeem the cognitive standing of first-person reports whilst dispensing with (positing) the immaterial. What makes such reports true, then, is that they refer to something *real* (that 'exists'). Now, as Rorty remarks, "entities referred to by expressions in one Rylean category may also be referred to by expressions in another […] [and] expressions in the first category may drop out of the language [be eliminated] once this identity of reference is realized" (Rorty 2014, p. 129). In order for brain-states talk to take over from sensation-talk, then, we need some criterion to establish that the identity of reference has in fact been realised. This returns us to Rorty's adherence to *EET*. *EET* stipulates the sorts of interests that determine the criteria for establishing a genuine identity of reference; namely, those relating to our ability to describe and predict, for which the candidate vocabulary is physico-neurological. Now, it's unsurprising that an identity theory purporting to be a version of *materialism* privileges such a vocabulary, but that leads to two conclusions. Firstly, that Rorty is not so much offering a vindication of materialism as presupposing its truth. Secondly, that the above argument reveals that if one supposes that the identity of reference is fixed by means of a vocabulary that constitutes an ontologically privileged standpoint, then it is natural to supposes that when one refers to something by making true statements that mention it, the thing referred to *exists*. Since (5) is implied by (4) and (1), then insofar as he is a *materialist who accepts (R)* Rorty is committed to (6): sensations are reduced to brain-states rather than eliminated.

5. At Home with the Antipodeans

In the introduction we noted that Rorty comes to the view that both reductionist *and* eliminativist/disappearance versions of identity should be rejected but that we should nevertheless sign up to materialism. The relevant sentence is worth quoting in full:

> The proper reaction to the Antipodean story is to adopt a materialism which is not an identity theory in any sense, and which thus avoids the artificial notion that we must wait upon "an adequate theory of meaning (or reference)" before deciding issues in the philosophy of mind. (Rorty 1979, p. 119).

The 'Antipodeans' are Rorty's imagined race who are raised from birth to employ widely available technology to report directly on states of their central nervous system. Since the contents of such reports are by hypothesis public, self-reporters are accorded no epistemic authority of the sort that could be assimilated to 'our' mental states[11]. As this makes clear, then, fealty to (ID) presupposes the possibility of completing some substantive semantic programme which would in turn *vindicate* materialism. Since the idea of founding identity on a theory of meaning was never an option, what Rorty comes to see is that adherence to the following causes the problem:

(R) 'X' is a referring term just in case Xs exist.

This implies that philosophy has access to an ontologically privileged standpoint; that one can establish identity of reference only if one has a prior sense of the lie of the ontological land. It is the *improper* response to an encounter with the Antipodeans because it conflates a materialist thought-experiment with the idea—the critique of which is central to *Philosophy and the Mirror of Nature*—that "philosophy should provide a permanent matrix of categories into which every possible empirical discovery and cultural development can be fitted without strain" (Rorty 1979, 123). To repeat: the error is to confuse the imagined 'cultural development' that the Antipodeans represent with the assumption that some fundamental discovery has been made about *what there really is*. Such a materialism is just an otherwise unwarranted decision to privilege the set of interests (prediction; control) constitutive of a certain family of vocabularies (the natural sciences).

In subsequent work Rorty aims to show that philosophy can reject as unrealisable—or as pragmatically unhelpful—the longing for a standpoint that transcends the vagaries of human culture whilst still accommodating a perspective from which conceptual changes of the sort the 'disappearance' eliminativist aimed to explicate can be evaluated[12]. Indeed, it is something along these lines that Brandom has in mind when he lauds Rorty's contribution to the mind-body problem and not the materialism/identity. As we noted in section 2, Rorty aims to naturalise the

authority of first-person reports—their 'epistemic peculiarity'—by showing how that authority might come into being through the instigation of a new sort of cultural practice: a new language-game. In this sense, what there 'is' is not a matter of representing reality but of pragmatic positing[13]. Of course, once one decides that what there 'is' is determined by social norms there is no reason to restrict what 'is' to the norms that happen to express our interests in prediction and control as the metaphysical materialist supposes. Monism gives way to pluralism; and from this pluralistic perspective the thought-experiments that philosophers and others devise—worlds in which 'all people are born equal', or 'God is dead', or 'general relativity and quantum mechanics are harmonised through M-Theory'—are all projections from where we are now, rooted in current norms, and prophetic gambles on what we might think and what there might as a consequence 'be' in the future.

Returning, then, to section 1, how are we to evaluate Rorty's eliminative materialism? According to Option 1, although his 'disappearance' theory is regarded as having "played an extremely important role in the formation of eliminative materialism" it amounts to nothing beyond that of the "conventional reductive materialist" (Ramsay 2020, p. 41). In other words, the very element Brandom celebrates in Rorty's eliminative materialism—the Sellarsian twist on Quine's 'witnessing to one's own bodily condition' that naturalises the normativity indicative of their 'epistemic peculiarity'—is regarded as having had nothing to do with the (albeit failed) elaboration of the position. To a degree this can be regarded as cherry-picking: it suits the contemporary proponent of eliminativism to ignore the implications for their own materialist position of Rorty's redescription of normative statuses in social terms; and it suits Brandom to see Rorty's career-launching contribution in terms of a trajectory that leads to his own project. Nevertheless, the questions that help delineate Option 2 still remain. Was Rorty a 'conventional' materialist at the time he offered his 'disappearence' theory, and by 'conventional' materialism should we understand Quine's 'mildly conceived physicalism' or something more metaphysically demanding? And what sort of materialist/physicalist did he subsequently become?

Let's start with the latter question. Recall the formula:

• *What people used to call* Xs *are really* Ys,

can be 'paraphrased' as the *Explanatory Equivalence Thesis* (*EET*). What Rorty comes to aver is that what makes Xs and Ys *what they are* is determined by the true sentences in which 'X' and 'Y' appear as signifying terms. And since the norm of truth is understood in social terms as use/justification, it is the overall pattern of such use—the pattern of truth—that determines *what* we're talking about when we talk about particular Xs and Ys. If we assume that Y-talk covers physical states, then the identity of the physical state in question is of course fixed in this way;

and likewise any item of X-talk. But now we see that since there's no standpoint from which to identify what there *really* is; or, to put it slightly differently, no language—'Z'-talk, say—that punches through merely social norms to the underlying ontology; there are no criteria for determining when an X has been either eliminated in favour of Y or reduced to Y. X-talk might "just fade away" but because "nobody any longer has a use for this sort of talk […] not because someone has made a philosophical or scientific discovery" (Rorty 1991, p. 115). Contra *EET*, there is no privileged standpoint from which we can say of Ys that they are *identical* in predictive function to Xs and that as a consequence the referring function of Xs can be taken over by Ys. What then does it mean to be a materialist/physicalist? The thought is that every event is a candidate for microphysical description: there is nothing characterizable as a something that is not in principle a something about which one can offer an explanation in terms of the physicist's favoured elementary particles.

This is a *non-reductive* physicalism: there is no sense in which elementary particle talk (Y-talk) can eliminate other forms of (X-) talk. Rorty opposes it, then, to the physicalism he ascribes to Quine, which he claims is asserted on metaphysical grounds (Rorty 1991, p. 116)[14]. One might think that Rorty concedes too much when he claims that "Mozart composing a melody" or "Euclid seeing how to prove a theorem" (Rorty 1991, p. 114) are events that can be brought under a physical description. After all, what makes them the events they relate to interests that have nothing to do with the explanatory sciences. But of course there's nothing to prevent a scientist scanning the brain of an active composer and claiming that they've finally explained the mystery of the compositional process. The philosopher can point out that the fact that both the scientist and the musicologist talk of 'the process of composition' doesn't mean that they are talking about the same thing. Without reduction the scientist can say that, and folk might believe them; but there's no warrant from philosophy for the strict *identification*. In summary, then, one might therefore regard Rorty's physicalism as a tactical appropriation of the term, something that aims to capture our anti-non-naturalistic intuition that there's nothing in the world than cannot be brought under a scientific descriptions of some sort (though we might still cavil at promoting physics in particular) whilst rejecting the sort of metaphysical appropriation of that conviction that undergirds the programmatic aspirations of contemporary eliminativists and scientific realists.

This brings us to the final requirement: how to characterise Rorty's materialism at the time of the 'disappearence' theory. Was he, as Ramsay suggests, a metaphysical materialist of the 'conventional" sort? And here I think the response has to be 'No!'. That might seem at odds with the conclusion drawn at the end of section 4 that his position amounted to reductive

naturalism. But the error Rorty came to see himself as having made was not that of proposing as a possibility the standpoint of (what is in effect) the Antipodeans on our practices, but having regarded that standpoint as a philosophical one: of combining, as it were, a prophetic proposal/projection from where we are now with (R). When at the end of section 4 I charged Rorty with having presupposed the truth of materialism, then, the intention was not to imply that he presupposed the truth of a *metaphysical* thesis which would be in stark conflict with his pragmatism. That would be to support Option 1. Rather, the intention was to draw attention to what Rorty envisaged such a standpoint on present/past/other practices afforded one. That is to say, he assumed that the proposed standpoint would furnish one with the criteria necessary to make the appropriate identities. And that, he came to determine, would presuppose the availability of the sort of substantive theory of meaning or reference with which the pragmatist should have no truck[15].

6. Conclusion

In this chapter I've attempted to show that Rorty's adherence to the term physicalist should be regarded as metaphysically deflationary in intent; but it may well have been more useful to have abandoned the term altogether and to have concentrated on what in contemporary culture is a more evocative concept: "naturalism". That aside, my aim has been to demonstrate that qua materialism Rorty's "disappearence" theory was not expressive of a conventional materialism as per Option 1 but of an erroneous presentation of what is after all an imagined possibility represented by the Antipodeans: that one might come to speak 'Physicalese'. By categorising the 'disappearence' theory as a benighted attempt to formulate an eliminativist position of the sort associated with the Churchlands, Stich and himself, philosophers like Ramsay have excluded Rorty's subsequent work from consideration. But one implication of the interpretation on offer is that it is not so much Rorty's early theory that is relevant to the plausibility of contemporary versions of eliminativism but his excavation of the error he came to see himself has having made and the anti-reductive account it led to[16].

Bibliography

Bernstein, R. J. 1968/2000. The Challenge of Scientific Materialism. In *Materialism and the Mind–Body Problem*, 2nd edition. Ed. D. M. Rosenthal, 200–222. Indianapolis: Hackett.

Carnap, R. 1956. The Methodological Character of Theoretical Concepts. *Minnesota Studies in the Philosophy of Science* 1: 38–76.

Brandom, R. Ed. 2000. *Rorty and His Critics*. Cambridge: Blackwell.

Bush, E. 1974. Rorty Revisited. *Philosophical Studies* 25: 33–42.

Churchland, P. M. 1979. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.

Churchland, P. M. 1981. Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy* 78(2): 67-90.

Churchland, P. M. 1984. *Matter and Consciousness*. Cambridge, MA: MIT Press.

Churchland, P. S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, Massachusetts: MIT Press.

Cornman, J. 1962/2000. The Identity of Mind and Body. In *Materialism and the Mind–Body Problem*, 2nd edition. Ed. D. M. Rosenthal, 73–9. Indianapolis: Hackett.

Cornman, J. 1968a. On the Elimination of "Sensations" and Sensations. *The Review of Metaphysics* 22: 15–35.

Cornman, J. 1968b. Mental Terms, Theoretical Terms, and Materialism. *Philosophy of Science* 35(1): 45–63.

Cornman, J. 1971. *Materialism and Sensations*. New Haven and London: Yale University Press.

Feigl, H. 1958. The 'mental' and the 'physical'. *Minnesota Studies in the Philosophy of Science* 2: 370–497.

Feuerbach, L. 1890. *The Essence of Christianity*, 2nd edition. Trans. M. Evans. London: Kegan Paul

Feyerabend, P. 1963. Mental Events and the Brain. In *Materialism and the Mind–Body Problem*, 2nd edition. Ed. D. M. Rosenthal, 172–173. Indianapolis: Hackett.

Gascoigne, N. 2008. *Richard Rorty: Liberalism, Irony and the Ends of Philosophy*. Cambridge: Polity.

Gascoigne, N. Forthcoming. After Metaphysics: Eliminativism and the Protreptic Dilemma. In *The Cambridge Companion to Rorty*. Ed. D. Rondel. Cambridge: Cambridge University Press.

Grice, H. P. and Strawson, P. F. 1957. In Defense of a Dogma. *The Philosophical Review* 65(2): 141–158.

Hiley, D. R. 1978. Is Eliminative Materialism Materialistic? *Philosophy and Phenomenological Research* 38: 325–337.

Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.

Lycan, W. G. and Pappas, G. S. 1972. What is Eliminative Materialism? *Australasian Journal of Philosophy* 50(2): 149-159.

Place, U. T. 1956. Is Consciousness a Brain Process? *British Journal of Psychology* 47: 44–50.

Quine, W. V. 1952/1976. On Mental Entities. In *The Ways of Paradox*, 2nd edition, 221-227. Cambridge, MA: Harvard University Press.

Quine, W. V. 1980. *From a Logical Point of View*, 2nd edition, revised. Cambridge, MA: Harvard University Press.

Quine, W. V. 2013. *Word and Object*. Cambridge, MA: MIT Press.

Ramsay, W. 2020. Was Rorty an Eliminative Materialist? In *A Companion to Rorty*. Ed. A. Malachowski, 25-42. Oxford: Blackwell.

Rorty, R. 1976. Realism and Reference. *The Monist* 59: 321–340.

Rorty, R. 1979. *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.

Rorty, R. 1991. *Objectivity, Relativism, and Truth*. Philosophical Papers, vol. 1. Cambridge: Cambridge University Press.

Rorty, R. 2014. *Mind, Language, and Metaphilosophy: Early Philosophical Papers*. Eds. S. Leach and J. Tartaglia. Cambridge: Cambridge University Press.

Ryle, G. 1929. Review of *Sein und Zeit*. *Mind* 38(159): 355-370.

Ryle, G. 1954. *Dilemmas: The Tarner Lectures 1953*. Cambridge: Cambridge University Press.

Savitt, S. 1974. Rorty's Disappearance Theory. *Philosophical Studies* 28(6): 433-436.

Sellars, W. 1997. *Empiricism and the Philosophy of Mind*. Ed. R. Brandom. Cambridge, MA: Harvard University Press.

Smart, J. J. C. 1962/2000. Sensations and Brain Processes. In *Materialism and the Mind–Body Problem*, 2nd edition. Ed. D. M. Rosenthal, 53-66. Indianapolis: Hackett.

White, M. 1950. The Analytic and the Synthetic: An Untenable Dualism. In *John Dewey: Philosopher of Science and Freedom*. Ed. S. Hook, 316-331. New York: The Dial Press.

Further Reading

Brandom, R. 1998. *Making it Explicit*. Harvard: Harvard University Press.

    A systematic development of the innovative approach to the practice-based explication of social norms that Brandom ascribes to Rorty's eliminative materialism.

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.

    Essays outlining the 'reason can be causes' thesis that Davidson calls 'Anomalous Monism' and which is one of the sources of Rorty's non-reductive physicalism.

Gascoigne, N. 2015. Richard Rorty. *Oxford Bibliographies in Philosophy*. Ed. D. Pritchard. New York: Oxford University Press. doi: 10.1093/obo/9780195396577-0190

    A critical bibliography of works on and by Rorty.

McDowell, J. 1998. Reply to Commentators. *Philosophy and Phenomenological Research* 53(2): 403–432.

    On the 'baldness' of Rorty's naturalism amongst other things.

Ramsey, William. 2013. Eliminative Materialism. In *The Stanford Encyclopedia of Philosophy*, summer edn. Ed. E. N. Zalta. http://plato.stanford.edu/archives/sum2013/entries/materialism-eliminative.

An informative overview of how eliminativism developed in the aftermath of Rorty's early work.

Notes

[1] Rorty first proposed what he refers to as the "disappearance" form of Identity Theory in his 1965 essay "Mind-Body Identity, Privacy, and Categories". The position is extended and defended in the "Incorrigibility as the Mark of the Mental" (1970) and "In Defense of Eliminative Materialism" (1972). These and other important essays from the period are collected in Rorty 2014 (chapters 6, 8 and 10 respectively). The more popular descriptor 'eliminative' materialism was coined by Cornman (1968a).

[2] I'll say more about the Antipodeans in section 5.

[3] Cf. Quine (2013, 1966) and Feyerabend (1963).

[4] Cf. P.M. Churchland (1979, 1981, 1988), P.S. Churchland (1986), Stich (1983).

[5] Ryle was of course familiar with the German phenomenological tradition and famously reviewed *Being and Time* (Ryle 1929).

[6] For an influential contrasting view of the status of such identities see Kripke 1980.

[7] For a response to this from the 'Oxford' school see Grice and Strawson 1957.

[8] "for my part I do, qua lay physicist, believe in physical objects and not in Homer's gods; and I consider it a scientific error to believe otherwise. But in point of epistemological footing the physical objects and the gods differ only in degree and not in kind. Both sorts of entities enter our conception only as cultural posits" (Quine 1980, p. 44).

[9] Cf. Cornman 1968a, 1968b; Bernstein 1968; Savitt 1975.

[10] Adapted from Hiley (1978) after Lycan and Pappas (1972).

[11] The name is of course a tongue-in-cheek homage to the Australasian materialists who first propounded the reductive version of the identity theory.

[12] See Gascoigne (forthcoming) for a fuller account.

[13] The positing of what we need to 'talk about'. See Rorty 1976.

[14] "If we are limning the true and ultimate structure of reality, the canonical scheme for us is the austere scheme that knows no quotation but direct quotation and no propositional attitudes but only the physical constitution and behavior of organisms" (Quine 213, p. 202).

[15] In Gascoigne 2008, ch. 3 I melodramatise this as Rorty's 'Kehre'.

[16] My thanks to Martin Müller for 'encouraging' me to finish this and for his editorial acumen!