Eliminativism, Reference and Vocabulary Replacement:

Sellarsian Roots of Rortian Pragmatism

(Draft)

Martin Gustafsson

Stockholm University

It may be hard to see the connecting threads between the Princeton professor whose tightly argued "Incorrigibility as the Mark of the Mental" (1970) and "Functionalism, Machines, and Incorrigibility" (1972) were aimed specifically at the smallish clan of analytic philosophers of mind, and the international man of letters described by Harold Bloom as the most interesting philosopher in the world. Can we see the stirrings of Rorty's later ideas in between the lines of his early papers in the philosophy of mind? Perhaps, but that will not be my topic. (Dennett 2000, 91)

What is not Dennett's topic in the paper from which this quote is taken will indeed be the topic of the paper you are just about to read. My aim in what follows is to find some of the "connecting threads" between those publications from the 60's and early 70's by which Rorty built himself a position in the American analytic-philosophical community, and the book by which, about a decade later, he came to be known as a detractor of ideas central to that community: *Philosophy and the Mirror of Nature* from 1979. I will argue not only that there are such connections to be found, but also that seeing those connections is crucial for an

adequate understanding of the nature of Rortian neo-pragmatism. A full study of this topic would have to involve a discussion of his later, post-1979 production as well, but in this paper I will leave those later works largely aside. In fact even Rorty's development during the 60's and 70's is much too complex to be fully covered within a paper such as this. What I will do is to highlight some of the most intriguing features of that development.

In his recent sociological study of Rorty's career, Neil Gross points out that the standard story, according to which "Rorty started out as a hard-nosed analytic philosopher and only later came to doubt the value of the analytic program," is wrong. (Gross 2008, 308) Gross notes that Rorty did his undergraduate and graduate work at Chicago and Yale – departments at which, in the 50's, analytic philosophy was looked at with considerable skepticism. In his Doctoral dissertation on the concept of potentiality, Rorty does engage extensively and critically with logical empiricism, noting that the post-war works of thinkers such as Hempel, Goodman and Sellars have freed the movement "from some of the more dogmatic aspects of positivism." (Rorty 1956, 413; quoted in Gross 2008, 143) However, the dissertation as a whole is certainly not the work of a "hardnosed analytic philosopher", but rather of a historically oriented metaphysician.

Gross also emphasizes Rorty's early exposure to pragmatist ideas. At Chicago, the influence of Dewey was palpable, but the young Rorty seems to have been more attracted to Peirce. This attraction was encouraged by his supervisors for his Master's thesis (Charles Hartshorne) and his Doctoral dissertation (Paul Weiss), who were the editors of the first edition of Peirce's *Collected Papers*. Hartshorne in particular was deeply influenced by Peirce's philosophy.

In his attempt to play down Rorty's analytic heritage, Gross goes as far as describing the papers on mind-body identity and eliminativism from the 60's and early 70's as an isolable part of his work – a part the character of which was largely conditioned by career

tactics. According to Gross, these papers

> are best read as a distinct piece of his oeuvre. They represent Rorty's attempt to make contributions to analytic thought of a piece with those that other bright, young analytic philosophers of his generation were making. They were, in other words, part of Rorty's efforts to position himself even more squarely within the mainstream philosophical establishment. [...] it is suggestive of a connection between his works on these topics and his interest in promotion that he began writing some of the relevant articles only a few years before he had to prepare his tenure file. (Gross 2008, 185)

This passage seems to me to manifest the limitations of Gross's sociological approach. A careful study of papers such as "Mind-Body Identity, Privacy, and Categories" (1965) and "In Defense of Eliminative Materialism" (1970) shows that if we want to understand Rorty's *philosophical* development, these papers should *not* be seen as a "distinct" piece of Rorty's oeuvre. Rather, they constitute crucial steps toward the specifically Rortian type of pragmatism that is found in his later works. In particular, these early investigations into mind-body identity and eliminative materialism display very clearly how Rorty is eventually led to a methodological view characteristic of his mature philosophy – a view which includes as its central conception that of *vocabulary replacement* as a way of getting rid of philosophical problems. My aim in this paper is to clarify this particular aspect of Rorty's development.

In accounts of Rorty's philosophy, his earliest publications are often simply ignored. For example, in his book on Rorty, Alan Malachowski pays very little attention to this phase of his development. Some readers, however, do emphasize the connection between these papers and what happens later. (See, e.g., Brandom 2000, 158). The most ambitious attempt to do so is Neil Gascoigne's recent book-length study of Rorty's philosophy (Gascoigne

2008). Gascoigne devotes the first 100 pages to a detailed investigation into Rorty's early eliminativism and what he describes as Rorty's "Kehre" at the beginning of the 70's, leading up to the sort of conception presented in *Philosophy and the Mirror of Nature* (the bulk of which was written already in 1974).

I am to a large extent in agreement with Gascoigne's presentation of these matters. In my discussion, however, the point of emphasis will be somewhat different. In Gascoigne's story, Quine and Davidson appear as the deepest influences on Rorty during the relevant period. In what follows, I will instead stress the importance of Sellars. Of course, I am not denying that Quine and Davidson played important roles too (as did Kuhn, Geertz, Foucault, Dewey, and later Wittgenstein). But I tend to think that Rorty's perception of the central problems that he was dealing with at this time was even more colored by his reading of Sellars's work. For one thing, Sellars seems to have been the analytic philosopher with whom Rorty himself felt most affiliated. In his "Intellectual Autobiography" written for the forthcoming volume on his work in the *Library of Living Philosophers*, Rorty writes:

> Even at Yale the suspicion was growing that Carnap and Quine might be riding the wave of the future. So I began looking around for analytic philosophers who were less reductionistic and less positivistic than they, less convinced that philosophy had only recently come of age. This led me to the work of Sellars, whose work set me on the paths that I have spent the rest of my life trying to clear and broaden. Sellars combined a Carnapian style (lots of numbered premises, bedecked with lots of quantifiers) both with a thorough acquaintance with the history of philosophy and with an exuberant metaphysical imagination. That mixture of logic-worship, erudition, and romance was reminiscent of Peirce, with whose writings I had spent a lot of time, hoping to discover the non-existent secret of his non-existent "System," and, in particular, to figure out

what he meant by "Thirdness is real." Sellars and Peirce are alike in the diversity and richness of their talents, as well as in the cryptic character of their writings. But Sellars, unlike Peirce, preached a fairly coherent set of doctrines. (Quoted in Gross 2008, pp. 312-313)

To be sure, Rorty made objections against much of what Sellars had to say. So, their special relation does not seem to have consisted in Rorty's accepting more Sellarsian doctrines than, say, Quinean or Davidsonian ones. My hunch is rather that, in many cases, even those Sellarsian doctrines that Rorty rejected were of a special importance to him; for those were often doctrines to which he himself would feel genuine *attraction*. If this is right, Rorty's objections against Sellars can plausibly be construed also as more or less hidden self-criticisms. Rorty's engagement with Sellars during this period is at the same time an engagement with his own developing views. Hence it provides an especially useful key to understanding what this development involved.

1. Sellars, Science and Common Sense

In 1971, Rorty published a review of Alfred Ayer's book on Peirce and James, *The Origins of Pragmatism*. The review is courteous, but it is clear that Rorty thinks Ayer misses the depth of pragmatist viewpoints. For example, Ayer criticizes Peirce's notion of truth as what will be believed at the end of inquiry, by pointing out that there are many statements whose truth will simply not be the subject of future inquiry – such as what clothes I'm wearing today. According to Ayer, Pierce's definition manifests an unfounded preference for scientific method over other ways of fixing belief.

"But surely," Rorty comments, "there is more than this to be gotten out of the topic." He continues:

When we consider the problem of how to analyze changes in conceptual framework (as this problem is raised by, for example, Kuhn and Feyerabend), we run into truths which change to falsehoods even though the criteria for truth within the framework in which they were originally proposed are still satisfied. To aid in analyzing this situation, writers like Sellars (in his recent *Science and Metaphysics*) have made good use of Peirce's notion of the ultimately adequate conceptual framework as a regulative ideal. Indeed it is hard to see how the notion of truth as something which transcends conceptual frameworks can survive without some such notion. (Rorty 1971, 97)

In its context, this passage may be taken to suggest that the problem gestured at arises only when we want to analyze changes in, or shifts between, different *scientific* conceptual frameworks. But in Sellars, and, as we shall see, in Rorty as well, the problem in question is central also when it comes to the relation between science and common sense. In *Empiricism and the Philosophy of Mind*, Sellars famously claims that there is a sense in which "the scientific picture of the world *replaces* the common-sense picture; a sense in which the scientific account of 'what there is' *supersedes* the descriptive ontology of everyday life." (Sellars 1997[1956], 82; original italics.) The problem Rorty thinks Ayer should have discussed arises when we try to understand what such "replacing" or "superseding" may consist in.

Sellars warns that one must be cautious here. He distinguishes between a right and a wrong way of conceiving the sort of replacement that is at issue. Consider, for example, the common sense view that physical objects have colors. Has science refuted this idea? Has science shown that physical objects aren't really colored? Well, Sellars argues, that depends on what you mean. If you mean that the sentence 'Physical objects have colors' expresses an

empirical proposition which is widely believed to be true but which science has proven false, then what you are claiming is "of course [...] absurd." For your claim to be correct, Sellars continues, you must instead think of the sentence 'Physical objects have colors' as part of the framework within which common sense talk of physical objects – including ordinary color judgments such as 'This desk is brown' – make sense at all. This, in turn, means to conceive of the counter-assertion, 'Physical objects aren't really colored',

> only as a clumsy expression of the idea that there are no such things as the colored physical objects of the common-sense world, where this is interpreted, not as an empirical proposition – like "There are no nonhuman featherless bipeds – *within* the common-sense frame, but as the expression of a rejection (in *some* sense) of this very framework itself, in favor of another built around different, if not unrelated, categories. (Sellars 1997[1956], 82)

But still, the question remains: In *what* sense does science reject common sense, if not in a straightforwardly empirical manner? What is it to reject a "framework" rather than a set of ordinary falsehoods? Sellars makes two further points in this connection. First, he says that the rejection in question does not have to recommend any actual changes of everyday linguistic practice: "It need not [...] carry with it a proposal to brain-wash existing populations and train them to speak differently." Second, he notes that from the viewpoint of a participant who is immersed in these established practices, it will remain wrong to deny that objects are colored:

> [O]f course, as long as the existing framework is used, it will be *incorrect* to say – otherwise than to make a philosophical point *about the framework* – that no object is

really colored, or is located in Space, or endures through time. But, *speaking as a philosopher*, I am quite prepared to say that the common-sense world of physical objects in Space and Time is unreal – that is, that there are no such things. Or, to put it less paradoxically, that in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, of what is not that it is not. (Sellars 1997[1956], 82-3; original italics)

Sellars is trying to avoid three different positions that he finds mistaken. One is a straightforward error theory about common sense color ascriptions, according to which such ascriptions are simply false. Another is a reductionist view of common sense color ascriptions, according to which such ascriptions are somehow translatable into scientific statements about micro-particles, light-waves etc. The third position is one according to which common sense talk of colored objects and scientific talk of light-waves belong to conceptual frameworks which are so different or "incommensurable" that they cannot be compared and evaluated other than according to purely "pragmatic" criteria (in Carnap's sense). According to this third sort of position, there is nothing like "*the* dimension of describing and explaining the world" by reference to which different frameworks can be compared and ranked, and (it is concluded) hence no cognitive conflict between the judgments of the common man and those of the scientist.

In his attempt to avoid all these three conceptions, Sellars thinks he must somehow combine two claims. On the one hand, he argues that to say that the scientific picture of the world *supersedes* the common-sense picture is not to say that science wins a mere power struggle. Science wins the day, not just because of, say, the cultural authority or the rhetorical or economic strength of its spokesmen, but due to persuasion that is in some sense rational. The replacement of common sense by science is in some sense a cognitive achievement. On

the other hand, Sellars maintains that it is still in some sense correct within the common sense frame to say of things that they are colored. How on earth can these two claims be compatible?

Another way of putting the difficulty is in terms of an apparent tension within Sellars's use of the term 'frame' or 'framework'. As we have seen, this is a term he is happy to use. And he wants to think of the scientific and common sense frameworks as separate enough to allow us to say that it is correct "within" the common sense framework and incorrect "within" the scientific framework to say that things are colored. On the other hand, he does not want his notion of framework to open the doors to full-fledged tolerance or incommensurability. In fact, it is the central point of Sellars's rejection of what he calls the "positivistic" or "peninsular" conception of science that science and common sense are in a deep sense inseparable. More precisely, Sellars thinks science is a development and refinement of capacities or tendencies already present in everyday discourse. According to Sellars, the transition from pre-scientific to scientific discourse involves the exploitation of resources somehow available already before that transition was made:

> [W]hat we call the scientific enterprise is the flowering of a dimension of discourse which already exists in what historians call the "prescientific stage," and [...] failure to understand this type of discourse "writ large" – in science – may lead, indeed, has often led to a failure to appreciate its role in "ordinary usage," and, as a result, to a failure to understand the full logic of even the most fundamental, the "simplest" empirical terms. (Sellars 1997[1956], 81)

The "failure to understand the full logic of even [...] the 'simplest' empirical terms" that Sellars is talking about here is the failure of falling prey to the Myth of the Given – roughly,

the failure of thinking that terms like ordinary color words get their meaning simply from procedures of ostension the functioning of which is supposed to be independent of other linguistic habits that the language learner happens to have achieved. According to Sellars, someone who falls prey to this Myth will be inclined to think that such simple, primitive strata of linguistic practice are immune to improvement – and, in particular, that scientific results can never motivate linguistic reform at this basic level. Rather, science will be seen as an additional layer whose meaning and purpose are to allow us to deal more efficiently with the basic raw material that constitute its independently given data. To abandon this Myth, on the other hand, means to realize that scientific developments may well lead us to question the adequacy even of the seemingly most fundamental and simple levels of ordinary language use. For example, it means to realize that the development of science may well undermine the wider framework of linguistic habits within which our use of, say, color words, makes clear sense.

But the difficulty crops up again: What, exactly, does such undermining involve? This difficulty becomes even more pressing for Sellars in later works, such as in his 1968 *Science and Metaphysics*, to which Rorty refers in the passage from the Ayer review quoted earlier. In this series of lectures, Sellars allows himself to spell out the "correctness" of common sense color ascriptions in terms of *truth*. So, now Sellars is prepared to say that, within the frame of common sense, it is indeed *true* to say that things are colored. The problem is that this seems *plainly* incoherent with his claim that science has shown that the common sense world of colored, physical objects is unreal. After all, this latter claim seems to entail that our common sense talk of colored objects is on a par with talk of witches and unicorns –and such talk, one wants to say, is just *false*.

In 1970, Rorty reviewed *Science and Metaphysics*. His focus in this review is precisely on Sellars's attempt to handle this apparent contradiction. Sellars's viewpoint is very intricate,

and his presentation of it in the lectures is quite obscure. I will not attempt a detailed exegesis of Sellars's actual conception, but only point out its main features as Rorty understands them.

Fundamental to Sellars's account is a distinction between, on the one hand, the notions of *truth* and *denotation* – notion that he thinks are framework-relative, and thus applicable only from *within* some particular framework – and, on the other hand, the notion of the *adequacy* of a framework as a whole, as compared to other frameworks. According to Sellars, as Rorty understand him, if I say, '"Plato" denotes the teacher of Aristotle', I am not making a statement about the relation between the word 'Plato' and some non-linguistic entity. Rather, I am saying that the word 'Plato' plays a certain role within a given a conceptual structure, a given language-game; my utterance is analogous to 'This bottle cap is a bishop', uttered when we are about to play a game of chess. Similarly, if I say, '"There is a brown desk here" is true', I am not saying that a certain relation obtains between the sentence 'There is a brown desk here' and some non-linguistic fact. Rather, I am saying that the assertion of 'There is a brown desk here' is a proper move within the language-game currently played.

This means that, for Rorty's Sellars, sentences such as '"Satan" denotes the Devil' and '"Satan is the ruler of Hell" is true' are in a sense perfectly correct, since all they aspire to do is to describe relations or proprieties within a language-game (assuming that there *are* coherent language-games in which the described relations and proprieties do exist). As Rorty puts Sellars's view:

> There is no ground-floor level of truth where we have direct confrontation between reality and thought or language unmediated by justifying assertions. Rather, truth is by coherence. There once was played a coherent language-game in which it was legal to say there are devils, and in that language-game we now play in English it is true to say there are brown desks. But science may some day replace this language-game by

another. (Rorty 1971, 68)

And now the problem arises: How can this "replacement" of one language game by another be in some sense rational? How can it constitute a genuine cognitive achievement? Why is it less arbitrary than a replacement of, say, soccer by hockey?

According to Rorty, it is at this point that Sellars brings in the idea that one language-game can be more *adequate* than another. Explicating this notion of adequacy is central to Sellars's whole project, but it is quite difficult to understand how it can be done. Given the definition of truth in terms of assertability within a framework or language-game, it seems that Sellars must provide us with an explication of "more adequate" which, as Rorty puts it, "does not collapse into the senseless 'more true' or the question-begging 'containing more truths than'. If the fundamental sense of 'true' is language-game-relative, we cannot use possession of more truths as a mark of the superiority of one language-game over another." (Rorty 1971, 68)

Rorty says Sellars recognizes very clearly the difficulty of finding some notion other than his notion of truth, in terms of which he can explain what it is for one language-game to be more adequate than another. The notion Sellars does use for this purpose is that of *picturing*, which is supposed to "belong in a quite different box from the concepts of denotation and truth." (Sellars 1968, 135) This notion is very complex and hard to understand. Sellars makes an analogy with maps: one map may perhaps be more adequate than another, even if the less adequate map does not contain any straightforwardly *false* information. But this analogy is of little help, for in the case of maps we do take ourselves to have access to the common reality that is getting mapped, and to have such access independently of the maps whose adequacy we are comparing. Whereas the problem we confront in Sellars's case is precisely that we cannot take any language-game-independent access to reality for granted.

As Rorty puts it, Sellars's notion of picturing

> just seems to postpone the problem: we now want to know in what parameters we are to
>
> describe these objects common to language-games in order to see that they have been
>
> better mapped. This looks like an unanswerable riddle, since it is equivalent to asking
>
> how the thing-in-itself can be known—how objects can be described which are no more
>
> the objects of some given language-game than of any other. What Sellars needs here is a
>
> vocabulary which is common to all possible language-games, and which is suitable for
>
> formulating criteria of adequacy of mapping. (Rorty 1971, 69)

As Rorty reads him, Sellars bites the bullet that is pointed out in the last sentence of the just quoted passage. Rorty's Sellars purports to offer a vocabulary that *is* common to all possible language-games – namely, one that is derived from "the purely formal aspects of logical syntax" in such a way that it allows us to speak in abstraction from "those features which differentiate specific conceptual structures, and enables us to form the concept of a domain of objects which are pictured in one way (less adequate) by one linguistic system, and in another way (more adequately) by another." (Sellars 1968, 140; quoted in Rorty 1971, 69) The problem with Sellars's account, according to Rorty, is that it is much to brief to be convincing. Sellars never provides any account of the "formal aspects" that comes even close to indicating that they can do the required work. All he gives us at this crucial juncture are mere gestures toward "the logical or 'formal' criteria of individuality which apply to any descriptive conceptual framework" and "the logical criteria which differentiate, say *n*-adic from *m*-adic predicates generally." (Sellars 1968, 139-40) Hence, Rorty says, "[u]ntil Sellars fills out this passage [...] we must say that his project of giving a sense to 'picturing' remains up in the air. [...] I conclude that Sellars has not, in this book, solved the main problem he sets

himself." (Rorty 1971, 69)

In the passage I quoted earlier from his review of Ayer's *The Origins of Pragmatism*, Rorty said that "writers like Sellars (in his recent *Science and Metaphysics*) have made good use of Peirce's notion of the ultimately adequate conceptual framework as a regulative ideal." In fact, however, Rorty does not think that Sellars's use of such a notion is good enough to carry conviction. Indeed, at this time Rorty was rapidly steering toward the view that the dream of making any sense of such a notion of "adequacy", or of any other philosophically powerful notion of language-game-transcendent correctness or truth, was hopeless – a vain attempt to answer a genuinely unanswerable (and ultimately disposable) riddle. It is, however, interesting to see how much his way of conceiving and struggling with the problem owes to Sellars. Let us look a bit closer at how this struggle is manifested in some of his writings.

2. Rorty: From Eliminativism to the Pointlessness of Philosophical Jargon

I will focus on four of Rorty's pieces from the relevant period: "Mind-Body Identity, Privacy and Categories" (1965; Rorty's first and most famous defense of what came to called "eliminative materialism" in the philosophy of mind); "In Defense of Eliminative Materialism (1970; Rorty's reply to criticisms of the 1965 paper); "Realism and Reference" (1976); and chapter 2 of *Philosophy and the Mirror of Nature* (1979; here, Rorty rejects his earlier eliminativist view in favor of what he calls "materialism without mind-body identity").

We saw Sellars discussing whether common sense talk of colored objects may be replaced by scientific talk of light-waves and micro-particles, and what such replacement may mean. Rorty raises a similar question: Can ordinary talk of sensations be replaced by scientific talk of brain-processes, and, if so, what would such replacement involve? In "Mind-Body Identity, Privacy, and Categories", Rorty defends the claim that such replacement is possible. In defending this claim, he takes himself to be defending the central thesis of the so-

called mind-body identity theory. More precisely, Rorty thinks the thesis common to mind-body identity theories of any sort is this: It makes sense to think that empirical inquiry might lead to the discovery that sensations are nothing but brain-processes. This characterization may seem idiosyncratic; for, as Rorty notes, "a proponent of the Identity Theory is usually thought of as one who predicts that empirical inquiry *will* reach this result[.]" However, he continues, "few philosophers in fact stick their necks out in this way. The issue is not the truth of the prediction, but whether such a prediction makes sense." (Rorty 1965, 25 n. 1)

This by itself suggests that Rorty's basic interest in this paper is in fact not the nature of the mind. Indeed, I would go as far as saying that he does not care very much at all about whether sensations are in fact brain processes or not. Rather, as he puts it in the very first sentence of the paper, his interest is first and foremost methodological: controversies over mind-body identity matter to him primarily because they "form a case study for the investigation of the methods practiced by linguistic philosophers." (Rorty 1965, 25) It is these methods, and the general view of philosophy and conceptual change to which they are congenial, that constitute Rorty's wider and primary target in this and other papers ostensibly dealing with narrower issues within the philosophy of mind.

The "linguistic philosophers" Rorty is talking about had argued that any version of the mind-body identity theory must involve category mistakes. For if sensations are identical with brain processes, then, according to the law of strict identity, it would seem that predicates applicable to sensations must be applicable also to brain processes, and vice versa. If so, the identity theorist would have to allow that certain brain processes are dim or fading or nagging, and that after-images are publicly observable, physical and spatially located; and that seems absurd. (See, e.g., Cornman 1962) In response to this objection, identity theorists such as Smart claimed that sensation terms can be translated into a "topic neutral" language and thereby shown not to be subject to categorial conflation of the abovementioned sort. It soon

turned out, however, that this sort of response runs into severe difficulties.

So Rorty instead tries to handle the category mistake charge by developing a version of the identity theory where the identity claim in question is not understood in terms of strict identity. He calls this version the *disappearance* form of the identity theory, since it claims that what science might discover is that sensations *do not exist*. More precisely, Rorty argues that the relation of identity is "the sort of relation which obtains between, to put it crudely, existent entities and non-existent entities when reference to the latter once served (some of) the purposes presently served by reference to some of the former[.]" (Rorty 1965, 26) As examples of statements involving such a relation, he gives the following:

Caloric fluid is nothing but kinetic energy of molecules.

Zeus's thunderbolts are nothing but discharges of static electricity.

Demoniacal possession is nothing but a form of hallucinatory psychosis.

Less crudely – avoiding talk of relations between non-existent and existent entities – Rorty thinks such statements should be conceived as elliptical for:

What people used to call 'caloric fluid' is mean kinetic energy of molecules.

What people used to call 'Zeus's thunderbolts' are discharges of kinetic energy.

What people used to call 'demoniacal possession' is a form of hallucinatory psychosis.

In *these* statements, Rorty says, the relation of identity *is* strict identity; but "[s]ince there is

no reason why 'what people call "X"' should be in the same 'category' […] as 'X', there is no need to claim […] that topic-neutral translations of statements using 'X' are possible." (Rorty 1965, 27-28)

All these examples seem to refer to a scientific discovery, and we are inclined to say that the discovery was a discovery to the effect that certain entities do not in fact exist. It turned out that there is no caloric fluid, but only mean kinetic energy of molecules; Zeus's thunderbolts were just a myth explicable in terms of discharges of kinetic energy; demoniacal possession is a mere superstition – people who were allegedly "possessed" were in fact undergoing a form of hallucinatory psychosis. Rorty is defending the view that science may lead us to adopt a similar view of sensations. According to the disappearance form of the identity theory, it makes sense to think that science may one day lead us to conclude that what people used to call 'sensations' are just brain-processes, and that this conclusion may be seen as entailing that sensations do not exist.

At first sight, this might seem like a straightforward error theory about ordinary sensation talk. If this interpretation were correct, it would in fact be difficult to understand why Rorty would want to call his theory a form of the *identity* theory. It would also be difficult to see any interesting connection with the Sellarsian problematic described in the previous section. In fact, however, Rorty's eliminativism is not meant to be a mere error theory, and is indeed quite closely related to the Sellarsian problematic. For, just as Sellars thinks it is absurd to say that ordinary color ascriptions are incorrect, Rorty is unwilling to say that people who report sensations might be saying something that is straightforwardly mistaken: "we do not wish to say that people who have reported sensations in the past have (necessarily) any empirically disconfirmed beliefs. [...] a term may cease to have a referring use without those who made such a use being convicted of having held false beliefs." (Rorty 1965, 33) And even if Rorty claims that scientific developments would eventually lead us to

conclude that sensations do not exist, he is nonetheless prepared to say that, as sensation words are currently used in ordinary language, those words *do* denote something. At the same time, it is clear that he does not want to embrace any form of reductionism, according to which sensation talk is somehow translatable into talk about brain processes. Thus, he may seem to be trying to occupy a quite awkward or even straightforwardly incoherent position. It is as if he wants both to have his cake and eat it: On the one hand, he does not want to say that ordinary sensation talk is false or without subject matter. On the other hand, he does want to say that future science may show that sensations do not exist at all. As Cornman puts it, Rorty seems to be arguing, paradoxically, that "'sensation' denotes, but […] what it denotes are brain-processes rather than sensations[.]" (Cornman 1968, 48; cf. also Lycan and Pappas 1972)

The affinity with Sellars should be clear. In order to avoid such incoherence it seems Rorty must be working with framework-relative notions of truth and denotation. "Within" the practice of present-day, ordinary talk of sensations, it is indeed true to say that there are pains, tickles and after-images – whereas from the vantage point of future science it may be true to say that pains, tickles and after-images were just myths. But then, can Rorty avoid the seemingly relativist conclusions that seem to threaten us here? Is there any attempt on Rorty's part to provide us with something like Sellars's framework-transcendent notion of "adequacy"?

Well; there is, or appears to be, something like that going on in the text. According to Rorty, the imagined scientific development, ending in the denouncement of sensations as mere myths, is analogous to the following case. We can imagine a tribe where it is held that illnesses are caused by various demons. If a member of the tribe is ill, his friends or family take him to a witch-doctor. After a meal of sacred mushrooms, the witch-doctor says he sees a demon close to the patients body – and, depending on what demon he sees, he recommends a

certain sure. When the patient has pneumonia, witch-doctors report that there is a blue demon with a big nose close to his body; when a patient has diabetes, the demon in question is yellow with black hair; in cases of epileptics, the demon is green with a tail; and so on and so forth. We can imagine that the practice is surprisingly successful: the various treatments recommended usually make the demons go away and the patient recovers.

According to Rorty, if we encountered this tribe, we would be inclined to say that these demons are mere hallucinations, and that the illnesses have quite different causes: germs, viruses, and so on. However, if the practice of the tribe is successful enough, Rorty says there would be no strictly *empirical* criteria that this practice fails to satisfy. We can imagine that the predictions made by witch-doctors mostly turn out to be true. Moreover, the doctors claim to have direct observational evidence that demons exist. According to Rorty, a sophisticated witch-doctor may argue, against our "eliminativist" inclinations, that all modern science has shown is that there is a constant correlation between the presence of demons and the presence of germs, viruses, and so on, and that eating the sacred mushrooms *sometimes* causes hallucinations (but this second point is presumably something already known by the witch-doctors). In other words, such a witch-doctor would claim that the difference between a theory which makes no reference to demons and a theory which allows the existence of demons (together with germs, viruses and so son) is only a matter of simplicity. Demon eliminativists are able to provide an account of the cause and cure of diseases simply in terms of germs, viruses, and so on. The witch-doctor, after having studied moderns medicine, will agree with us that there are germs and viruses, but he will claim that in addition to these things there are also the demons he has repeatedly observed in his own practice.

According to Rorty, this sort of defense against our elimination of demons should be met simply by saying that "the simplicity of the accounts which can be offered if we forget about demons *is* an excellent reason for saying that there are no demons." (Rorty 1965, 29) To

be sure, we might tack the demon theory on to modern science, holding that "diseases are caused by the compresence of demons and germs (each being a necessary, but neither a sufficient, condition), and, second, that the witch-doctors (unlike drunkyards and psychotics) really do see intangible beings (about whom, alas, nothing is known save the visual appearances)." (Rorty 1965, 29-30) The resulting compound would not be more vulnerable to straightforward empirical falsification than our purely scientific theory. Rather, the disadvantage is that we would "be burdened with problems which we did not have before: the problem of why demons are visible only to witch-doctors, and the problem of why germs cannot cause diseases all by themselves. We avoid both problems by saying that demons do not exist." (Rorty 1965, 30)

Rorty suggests that, once a sufficiently sophisticated neural science has been developed, an analogous argument can be made in favor of saying that sensations do not exist. With respect to demons, they add no explanatory power to modern medicine, and we can give a good account of what the witch-doctor is reporting when he claims to be reporting on the existence of demons: he is in fact talking about mere hallucinations. Similarly, Rorty suggests, neural science may become so sophisticated that (1) any explanatory role played by sensation talk can instead be played by talk of brain-processes, and (2) it becomes possible to give a good account of what people are reporting when they claim to be reporting sensations: they are in fact reporting the occurrences of particular brain-states. In such a situation, Rorty argues, the neurologist would be in the same position in relation to ordinary talk of sensations as we are in relation to the imagined tribe. Just as it would be to our advantage to say that demons do not exist, the best thing for the neurologist to say would be that there are no sensations.

Much of Rorty's paper is devoted to answering a natural objection to this parallel, namely, that while it is clear what it is to mistakenly think that one is observing a real demon

rather than an hallucination, it is fundamentally unclear what it is to mistakenly think that one is observing a sensation rather than the occurrence of a brain process. This, however, is a topic that I will leave aside. What I will focus on is rather the following question. In his discussion of the parallel between demons and sensations, it may seem as if Rorty is working with criteria that is supposed to fulfill a purpose similar to Sellars's notions of "adequacy" and "picturing" – that is, criteria in terms of which everyday talk of sensations and scientific talk of brain processes can be compared and assessed. Is this impression correct? And, if so, what is this common measure supposed to be? And are there any problems with it?

The impression that Rorty is working with such "vocabulary-transcendent" criteria of assessment is due to passages such as the following:

"There are no demons" and "What people call 'sensations' are nothing but brain processes" can both equally well be paraphrased as "Elimination of the referring use of the expression in question ('demon,' sensation') from our language would leave our ability to describe and predict undiminished." (Rorty 1965, 31-32)

The inconvenience of ceasing to talk about sensations would be so great that only a fanatical materialist would think it worth the trouble to cease referring to sensations. If the Identity Theorist is taken to be predicting that some say "sensation," "pain," "mental image," and the like will drop out of our vocabulary, he is almost certainly wrong. But if he is saying simply that, at no greater cost than inconvenient linguistic reform, we *could* drop such terms, he is entirely justified. (Rorty 1965, 37; orginal emphasis)

The question is about what Rorty means by "*could*" here. (In another place he even uses the locution "could *in principle*" (34; original emphasis).) Passages like the first one suggests that

he takes the notions of descriptive and explanatory power to be the key ones here, and, hence, that these are the notions in terms of which vocabularies can be compared and assessed. So, the idea would be that irrespectively how "inconvenient" it would be to replace sensation talk with talk of brain processes in real life, the replacement may still be *possible* in the sense that neural science may develop in such a way that it could in a simpler way accommodate all the descriptive and explanatory power of ordinary talk of sensations (plus much more, of course).

So, it may seem as if Rorty wants to employ the notions of simplicity and of descriptive and explanatory power to a job similar to Sellars's notion of picturing. It requires little thought, however, to realize that this is deeply problematic, given Rorty's Sellarsian wish to steer a middle course between the Scylla of reductionism and the Charybdis of a straightforward error theory about everyday sensation talk. Remember Sellars's attempt to steer such a similar middle course in the case of talk of colored objects. He did so by making the notions of truth and denotation relative to frameworks or language-games. Then, in order to avoid the threatening relativism, he introduced the notion of adequacy of frameworks in terms of picturing. And here it was of course crucial that the notion of picturing belonged "in a quite different box from the concepts of denotation and truth". If Rorty wants to pull off a similar trick, it is difficult to understand how the notions of descriptive and explanatory power could be of much help – for they would seem to be precisely in the *same* ballpark as denotation and truth. After all, it is difficult to understand what it would mean to talk about the descriptive power of a given vocabulary, if we weren't also allowed to talk, *from the same viewpoint*, about the referents of its terms and the truth-values of its sentences. So, if Rorty thinks he can use the notions of descriptive and explanatory power as "vocabulary-transcendent" measures of the adequacy of language-games, it would seem that he must allow that the notions of truth and denotation have a similar, "transcendent" function. And then it seems impossible to keep his view from collapsing either into reductionism or into a

straightforward error theory about sensations. For then he must assume that we can meaningfully ask, from "outside" any particular framework: Does our everyday talk of sensations *refer* to anything? And Rorty would then seem obliged to answer *either* "Yes, it refers to brain processes", in which case his view would amount to a form of reductionism; *or*, "No, it does not refer to anything – there are no sensations", in which case his view would amount to an error theory (which is similar to what Lycan and Pappas call "Strong Eliminative Materialism" (Lycan and Pappas 1972).)

Another problem is that, to the extent that Rorty thinks notions of descriptive and explanatory power are available that are independent of particular vocabularies, it remains for him to explain these notions without falling prey to some variety of the sort of Myth Sellars wanted to reject in his criticism of the "positivistic" conception of science. For the questions immediately arise: *What* is the vocabulary-independent measure of descriptive and explanatory power? And it seems hard to answer this question without introducing something like independently given "data"; and this is, of course, precisely what Sellars and Rorty want to avoid.

I think these are matters on which Rorty was far from clear in 1965. Other passages in his paper suggests that a quite different reading is more appropriate, according to which Rorty's notions of descriptive and explanatory power are not at all meant to be vocabulary-transcendent in the just described sense. In fact, this was an issue on which people asked him for clarification. Thus, already in 1968, Richard Bernstein described a tension in Rorty's eliminativism, leading up to an interesting dilemma:

On the one hand [...] Rorty [...] seems to be presupposing a metalanguage or metatheory in which we can evaluate different types of descriptive expressions and determine whether our ability to describe is or is not diminished. But on the other hand, Rorty

sometimes writes as if the radical displacement of languages takes place without any inter-theoretical justification. Like Marx's concept of the state, one form of discourse withers away (in fact or in principle) when it no longer serves any function or purpose that isn't better performed by another mode of discourse. But then it is no longer clear what it means to say that our ability to describe is undiminished. (Bernstein 1968, p. 272-273)

In "In Defense of Eliminative Materialism", Rorty makes it clear that his notions of descriptive and explanatory power are not vocabulary-transcendent in the strong sense Bernstein is suggesting in the first half of the just quoted passage. Rorty's response is intriguing, and I will quote it at length:

To say that our ability to describe is undiminished is merely to say that by using some portion of language common to the competing vocabularies (e.g., "What do you experience when I do *that* to your arm?") we can isolate the questions to which alternative answers might be given and note that both vocabularies offer something to say in reply. No general metalanguage is needed, but merely some way of locating the place in the language-game which is to be filled by either of the alternative candidates. I quite agree with Bernstein's implicit suggestion that any general metalanguage or metatheory would be question-begging, and in particular any which always awarded the prize to the "scientific" alternative would be. Therefore I grasp the second ("withering away") horn of the dilemma he sketches. But I take no sides on the question of whether the materialist is right in his prediction that the ordinary ways of reporting on introspections will wither away. In my view, the truth of the prediction is of much less philosophical interest than the fact that the prediction is itself a coherent suggestion.

(Rorty 1970a, 120)

This long passage should be juxtaposed with another passage that occurs only one page earlier and which, Rorty says, makes the same point:

> I am not in any sense claiming that the customary vocabulary of introspection is "illegitimate." Rather, I am merely claiming the same legitimacy for the neurological vocabulary – where "legitimacy" means the right to be considered a report of experience. My attitude is not that some vocabularies are "illegitimate," but rather we should let a thousand vocabularies bloom and then see which survive. The materialist predicts that the neurological vocabulary will triumph. He may be right, but if he is, it is not because of some special feature of this vocabulary which consists in its having originated in theoretical science. Given different cultural conditions, one can imagine the neurological vocabulary having been the ordinary familiar one and the mentalistic one the "scientific" alternative. (Rorty 1970a, 119)

It is difficult to know what to make of these passages. It is very clear that Rorty rejects the idea of a meta-language in which notions such as descriptive and explanatory power serve as neutral measurements of the relative adequacy of language-games or vocabularies. The rhetoric of the second passage sounds very tolerant: let a thousand vocabularies bloom, and if the neurological vocabulary triumphs over vocabulary of sensations the materialist's prediction happened to be true. But our question is precisely what this possible "triumph" is supposed to consist in; and on this issue, Rorty is still quite obscure.

The first passage makes it sound as if the triumph may be something else than a matter of who is most powerful or who can provide the most enticing rhetoric. More precisely, Rorty

seems to be striving, like Sellars, to construe this triumph as a matter of a development that is in some sense justified, not in terms of independent criteria given in a meta-language, but in terms of criteria accessible from *within* that development. The suggestion seems to be that the two language-games or vocabularies have enough in common to allow a comparison between introspective and neurological accounts of what it is that is experienced, for example, when someone pricks my arm with a needle. A user of the vocabulary of introspection answers "Pain!", whereas a user of the imagined vocabulary of future neural science answers "The firing of my C-fibers!". And it might seem to be Rorty's view that, according to common criteria of simplicity and descriptive and explanatory power, a decision can somehow be made in favor of the latter answer.

However, it is still unclear how this is supposed to work, For nothing in Rorty's account explains why an adherent to the vocabulary of introspection should not say that his vocabulary has more descriptive power than a vocabulary that does not refer to sensations. After all, from within the vocabulary of introspection, Rorty thinks it is *correct* to say that there are sensations. Hence, from this viewpoint, it would indeed be correct to say that replacing the vocabulary of introspection by a purely neural vocabulary is to *lose sight* of a whole domain of objects: pains, tickles, after-images, and so on. In other words, Rorty must admit that from this viewpoint, it will be *correct* to say that the descriptive power of the purely neural vocabulary is much *less* than the descriptive power of a vocabulary that made reference to sensations in addition to brain processes. Of course, this is not what it will look like once the replacement has been done. *Then* it will be correct to say that sensation talk does not, in fact, refer to anything, and that all that people were reporting when they said things like 'I'm in pain' and 'I have a green after-image' were the occurrences of certain brain-processes. But what we want in order to solve the present conundrum is a notion of reference which can serve to rationally persuade the adherent of the vocabulary of introspection of the

mythical status of sensations *prior* to his conversion to a the purely neural language-game. And it remains difficult to understand how Rorty's account can leave room for any such notion of reference.

I suggest that in the years around 1970, Rorty gradually came to realize that his conception is indeed incompatible with the idea that there is a notion of reference that can play this sort of role. More precisely, he came to realize that the only notion of reference he can allow that can motivate the claim that sensations do not exist is a notion that is applicable *in retrospect*, while looking back at the vocabulary of introspection from the viewpoint of the neural vocabulary after the latter has replaced the former. This idea is in fact hinted at already in his review of Sellars's *Science and Metaphysics*. Right after having noticed that Sellars's conception of adequacy in terms of picturing does not seem to work, Rorty goes on by suggesting the following couple of alternatives:

> Perhaps [Sellars] should just say that the mere notion of "better picturing" is sufficient to give us the notion of "more adequate" (and thus of truth by correspondence, rather than intra-structural truth by coherence) even if we do *not* have a vocabulary in terms of which we can isolate the objects pictured in a way that is neutral between language-games. Or perhaps he should just construe "more adequate" in terms of the familiar though complex criteria by reference to which we now say that our science is more adequate than Greek science. (Rorty 1971, 69)

It is clear, I think, that the second of the two alternatives is the one Rorty himself is going to develop and embrace in the 70's. Further clarity on this matter is reached, first in "Realism and Reference" and then in *Philosophy and the Mirror of Nature*.

In "Realism and Reference", Rorty distinguishes between three notion of referring. The

first sense – 'reference$_1$' – is equivalent to the ordinary notion of 'talking about'. This notion of reference is quite similar to Sellars's notion of 'denoting', as Rorty understands it. In this sense of 'reference', people can refer to things that do not exist. Thus, in the past, people referred$_1$ to Satan, Hell, caloric fluid, Zeus' thunderbolts and demoniacal possessions. And we can imagine future neurologists saying that in the 20th century, people referred$_1$ to sensations, even if sensations do not in fact exist.

The second notion of reference – 'reference$_2$' – is, by contrast, one by means of which we can retroactively say things such as

There are no demoniacal possessions; what people were referring$_2$ to by 'demoniacal possession' is a form of hallucinatory psychosis.

Zeus' thunderbolts do not exist; what people were referring to by 'Zeus' thunderbolts' were discharges of static electricity.

Importantly, these statements do not entail that demoniacal possession is strictly identical with hallucinatory psychosis, and that Zeus' thunderbolts are strictly identical with discharges of static electricity. The truth of 'Jane is undergoing a hallucinatory psychosis' does not entail the truth of 'Jane is undergoing a demoniacal possession'. As Rorty puts it, "the truth of one's remarks is not determined by the discovery of what one is talking about. Rather, the subject is changed." (Rorty 1976, 325)

Reference$_1$ and reference$_2$ are both different from the philosopher's favorite notion of reference – what Rorty calls 'reference$_3$'. Like with reference$_2$, you can only refer$_3$ to what really exists. Unlike reference$_2$, however, reference$_3$ does not allow you to say *both* that what people were referring to by 'Zeus' thunderbolts' are in fact discharges of static electricity *and*

that Zeus' thunderbolts do not exist. For it is constitutive of reference$_3$ that if 'Zeus' thunderbolts do not exist' is true, then there is nothing to which 'Zeus' thunderbolts' refers$_3$. In particular, if 'Zeus' thunderbolts do not exist' is true, 'Zeus' thunderbolts' cannot refer to discharges of static electricity.

It is precisely because we tend to assume that 'reference' must mean reference$_3$ that it seems as if Rorty *has* to choose between a reductionist view according to which sensations *do* exist but are reducible to brain-processes, and a strongly eliminativist or error-theoretical view according to which people who employ the vocabulary of sensations simply fail to refer to anything at all. Rorty agrees that if we accept that the notion of reference$_3$ is central to the sort of problem he is interested in, his attempt to steer between reductionism and strong eliminativism must indeed fail. The conclusion he draws, however, is not that this attempt was misguided. Rather, he thinks it is the assumption that reference$_3$ is relevant here that needs to be questioned:

> As Pappas and Lycan correctly point out [...] there is no way in which one can make sense of the difference between [my] sort of "eliminative" materialism and "reductive" materialism" (the sort which depends upon "topic-neutral translations" of mentalistic terms) if "talking about" is construed as reference$_3$. Unlike Pappas and Lycan, I would draw the moral "so much the worse for reference$_3$." (Rorty 1976, 338 n. 5)

According to Rorty, reference$_1$ and reference$_2$ are the only notions of reference we need in order to handle the problem of ancestral error. Indeed, Rorty thinks it is precisely the notion of reference$_3$ that creates specifically *philosophical* problems for us here, as opposed to concrete historiographical ones. In the absence of plausible and non-question-begging reductions of terms such as 'demoniacal possession', 'caloric fluid' and 'phlogiston', thinking

of reference as reference$_3$ forces us to say that such terms did not refer at all, and that, in using such terms, our ancestors were therefore out of touch with the world. But then, how can we describe, say, the development of phlogiston theory as an instance of genuine scientific progress? And how can we do justice to the fact that our ancestors were no less careful and judicious than we are? Indeed, how can we even be sure that our own terms refer to anything? Why assume that we happen to be so fortunate, if our ancestors were not? We may try to handle such difficulties by distinguishing between observation and theory, and argue that it was only the theoretical concepts of our ancestors that failed to denote, whereas the observational part of their language – a part they have in common with our language – referred to real entities. But this, of course, is a step Rorty urges us to resist. Indeed, he would see this temptation to fall back on the Myth of the Given as a clear indication that it is a mistake to think that reference$_3$ can help us get clear about scientific development and ancestral error. Reeference$_3$ might have a legitimate use in technical semantics, but it only creates confusion if imported into epistemological contexts.

So, Rorty concludes that if one's aim is to understand the history of inquiry, the best things to do is to be satisfied with reference$_1$ and reference$_2$. Certainly, these two notions of reference offers different perspectives on what our ancestors were doing. We can say, either that "our ancestors referred$_2$ to what we do, but didn't know it and hence said mostly false things about it", or, alternatively, use "the more sophisticated don't-let's-be-beastly-to-other-conceptual-frameworks view" that "our inquiring ancestors were referring$_1$ to things that didn't exist, but mostly speaking truths about those fictitious things." (Rorty 1976, 335) The important point is that as long as reference$_3$ is kept out of the picture, there will be no need to find univocal rules for when to say one of these things rather than the other. The longer the distance between us and the ancestors whose views we are trying to capture, the more inclined we will be to choose the second type of description; the closer we get to present science the

more the first type of description will take precedence:

> [I]t would be pointless to have said to Aristotle "There is no such thing as natural motion; what you're talking about is gravity," but it would be in point to say this, circa 1680, to someone who had patched up a confused synthesis of Aristotelian physics and Galilean mechanics. The difference is, roughly, that there are very few statements of Aristotle's in which "gravity" can be substituted for "natural motion" to produce a truth, whereas in the imagined synthesis there would be many. (Rorty 1976, 326)

It seems clear, however, that there will be many cases in which we can just as plausibly say one thing as the other, and that our choice will depend on the specific historiographical purposes of our description. The aptness of these various descriptions is a matter of degree and of context, and no philosophical theory of reference$_3$ can help us in the making of such choices.

I said earlier that reference$_1$ is similar to Sellars's notion of denotation, as Rorty understands it. What Rorty gives us instead of Sellars's conception of adequacy is reference$_2$. That is, he does precisely what he suggested Sellars might have done instead of construing adequacy in terms of picturing: he construes it in terms of the familiar though complex criteria by reference to which we now say that our science is more adequate than that of our ancestors. And in doing so, he of course gives up the heroic Sellarsian project of finding something over and above the time-bound, messy, ordinary criteria of adequacy that are employed by historians of science. One might say that Rorty gives up the idea of a special task for philosophy in dealing with ancestral error, in favor of what he calls a *pragmatical* conception:

On this relaxed and unphilosophical view, the problems raised by our ancestors' errors dissolve. […] We are bound to treat our present views on nature and morals as true, for we know no better. But the invidious distinctions we draw between ourselves and the Trobrianders, or between our chemists and those who believed in phlogiston, are to be backed up in the detailed and humdrum ways in which we explain the advantages of the rule of law, or of thermodynamics, over any alternatives so far canvassed. There is nothing particularly *philosophical* to be said. (Rorty 1976, 323)

In this passage, Rorty formulates what is essentially the view presented in more detail in chapter 2 of *Philosophy and the Mirror of Nature*. In that chapter, he is even clearer on the following issue: The point he tried to make in his earlier papers by presenting what seemed to him like a defense of a philosophical view called "the disappearance form of the mind-body identity theory", is much better made by not defending any positive philosophical view at all. What he was hazily making his was toward in the 60's and 70's was rather the insight that in attempts to understand ancestral error and the nature of vocabulary replacement, philosophical talk of reference, identity, ontological status, and so forth, is just an obstacle. All we need here is the detailed and hard work of the careful historian; there is no point in pressing him to formulate his conclusions in philosophical jargon.

In chapter 2 of *Philosophy and the Mirror of Nature*, Rorty does not employ the distinctions between reference$_1$, reference$_2$ and reference$_3$. Presumably he thinks that this classification still looks too much like a debatable philosophical schematism. Instead, his famous overall strategy is to describe in detail an imagined civilization – the Antipodeans – living on a planet far away from the Earth. The biology of the Antipodeans is very similar to that of Earthlings. However, Antipodeans do not talk about sensations. Rather, due to their impressive development of neurology and biochemistry, they talk of brain processes. Thus,

antipodean mothers cry out 'He'll stimulate his C-fibers!' when their children approach hot stoves. When people look at clever visual illusions they say things such as 'How odd! It makes neuronic bundle G-14 quiver, but when I look at it from the side I can see that it's not a red rectangle at all'. And so on and so forth. (Rorty 1979, 71.) Otherwise, the Antipodean culture is much like ours. They have poetry, science, religion, and some even believe in immortality (albeit as a straightforward matter of bodily resurrection). Rorty's point is precisely to show how marginal the difference between us and the Antipodeans is, and that it is only the philosophical worries about reference and existence that can make this difference seem to be of great – indeed monumental – significance.

To make this point vivid, Rorty imagines that a human expedition of scientists lands among the Antipodeans and starts interacting with them. According to Rorty, such interaction will create no special problems – except among the philosophers. Among the philosophers – and only there – seemingly unsolvable difficulties will arise. Do the Antipodeans really have sensations or not? Do they have them without knowing it? Or do they know about their sensations, even if they keep quite about them? Or do they in fact talk about sensations, since sensations are identical with the brain processes they are talking about? And so on. Rorty goes through various ways in which the philosophers may try to decide these issues – for example, in terms of the distinction between corrigible and incorrigible reports – but concludes that there is no non-question-begging way of reaching agreement. The conclusion he wants us to draw from this is not that the issue of whether Antipodeans have sensations is important but ineffable, but that it is an artificial product of philosophical terminology.

As a corollary,

the reductive and eliminative versions of the identity theory are both merely awkward attempts to throw into current philosophical jargon our natural reaction to an encounter

with the Antipodeans. I do not think that the difference between the two should be pressed. Rather, they should both be abandoned, and with them the notion of "mind-body identity." [...] It is pointless to ask whether the fact that cerebroscopes correct Antipodean reports of inner states shows that they are not *mental* states, or shows rather that mental states are really neural states. It is pointless not just because nobody has any idea how to resolve the issue, but because nothing turns on it. (Rorty 1979, 120)

3. Conclusion: Vocabulary Replacement and the Elimination of Philosophical Problems

In the Introduction to this paper, I said that one of its central aims was to show how Rorty's early investigations into mind-body identity and eliminative materialism display how he is led to the methodological view characteristic of his mature philosophy – a view which includes as its central conception that of *vocabulary replacement*. What I have said in the previous section may seem to run counter to this aim, however. For isn't the upshot of Rorty's view, as I have described it, precisely that philosophy has *no* right to reject one vocabulary in favor of another – that there are no specifically philosophical reasons to change our ways of speaking?

In his discussion of Rorty's writings in the philosophy of mind, David Hiley seems to argue precisely for such a reading. Hiley rightly notes that "for Rorty, [...] the philosopher can neither determine which is the most adequate description of what there is, nor prescribe which ought to be. The former view is wedded to a nontrivial sense of 'the world' which conditions language and to which language must be adequate. The latter supposes that philosophy is a sort of superscience that can lay down in advance conditions and limitations on the development of inquiry." (Hiley 1978, 335) From this correct observation, Hiley goes on by concluding that all the philosopher can do is to let a thousand vocabularies bloom, noting that they are all "possible", and then passively wait and see which survives. Strictly thought through, Hiley thinks Rorty's so-called eliminative materialism "simply becomes a nihilistic

claim about descriptive vocabularies." (Hiley 1978, 337)

There is certainly some support for Hiley's reading in the passages from "A Defense of Eliminative Materialism" that I quoted earlier. Still, it is clear from many other passages that Rorty is not nihilistic in this extreme sense. Even if the philosopher abandons the aspirations to determine which vocabulary is the most adequate description of Reality and lay down in advance conditions and limitations on the development of inquiry, Rorty thinks there is still say something to be said in favor of certain vocabularies as compared to others. More precisely, Rorty thinks the philosopher can make clear that abandoning a certain vocabulary – say, one in which one talks about sensations – in favor of another – say, one in which one talks of brain processes – is a good thing to do because certain philosophical puzzles will then *disappear*, in the sense that the verbal resources required to formulate those problems will no longer be in place. Consider again the demon example. If we stop talking of demons and instead talk of germs and viruses, we will no longer be able to ask questions such as 'Where do the demons go when they are not making people ill?', 'Do demons reproduce, and, if so, how?', 'How is it that only witch-doctors can see demons?', 'Why is eating sacred mushrooms necessary for seeing demons?', and so on. That will certainly be a sort of gain. We can lay those worries aside and instead focus on other, presumably more fruitful issues. Similarly, by abandoning the vocabulary of sensations, Rorty thinks we can lay aside troubles about the existence of other minds, the interaction between the mental and the physical, and so forth – troubles that are irresolvable within the vocabulary that allows their formulation. Hence, Hiley is simply wrong in ascribing to Rorty the view that "[t]he most that can be said for eliminative materialism is that alternative vocabularies are possible thus a materialistic vocabulary is possible." (Hiley 1978, 335) According to Rorty, there *is* indeed a "philosophical" sense in which a materialistic vocabulary is better than a mentalistic one. It might not be a very important advantage, since Rorty thinks most people do not care about

philosophical problems anyway. But it is an advantage nonetheless, and one that makes it a quite legitimate task to encourage this sort of vocabulary replacement by showing how it can liberate us from philosophical worries. Such encouragement is a vital part of Rorty's mature way of doing philosophy, and it is self-consciously applied at least from *Philosophy and the Mirror of Nature* and all the way up to his last writings.

One might of course question whether Rortian vocabulary replacement can really be a satisfactory way of getting rid of philosophical problems. I have argued elsewhere that it seems quite unsatisfactory, since a person who is genuinely troubled by a philosophical difficulty can hardly experience a replacement of vocabulary as a solution. Rather, such a person will feel that his problem is simply swept under the rug – that to forget about a problem is not to solve it. It seems that Rorty's form of philosophical revisionism can have its intended effect only on a person who is already fed up with the sorts of questions Rorty wants us to stop asking – a person who already has a sense that the vocabulary Rorty wants us to get rid of is out of date. (Gustafsson 2001, 647)

To substantiate this criticism, however, must be the aim of another paper. My purpose here has just been to describe Rorty's development "from within", and thereby provide a better understanding of the difficulties and trains of thought that led up to that form of provocative form of neo-pragmatism to which he owes his notorious reputation.

References

Brandom, R. B. (ed.) 2000a. *Rorty and His Critics*. Malden, MA: Blackwell.

Brandom, R. B. 2000b. Vocabularies of Pragmatism: Synthesizing Naturalism and Historicism. In Brandom (ed.) 2000a, 156-183.

Cornman, J. 1962. The Identity of Mind and Body. *Journal of Philosophy* 59: 486-492.

Cornman, J. 1968. Mental Terms, Theoretical Terms and Materialism. *Philosophy of Science* 35: 45-63.

Dennett, D. 2000. The Case for Rorts. In Brandom (ed.) 2000a, 91-101.

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

Gross, N. 2008. *Richard Rorty. The Making of an American Philosopher*. Chicago: University of Chicago Press.

Gustafsson, M. 2001. Review of Brandom (ed.), *Rorty and His Critics*. Philosophical Review 110: 645-650.

Hiley, D. R. 1978. Is Eliminative Materialism Materialistic? *Philosophy and Phenomenological Research* 38: 352-337.

Lycan, W. G., Pappas, G. S. 1972. What Is Eliminative Materialism? *Australasian Journal of Philosophy* 50:149-159.

Malachowski, A. 2002. *Richard Rorty*. Princeton: Princeton University Press.

Rorty, R. 1965. Mind-Body Identity, Privacy, and Categories. *Review of Metaphysics* 19: 24-54.

Rorty, R. 1970a. In Defence of Eliminative Materialism. *Review of Metaphysics* 24: 112-121.

Rorty, R. 1970b. Incorrigibility as the Mark of the Mental. *Journal of Philosophy* 67: 399-424.

Rorty, R. 1971. Review of *The Origins of Pragmatism: Studies in the Philosophy of Charles Sanders Peirce and William James*, by A. J. Ayer. *Philosophical Review* 80: 96-100.

Rorty, R. 1972. Functionalism, Machines, and Incorrigibility. *Journal of Philosophy* 69: 203-220.

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

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

Rorty, R. 1982. *Consequences of Pragmatism*. Minneapolis: University of Minnesota Press.

Sellars, W. 1968. *Science and Metaphysics: Variations on Kantian Themes*. London: Routledge and Kegan Paul.

Sellars, W. 1997 [1956]. *Empiricism and the Philosophy of Mind*. Cambridge, MA: Harvard University Press.