Philosophy and Phenomenological Research
Vol. L, No. 1, September 1989

The Myth of Jones and the Mirror of Nature: Reflections on Introspection*

JAY L. GARFIELD
Hampshire College

1. Introduction

No piece of philosophical science fiction in this half of this century has been as significant in its impact on epistemology and the philosophy of mind as the Myth of Jones, the heart of "Empiricism and the Philosophy of Mind" [Sellars, 1956 (reprinted in {Sellars, 1963}, to which all page references refer)] in which Jones, the great apocryphal forebear of cognitive science, and his great discovery — the inner psychological episode — are introduced. Nonetheless, not only have the devastating implications of this myth for the Myth of the Given not always been taken to heart in contemporary epistemology, but considerable confusion reigns concerning the central moral of the story. The authors of no less than four recent

* The idea for this paper came from some remarks John Connolly made in a conversation. I am deeply indebted to him for this and many other useful philosophical conversations on this and related themes. Beyond the obvious debt to Wilfrid Sellars, I also owe special thanks to Meredith Michaels, who read and provided useful comments on a very early draft of the first half of this essay on incredibly short notice, and for providing useful comments on several subsequent drafts. Thanks also to Lynne Rudder Baker, Lee Bowie, Steven Weisler, Willem A. deVries, Dan Lloyd, Harold Skulsky and Raimo Tuomela for detailed and valuable comments on early complete drafts. I also thank the members of the Propositional Attitudes Task Force, particularly Lee Bowie, John Connolly, Murray Kiteley, Dan Lloyd, Meredith Michaels, Ken Taylor and Jonathan Vogel for a useful discussion of this paper and for many profitable and enjoyable philosophical discussions. I received extensive and useful comments leading to substantial revision in a discussion of this paper with the Triangle Philosophy of Mind Group. Particular thanks go to Jay Rosenberg, Timo Airaksinen, Michael Fairjohn, Christopher Hill, Douglass Long, Bill Lycan, David Sanford and Sydney Shoemaker. I have profited enormously from the contributions of all of these friends and colleagues, but I persist in many of my errors despite their counsel.

books ([Churchland, 1979], [Dennett, 1987], [Lycan, 1987], [Garfield, 1988]) in the foundations of cognitive science have noted Sellars' seminal role in the field, and each has claimed him as their own intellectual ancestor. But these authors agree on virtually nothing else: Whereas Lycan and Garfield are both realists with respect to belief (though Lycan is a functionalist and Garfield a naturalist), Dennett is an instrumentalist, and Churchland an eliminativist. This essay is an attempt to correct some of that confusion. In particular, it has been supposed by many (e.g., [Feyerabend, op. cit.] and Churchland, op. cit.]) who have embraced Sellars' attack on the given that the argument contained in the Myth entails that such cognitive states as belief are themselves mythical, or at least dependent for their status on suspect empirical theories concerning human psychology. Other Myth-embracers (notably [Rorty, op. cit.]) have gone so far as to claim that the Myth of Jones requires us to abandon the idea that there is any mental representation at all. I will argue that these claims mistake the Jonesean myth's central morals about the nature of introspective awareness, psychological categories, and privileged access, and that the Myth of the Given can be rejected without such a drastic revision of our self-image.

Central to these influential eliminativist misinterpretations of the Myth of Jones, I will argue, have been two conflations. The first concerns what Rorty has aptly called the idea that the mind is a "mirror of nature." According to this view, which Rorty correctly rejects, it is only to the images in the mirror that we have unproblematic epistemic access. This idea, I will be concerned to argue, is conflated by Rorty and his followers with the idea that our thoughts can be understood as representational. The Myth of Jones requires a rejection of the first idea, but requires not a rejection, but an acceptance, under a particular interpretation, of the second. The second conflation involves confusing spontaneous, sincere belief-assertions with introspective judgments.[2] Rejecting the Myth of the Given in favor of the Myth of Jones requires us to treat introspective judg-

---

[1] This confusion is reflected in some of the most important and influential recent essays in epistemology (esp. [Churchland, 1979] and [Rorty, 1979], but see also [Millikan, 1984], and [Feyerabend, 1981]).

[2] Though this distinction will be drawn and defended much more carefully below, for now it will do to think of sincere belief-assertions as those self-ascriptions wherein the "I believe . . ." could as well be dropped — where at most it contributes to the *illocutionary* force of the assertion — as in "I believe that snow is white." Introspective judgments, by contrast, are those in which the principal topic of the assertion is the doxastic state of the speaker, as in "I never realized it, but I in fact believe that my father wants to kill me. There is no other explanation for my behavior." In such cases, the *locutionary* force of "I believe" is salient.

ments as being in an important sense theoretical; but this does not force a similar treatment of all sincere belief-assertions. Ignoring these two important distinctions, I will argue, involves a Lockean interpretation of Sellars' myth which itself involves a subtle and pernicious reintroduction of the Myth of the Given at the heart of the account.

Before embarking on the substantive part of this investigation, it is important to recall the two philosophical myths at issue. First the Myth of the Given: The Myth of the Given comes in both empiricist and rationalist versions. They have in common a commitment to the view that there is something that we know immediately, without depending upon any process of inference, or on any further knowledge; that there are cognitive states or mental contents of which we are immediately and incorrigibly aware *as the kinds of states or contents* they are; that we have privileged access (in more than a pragmatic sense — in a principled sense) to at least some of our mental states or processes; or finally, that there are self-presenting claims about our inner states which are self-justifying. I take it that part of the importance of the Sellarsian analysis, an analysis I take for granted here, is that these ways of putting the central commitment of the Myth of the Given are at bottom equivalent.[3]

In its rationalist version, the Myth asserts that these privileged contents, these self-presenting states, are beliefs (or in more extreme views, that there is innate self-presenting general knowledge). So for the Cartesian, my beliefs are given to me as beliefs, my doubts as doubts, and moreover, as the very beliefs and doubts they are, individuated by their contents, about which error on my part is inconceivable. My knowledge that I entertain these beliefs and doubts is hence incorrigible, immediate, and self-justifying.

The empiricist version of the Myth involves similar claims regarding sensations or perceptual beliefs. On this version my sensations are given to me as the sensations they are (depending on the specific version the relevant sortal may be phenomenalistic or middle-sized-dry-goodsy). My knowledge that I am in a particular sensory state (or my perceptual belief) enjoys incorrigibility: They are known immediately, and are self-justifying.

---

[3] I should also state at the outset that the Myth of Jones is intended to, and does, strike at both empiricist and rationalist forms of the myth. I will, however, in this essay, be concerned only with the attack on the rationalist version, since my real concern is with the question of the relationship between eliminativism with respect to the propositional attitudes and the Sellarsian enterprise. An absolutely parallel essay could be written on the empiricist side concerned with the status of qualia. This is important, because I will argue below that one form of the Myth of the Given re-introduced by Sellars' Lockean commentators is the empiricist version.

Against this myth, which lies at the heart of much of the European epistemological tradition, Sellars levels a counter-myth, the Myth of Jones, whose interpretation is the focus of this study. In brief, the Myth runs as follows:[4] We are to imagine our fictional Rylean ancestors, who speak a Rylean behavioristic language which allows reference only to public, observable states and properties of persons (including dispositions to behave or to be in observable states). The language includes, however, not only the subjunctive conditional, but also such semantic predicates as "means" (which can, of course, always be paraphrased in the customary Sellarsian dot-quotational way). Now, among these Rylean ancestors, we are to imagine the emergence of a certain genius Jones, who, observing that the intelligent behavior of his fellows can frequently be explained by reference to the acompanying speech, develops a theory according to which *all* intelligent behavior is in fact so guided, even when the speech is not audible — that there is, in the cases of silent intelligent activity, a series of inner episodes that are causally responsible for that behavior in a way analogous to the way in which verbal behavior is responsible for guiding other intelligent behavior. Pursuing the analogy, Jones uses the same semantic categories already in use for describing overt verbal behavior to characterize these hypothesized inner episodes, and so can speak of their *meaning* or *content*. He calls these inner assertions "thoughts." Jonesean psychology, counterintuitive though it may have been (in virtue of the fact that none of Jones' contemporaries would ever admit to having had one of these "thoughts") caught on, because of its tremendous scientific successes. People began regularly explaining their fellows' behavior in terms of these new theoretical entities. By degrees, a strange thing happened: People began to get so good at the inferences to the existence of these entities posited by the theory that they began to explain their own behavior by saying such things as "I think that **p**." Indeed, they got so good at it, that without even observing their own *overt* behavior, they could reliably report their own thoughts with greater accuracy than others could. The language initially introduced as a highly theoretical third-person explanatory device attained a reporting role in which first-person reports, despite apparently more slender evidential bases, had greater authority.

The essential features and morals of the Myth for our purposes are these: (1) While introspective reports are in an important sense observational, they are couched in a vocabulary that is, in an equally important sense, theoretical. (2) The theory through which the vocabulary of thoughts is introduced and in which it at least initially is embedded is an

---

[4] An only slightly abbreviated version of the original text appears as an appendix to this paper.

empirical theory — hence possibly a false one — whose point is to explain human behavior. (3) The model of inner states and processes employed by this theory is a linguistic model, more particularly, a sentential model. (4) The only legitimate sense in which we enjoy privileged access to our thoughts is the pragmatic sense — our self-ascriptions and explanations of our own behavior in terms of our posited inner episodes are more reliable, and less frequently revised than are those generated by third party description and explanation. But in any particular case it is possible for me to be wrong and you right about the contents of an occurrent thought of mine.

In this essay I want to clarify these morals and defend them against those who have adopted the Myth in the service of views I believe are quite alien to the Sellarsian framework, alien because of their misconstrual of the importance of the indispensability of semantic discourse and of the observational role assumed by the vocabulary of thoughts. Ultimately, I will argue that they are alien because of their consequent eliminativism with regard to the mental, which leads to the unnoticed and pernicious adoption of the Myth of the Given — the target of the Myth of Jones. I will begin by discussing the manner in which the Myth of Jones — using a very Cartesian strategy — demolishes the vestiges of Cartesianism represented by the Myth of the Given. Following, I will consider the relationship of the Myth of the Given to the idea that the mind is a "mirror of nature," explaining the role of Jones in deflating that metaphor, but also criticizing Rorty's extension of the Sellarsian attack to an attack on representation in general. This will set the stage for a direct discussion of the misappropriation of Jones in service of the eliminativist views of Churchland, Feyerabend and Rorty. In discussing this misappropriation, I will diagnose their error as consequent upon a covertly Lockean reinterpretation of Sellars' account of introspection. I will conclude by arguing that this Lockean reading of Sellars is unwarranted by the text, and yields an incoherent version of the Myth of Jones — one that embodies the Myth of the Given. I will argue for a non-Lockean analysis of introspection and of belief-report which makes sense of Sellars' account, avoids a commitment to givenness, and does not undermine the reality of the mental.

2. The Devil and the Myth of Jones

Descartes used the device of the epistemically omnipotent evil deceiver to argue that while all of my knowledge of the external world is fallible, my immediate introspective knowledge of my occurrent internal states is not. It is important to see Descartes' revolutionary move as a first step, but emphatically not a final step, in the attack on givenness.

One way of making this point is to note that an important moral — perhaps *the* important moral, from our vantage point — of Descartes' myth is the vast epistemological *difference* between our perceptual knowledge and our introspective knowledge: Whereas claims about the external world rely for their evidential support not only upon our current sensory states and consequent perceptual beliefs, but also upon a *theory* of the causation of these internal perceptual states by external objects, our introspective claims about our own states, Descartes argues, require no such fallible theoretical premises in their justification. To put the point another way: for Descartes, whereas objects in the external world are not given to us directly as the kinds of objects they are, with the sensible and non-sensible properties they bear, our inner states, the objects of introspection, are given to us as the kinds of states they are, and as bearing all of the essential properties they bear.[5]

On this account, the deceiver argument is a double-edged sword. By demonstrating the dubitability and consequent theoretical nature of perceptual knowledge, Descartes convincingly undermines the Myth of the Givenness of the external world. At the same time, the very strategy he uses, by employing and emphasizing the putative contrast between perceptual and introspective knowledge, asserts and defends the givenness of our own internal states.

The Myth of Jones, seen in this (extended) historical context, can be viewed properly as the paradoxically simultaneous completion of the first part of the Cartesian project and the refutation of the second part which arguably motivated the first. Whereas Descartes emphasizes the heterogeneity of perception and introspection, Sellars emphasizes their homogeneity, assimilating, from the standpoint of epistemology, the latter to the former. For a consequence of the claim that thoughts are introduced as theoretical entities is that their legitimacy is vouchsafed only by the truth of the theory that posits them. A further — and no less important — consequence is that our apparently spontaneous and theoretically unencumbered ability to report them is no more evidence for their givenness than is the apparent spontaneity and atheoreticality of our perception of middle-sized dry goods. Finally, the claim that privileged access is a pragmatic and not a metaphysical matter of fact entails that from the standpoint of dubitability, and hence of givenness/inferredness, our own thoughts are no more revealed to us in principle than are the contents of our environ-

[5] To be sure, the historical claims I am making about Descartes' argument are both anachronistic in their formulation, and stand in need of careful textual exegesis for their support, but the anachronism is excusable in the context of this project, and the exegetical work is beyond its scope.

ments. Our thoughts are apperceived using the categories provided by the (at least initially) theoretical vocabulary of inner episodes, which provides the descriptive and explanatory conceptual framework in terms of which we represent to ourselves our inner life. Introspection hence, like perception, is a conceptual response to represented, and hence theoretically infected, data.[6] As a consequence, the doubt Descartes thought could only be raised for perception can also be raised for introspection, and consequently, the view that the inner is privileged must be relinquished.

3. The Myth and the Mirror

This somewhat paradoxically Cartesian attack on Cartesianism has led commentators like Rorty to adopt the view that Sellars is, or can be reinterpreted as, going after another Cartesian invention, the representational theory of mind. Rorty has characterized the view that thought is representational, together with a loosely allied set of theses about the nature of representation, as the view that the mind is a "mirror of nature." He argues that the Myth of Jones can be understood as undermining that view, and consequently, he argues, as undermining the view that thought is representational. I will argue that Rorty is correct in his assertion that the Myth does undermine the mirror metaphor, but incorrect in drawing the further conclusion that this requires a rejection *in toto* of a representational theory of mind.

Rorty characterizes the mirror metaphor in this way:

> . . . [T]o understand how to know better is to understand how to improve the activity of a quasi-visual faculty, the Mirror of Nature, and thus to think of knowledge as an assemblage of accurate representations. Then comes the idea that the way to have accurate representations is to find, within the Mirror, a special privileged class of representations so compelling that their accuracy cannot be doubted. These representations will be the foundations of knowledge. [163]

For the Cartesian, of course, the initial problem is that all we can know by directly inspecting the surface of the mirror is the pattern of light it emits. Any inference to the cause of that emission requires a theory about the degree and nature of the distortion the mirror induces, and that theory is the task of the *Meditations*. I think that Rorty is correct in attributing to

---

[6] It is important to emphasize that while perception on this account mobilizes concepts, they are mobilized perceptually as a *response* to sensation in the generation of a perceptual taking that can be understood as having the form of a judgment. This *response* contrasts with the theoretical use of concepts in *inference*. While perception is always *perception-as*, it is not inference to the best explanation of sensations, where such inference involves premises which are already judgments concerning the nature of the sensations. Hence a perceptual report, or an introspective report, while conceptual, is not thereby already inferential.

Descartes the mirror metaphor, and indeed much of Descartes' problematic can be illuminated through this metaphor. It is also then correct, if I am right in my analysis of the relation of the myths of Jones and of the deceiver to one another, that the Myth of Jones undermines the utility of this metaphor. But this would only undermine the view that thoughts are representational if one adopted either (1) the view that the Mirror of Nature theory of mind is the *only* version of a representational theory of mind, or (2) the view that the theory in which the vocabulary of thoughts is embedded is false.[7] Without the support of (1), the Myth of Jones only has as a moral that whatever the nature of the representational character of thought, it is not mirrorlike. Without the support of (2), there is no reason to believe that the unmirrorlike analysis of thoughts as representational states developed in the Myth is false. The succeeding sections of this paper will be devoted to discussing and rejecting (2). I now want only to make a few remarks directed toward casting doubt on (1).

Rorty's acknowledged debt to Sellars is great, and so is the space in [Rorty, 1979] devoted to the exegesis and development of Sellars' views. It is important to note that Rorty ignores Sellars' very different account of what it is to be representational. This account is startlingly different from Rorty's Mirror of Nature account, and when appreciated, stands as a compelling view of the nature of intentionality. Recall that Jones' model of a thought is as an internal (an unobserved) *assertion*. Recall that the categories used to construct the account of the intentionality of thought (its representational quality) are the *semantic* categories necessary for characterizing the reference or meaning of linguistic utterances. Now even Rorty never suggests that the assertion that language is meaningful presupposes a "mirror of nature" view of language. If the representational character of thought is analyzed as Jones would have it, then, it would hardly follow from the claim that the mirror metaphor is misleading that the claim that thought is representational is similarly misleading.

This is not the place to defend the Sellars-Jones account of the intentionality of thought,[8] but if it is at least a viable alternative, the claim that the Cartesian-Lockean mirror model of the mind provides the only model for mental representation is false. It is therefore important to examine the second of Rorty's theses, *viz.*, that the vocabulary of thoughts is a language which gets its sense from a theory in the context of which it occurs, which theory is radically false. This is a theme anticipated and developed

---

[7] If the vocabulary of thoughts is the theoretical vocabulary of a false theory, the only remaining analysis of thoughts as representational would be one whereon they are known to us directly, and that is the central tenet of the Mirror Theory.

[8] See [Garfield, 1988] for such a defense.

with great care in the work of Feyerabend and Churchland, and so I will focus primarily on their version of the argument, giving rather less attention to Rorty's.

4. Myth Appropriation

Churchland, following Feyerabend, offers an argument purporting to be an adaptation of the Myth of Jones to an eliminativist conclusion. The structure of the argument is as follows:

[1] The vocabulary of thoughts is a theoretical vocabulary for, and hence derives its meaning from, a theory of human behavior ("folk psychology," or, as Churchland calls it, the P-theory). (The Jonesean account of the semantics of psychological terms is true.) (see esp. § 12 of [1979])

[2] All observation, including observation of oneself (introspection) mobilizes theory-laden concepts. (The Myth of the Given is false.) [Ibid., §§ 3, 5, 16]

[3] Self-ascriptions of belief are introspective reports. (A consequence of Churchland's interpretation of the myth) [§ 13]

[4] Self-ascriptions of belief are true only if there are beliefs. (Any realistic theory of the truth-conditions of theoretical sentences, plus [2] and [3]) [§§ 4, 6, 16]

[5] There are beliefs only if something like the P-theory is true. ([1]) [§§ 13, 14]

[6] The P-theory is false. (An empirical claim) [§§ 14, 19, 20]

So,

[7] All self-ascriptions of belief are false. There are no beliefs. ([4], [5], [6]) [§§ 20, 21]

But the Myth of Jones does not force us to abandon our commitment to the existence of either mental representation or representational states like belief. Since Churchland's argument claims that we are forced to abandon this conception of ourselves and our relation to the world, there must be something wrong with the argument.

There is an interesting issue to be joined with regard to [6], and while there are interesting problems to raise for Churchland's defense of that premise, I have already discussed them at length elsewhere.[9] [4], [5], and

[9] See, for instance, chapters 6 and 7 of [Garfield, 1988].

[7] are consequences of [1]-[3]. It is hence these first three premises which represent Churchland's interpretation of the Myth of Jones, an interpretation shared by Rorty, that must be considered if the conclusion is to be undermined. I will argue that while there is an important sense in which each of these claims is true, and in which each is entailed by the Myth of Jones, [1] and [3] are ambiguous, and when read in the senses in which they are true are insufficient to support [5], and hence [7].

It is useful to begin with Churchland's formulation of the argument I have sketched:[10]

> . . . [W]hile it is one thing to claim that all feelings are felt, it is quite another to claim that all feelings are *conceptualized* as such, are *recognized* as feelings, and as feelings of a determinate sort. For this latter claim is clearly false. An infant, for example, is presumably subject to a substantial range of sensations and feelings — bodily, sensory, and emotional. But being an infant he has yet to generate or acquire the conceptual framework necessary to judge *that he is thirsty*, to recognize *that he is in pain*, or to be aware *that he is having a sensation of red*. No doubt the infant feels thirst, suffers pain, and senses redly, but a judgement to any such specific effect is as yet beyond his capabilities.
>
> In sum, the difference between having a φ-sensation, and judging that one has a φ-sensation, shows up in the following way. The making of a judgement necessarily involves the application of concepts . . . whereas the mere having of sensations and feelings does not require the application or even the possession of concepts at all. [98]

This account of introspection as the conceptual response to endogenous stimulation — as the application of concepts to experience — joins with the following neo-Sellarsian analysis of the theoretical nature of commonsensical psychological terms to yield the entire argument:

> . . . The facts are these: in the course of our daily affairs we display a systematic ability to explain, predict, and understand the behavior of certain animated particulars in terms of the wants, beliefs, pains, cogitations, and other psychological states and sequences to which they are presumed subject, and our facility in such matters is astonishing. . . .
>
> If we are to account for [this ability] at all, I do not see how we can avoid the suggestion that we share a command or tacit understanding of a framework of abstract laws or principles concerning the dynamic relations holding between causal circumstances, psychological states, and overt behavior. Bluntly, we share a moderately detailed understanding, or *theory* of what makes people tick. [91-92]

Churchland defends a particular adverbial theory of the semantics of theoretical terms the details of which needn't concern us. But it is important to note that on his analysis the meanings of theoretical terms are determined by their roles in the theories in which they are embedded,

---

[10] Churchland articulates the argument most directly with respect to feelings, and not beliefs, but he mobilizes its conclusion perfectly generally, to apply to all commonsense psychological predicates and self-ascriptions thereof.

more specifically, by their connections via nomic generalizations and theoretical postulates to other theoretical terms. The meaning of terms like "pain" and "belief," on this analysis, are given by the set of belief- or pain-laws comprised by the P-theory, and our mastery of these concepts issues directly from our (albeit tacit) mastery of this theory of human behavior. Only by mastering the P-theory do we acquire the concept of belief; only by acquiring the concept can we apply it to our internal states, and only by so self-ascribing some instance of the predicate-schema "belief that **p**" to ourselves can we report a belief.

With this more detailed understanding of Churchland's argument, we can return to the examination of premises [1] and [3]. Consider first [1]. This claim can be read either as asserting that the vocabulary of thoughts has its *origin* in a theory of human psychology or that *it currently functions as a theoretical term.* The Myth of Jones should convince us of the truth of the claim under its first reading, but not under its second. It is, however, the second reading which is required for the argument. This claim needs some explanation and defense.

It is first important to distinguish these two readings. In "Philosophy and the Scientific Image of Man" [1963] Sellars distinguishes between the manifest and the scientific images of man-in-the-world. There is considerable room for philosophical debate both concerning what Sellars intends the relation between these two images to be and concerning what it *ought* to be, independent of issues of Sellars exegesis. I will not take the time to argue here (though I defend this claim at length in [Garfield, 1988]), but will simply assume that in order to make intelligible our experience of the world, we must assume that these two images are to be preserved as coequal and mutually complementary — that a complete understanding of our epistemic situation requires what I have called in [1988] a "binocular vision." Roughly, for an entity to appear in the manifest image is for it to be part of the world as it is comprehended and perceived by sophisticated common sense and relatively naive perception. For an entity to belong to the scientific image, on the other hand, is for it to be introduced as an unobservable explanatory entity by a theory whose aim, ultimately, in concert with other theories, is to provide an explanation for the nature of and happenings in the manifest image. It is important to note (and this is an important consequence not only of Sellars' view, but also of the recent philosophy of science developed by Feyerabend and his followers) that inasmuch as *all* perception and conception — however "naive" and "commonsensical" — is theory-dependent to some extent, and inasmuch as our conceptions of theoretical entities *always* owe some debt to commonsensical concepts, this distinction is blurred and fluid. Nonethe-

less, there are clear cases: Chairs, tables, and trees appear in the manifest image; electrons, quarks, and spacetime singularities in the scientific.

Now by extension, we can assign terms places in one image or the other. Theoretical terms will be those whose meanings are dependent upon their roles in articulated scientific theories. Churchland, and Van Fraassen [1982] have joined Sellars in offering similar accounts of the semantics of theoretical terms. Putnam [1975] has offered a related account. Manifest Image terms, however, do not depend for their meanings on such rarified theoretical embeddings, or on the practice of identifiable experts, but rather on a broader network of social conventions.[11]

Terms can have origins in one image and migrate to the second. Consider, for example, "nervous," once a highly scientific term, which now has a home in everyday discourse, and no real theoretical or scientific role at all. Such semantic migrations typically reflect the tacit adoption of a theory, or at least its spirit, by the broader cultural milieu, so that persons-on-the-street come to use its theoretical terms as observation terms. (Or, for an example in the other direction, consider "force," once an everyday term which has taken on great scientific significance.)

Now, on one reading of [1], the one urged by the Myth, the vocabulary of thoughts is initially a theoretical language introduced by Jones, which, as a result of the mass conversion to Jonesianism, comes to be adopted as observation language by his neo-Rylean contemporaries. On the second reading, even after the language is widely adopted, and Jonesean psychology is the standard commonsense view, the vocabulary of thoughts retains its theoretical status (and moreover thereby acquires semantic allegiance to future scientific theories in psychology). On this second view —

---

[11] A distinction between two distinctions must be clearly drawn here: Sellars draws the theory/observation distinction *within each* image: A term is theoretical when it refers to an unobserved explanatory entity, whether or not that entity is posited by a scientific theory; a term is observational when it denotes a phenomenon that is inter-subjectively observable, even if observing a phenomenon *as satisfying* that term requires the mobilization of a theory (e.g. observing a particle in a cloud chamber). The distinction between the manifest and scientific images for Sellars is primarily a distinction between a conception of the world wherein persons, the normative predicables appropriate to them and their interactions, and the ontology of sophisticated commonsense are present on the one hand and a conception wherein the particles, forces, and properties posited by advanced scientific theories are present on the other. At issue in the present debate between realism and eliminativism with respect to the propositional attitudes, and the import of the Myth of Jones for that debate is whether the fact that our ordinary introspective access to our inner states is mediated by concepts — which we can understood as introduced as theoretical terms within the manifest image — entails that the acceptance of scientific theories which have no need of these concepts would falsify our current introspective reports.

one I will reject — the report of an inner psychological state, whether one's own or another's, involves a theoretical inference from data which provide inductive support for the report (as in a psychoanalytic argument for the claim that the analysand believes some claim that he was never conscious of believing). It is clearly this second reading which is required for Churchland's central argument to go through, for without it the claim that such reports are true only if the P-theory is true collapses. But suppose that I am right about the distinction between terms belonging to the two images. Then once the vocabulary of thoughts becomes spoken throughout the community, the Jonesean psychology becomes part of popular culture and thoughts become manifest entities (compare the fate of "nervous" discussed above).[12] It follows that [1] on this second reading is both unsupported by the Myth and is indeed false. A bit later on I will argue that things are even worse than this for the Churchland reading of the myth — that reading [1] in this way would commit one to the very Myth of the Given that the Myth of Jones is meant to replace.

Now consider [3], the claim that self-ascriptions of beliefs are introspective reports. Explaining why this claim is used by Churchland in a way reflective of a commitment not to Jonesean psychology but to givenness is a more complicated matter, and the full explication will not be clear until later in this essay. But for now, the following considerations will serve our purposes: Once we regard the vocabulary of thoughts as embedded in the manifest, and not the scientific image, and recall that the model of *thinking that* **p** is *saying that* **p**, it is clear that there are two ways in which one could intend, "I think that **p**." One could, as a consequence of introspective or other observations of oneself and one's behavior decide that the only plausible explanation for these observations is that one believes that **p** (as in the case of psychoanalytic explanation, or of the search for one's deepest philosophical commitments revealed only by careful examination of the suppressed premises in an enthymematic corpus). This is clearly the sense of self-ascription that Churchland has in mind. Or one could ascribe oneself the belief as *a way of asserting* **p**. For to believe **p** is, on this model to assert **p** *in foro interno*. And if one endorses this model, asserting **p** is one way of expressing one's belief in its

[12] This is not to say that observing or reporting thoughts is entirely non-conceptual. *That* would be to embrace givenness with a vengeance. Rather, when thoughts become manifest entities the use of the concept of a thought in reporting one's inner states becomes non-inferential. The application of the predicate is spontaneous — caused by one's inner state, and not *inferred* from theoretical premisses plus observations issuing in perceptual beliefs couched in a different, less theoretical vocabulary. (That, as John Connolly has remarked, is why there's no such thing as *justifying* one's own profession, "I believe that **p**," though if Churchland's view were correct, there would have to be such a thing.)

truth. Saying "I believe that **p**" is on this account but a way of saying **p**, though perhaps depending upon such variables as conversational context, stress, and dialect conventions, with somewhat different illocutionary force than a simple assertion of "**p**." This second form of self-ascription, what I will call "professing," will play a large role in the account to follow. For now note that it is one — perhaps the dominant — sense of self-ascription that one might envisage in post-Jonesian culture, and not the sense of self-ascription sufficient to render Churchland's argument sound. For on this account, the vast majority of belief-ascriptions to oneself are not introspective — they are merely assertoric.[13]

This distinction explains away Moore's paradox: In the case where "I believe that **p**" is read introspectively, I am describing the inner state of some epistemic subject [who happens to be me] as having the semantic properties of an assertion that **p**, which is surely not inconsistent, even pragmatically, with my avowal that **not-p**; in the case where the utterance is taken as an *expression* of my belief that **p** I am asserting that **p**, and a subsequent assertion that **not-p** would plunge me into contradiction. The paradox trades on this subtle ambiguity in "I believe that——," an ambiguity not shared by "She believes that——," a point to which we will return below. Note that the ambiguity in question is a peculiar one — perhaps more accurately characterized as a pragmatic, rather than a semantic ambiguity. For on the reading where "I believe that——" asserts that I and a proposition satisfy some relation, the sentence has quite ordinary truth conditions (*modulo* more general worries about the truth-conditions of sentences containing propositional attitude verbs). On the more common professional reading, however, the sentence as a whole has a more performative character: There is a sense in which it is self-confirming — if I assert it, I *ipso facto* assert its embedded sentence. There is another sense — that in which I evaluate the sentence merely as a fancy assertion, in which its truth-conditions are simply those of the embedded sentence.[14]

---

[13] There is, however, another important gloss available for other propositional attitude verbs: Consider, "I was just thinking that **p** (but of course **not-p**)." Here one is neither asserting **p** nor theorizing. One is responding perceptually — introspectively — to one's inner state. But both the analogy of perception and the primacy of profession are important to understanding the nature of this kind of introspective judgment: First, note that the response, while mobilizing the concept of an inner episode, does not in virtue of that fact issue from an inference from an introspective perceptual awareness plus a theory about the cause of that awareness. Second, the concept of *thinking that* used in this non-committal judgment is parasitic on that of belief, and is a subjunctive version of profession.

[14] It is not clear just how many propositional attitude verbs share this ambiguity. Some immediately come to mind, e.g. "hope" and "intend." I can infer that I *must* hope for some eventuality from my behavior, or my emotional reactions, or I can *profess* "I hope

One final eliminativist rejoinder must be considered at this point. It might be argued that while I have defended the meaningfulness of belief-ascriptions, I have not really made contact with the eliminativist positions regarding belief. For what is to be eliminated, according to these positions, is the existence of beliefs as explanatory entities, and while I have argued that the term "belief" may have a professing role and a parasitic commonsensical reporting role, none of this, it might be argued, entails the existence of genuinely explanatory inner episodes. At best, this line of argument would conclude, I have defended a *façon de parler*. But this would be to miss the point both of Churchland-Feyerabend eliminativism and of my argument. It is not for nothing that Churchland compares belief in belief to belief in phlogiston, *elan vital* and witches. These are uncontroversially nonexistent, in every non-Meinongian sense of that word. To defend *any* realistic account of belief is hence to challenge that position. Moreover, once one grants the existence, representational character and explanatory power of utterances of sentences, and a model of belief as a form of assertion — assertion that may or may not be actualized in speech — the conclusion that there are explanatorily useful representational inner episodes appears unavoidable. And it is just these episodes that the Myth of Jones demystifies and that Churchland argues are nonexistent.

I conclude that Churchland's reading is at least undersupported by the text, and that no eliminativist conclusion is warranted on these grounds from the Myth of Jones. I have not yet argued that no such conclusion is *compatible* with the Myth. I will argue for that stronger claim in subsequent sections.

The critical touchstone for Rorty's appropriation is the denial implicit in the Myth of Jones of the privileged quality of one's own inner states. For Rorty, essential to the idea of inner representational states ("the 'idea' idea") is their character as directly given *as the states they are* to their bearers. The epistemological role that they play on this Cartesian straw view is that of certain foundations of knowledge. Now to be sure, the moral of the Myth of Jones is that whatever inner states there are, and whatever the meanings of the terms in the vocabulary of thoughts, there are no such privileged inner states accessible to an inner eye. Now Rorty infers from this that any theory of human nature which ascribes inner representations must be false, and that there are hence no such representa-

that . . ." as a way of saying "Would that it were the case that . . ." Or again, I can report an intention to carry out an action, or I can announce my intention in acting. This suggests that Moore's paradox is quite general (I want a piece of chocolate cake for dessert, but I hope I don't get one — because I'm on a diet), and perhaps underlies certain problems about weakness of will. If so, this analysis of the paradox may provide an interesting way of understanding *prima facie* paradoxical assertions of contradictory intentions.

tions. But as I have argued, central to the Myth is the idea that inner representational states can be understood as representational not *despite*, but rather *because* of their status as non-privileged — that is — because of their quasi-linguistic character. If understood in this way, we will see, the eliminativist sting of Rorty's attack on the mirror model of the mind can be avoided.

5. Locking Up the Myth

How do Rorty and Churchland manage to get the Myth so wrong? The answer, I believe, is the tacit adoption, and the reading of Sellars as adopting, a Lockean view of introspection.[15] By a Lockean view I mean one whereon the warrant for judgments about our current internal states derives from self-observation via an inner sense, which, like an organ of outer sense, responds to endogenous stimulation to form perceptual representations.[16] On such a view our self-knowledge derives from self-perception through a faculty very much like a mind's eye. Locke puts this view quite nicely:

> Secondly, the other fountain from which experience furnisheth the understanding with ideas is, — the perception of the operations of our own mind within us, . . . which operations furnish the understanding with ideas which could not be had from things without. And such are *perception, thinking, doubting, believing, reasoning, knowing, willing*, and all the other different actions of our own minds; . . . This source of ideas every man has wholly in himself; and though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly be called *internal sense*. [123]

And again,

> The mind receiving the ideas mentioned in the foregoing chapters from without, when it turns its view inward upon itself, and observes its own actions about those ideas it has, takes from thence other ideas, which are as capable to be the objects of its contemplation as any of those it received from without. [159]

Leaving aside the adequacy or inadequacy of such a view of self-knowledge for a moment, let us see just why the Churchland-Feyerabend-Rorty interpretation of the Myth of Jones is committed to this Lockean view. Central to their indictment of the reality of inner representational states is

---

[15] To be sure, neither Rorty nor Churchland would endorse the naive view of introspection to which I will charge they are committed. Rather I will argue that their explicit view about introspection entails this Lockean model. The argument thus has the form of a *reductio*.

[16] This view — which I will argue is incoherent — is not to be confused with the genuinely Sellarsian (and anti-Cartesian) view that introspection is *in certain respects* analogous to perception — e.g. that both involve the (noninferential) conceptual *response* to stimulation.

their view that self-descriptions using the vocabulary of thoughts are *theoretical responses to current internal states*, which states ought to be described using some theory more adequate than the P-theory (advanced neuroscience, for instance). Now central to this indictment is the charge that the self-ascriber is *misdescribing* the current state s/he *observes* in him/herself, which state could be more accurately described in the language of a better theory.

On this view, the self-ascription of belief is a two-stage affair. First we observe our internal state. Then, using Jonesean psychology, the P-theory, or whatever, we theorize about the correct characterization, or perhaps the cause, of this current state. We conclude that we believe that p. Then we report that belief. The Lockean character of this picture should be obvious. It is important to be clear that the account of perception to which eliminativist proponents of this Lockean model of introspection are committed is also a two stage model. This is explicit in both Churchland's and Feyerabend's characterization of our sensory organs as measuring instruments whose readings we use in the fixation of (for lack of a better word — it is always hard to characterize the — dare I say it — beliefs of eliminativists) perceptual belief. On this view, in perception, our senses first react physically to external stimulation (stage one); then, we construct, on the basis of the character of this stimulation, a percept (stage 2). The nature of this percept according to folk psychology or its cognitivist descendent might be a representation of some kind, perhaps a perceptual belief. According to eliminativists, it is something we will be in a position to characterize only in the golden age of neuroscience. Presumably, though, it occurs neurologically upstream of both the immediate neural consequences of current stimulation and the more stable consequences of prior stimulation histories, maybe in something like Luria's tertiary projection regions.

The pernicious consequence of this Lockean view of introspection should be obvious. For essential to the two-stage model of introspective report is the supposition that we are immediately aware, through an inner sense, of our internal states, and that our report is an inference from this awareness. But that is to suggest that the contents of this inner sense are *given* to us independently of the theoretical apparatus through which we cognize them — that we are aware of them *as the contents of consciousness they are*, though we are unable to report them, save through language which is theoretically infected. To be sure, on this model, introspective *reports* are not self-evident, self-evidently true, or infallibly accurate accounts of our current psychological states. *They* are theoretical responses. This is, of course, the heart of the eliminativist attack on repre-

sentation. But — and this is the central point for our present purposes — the inner sensations to which these introspective reports are conceptual responses *are given to us in the very way that the Myth of the Given suggests sensory states are given*. As Churchland puts it, ". . . self-perception consists in the disposition-governed occurrence of *conceptual responses to one's internal states*, responses made within whatever matrix of self-understanding one has developed or acquired. Our current matrix of response is the P-theory . . ." [1979, p. 116, emphasis added]. One might reply at this point that while an inner φ-sensation must be given to us, it need not be given as a *φ-sensation*, and so that the model is not committed to givenness in the pernicious sense. But this reply is not available. For if the responses to these states are theoretical, and hence inferential, the premisses of the inferences must be available in a vocabulary rich enough to support the requisite theoretical logic. The initial awareness of the φ-sensation must hence be of a type sufficient to function as a premiss in an inference mediated by a theory about φ-sensations as such, and hence must be an awareness of it *as a φ-sensation*.

When the Feyerabend-Churchland model of perception is taken over as a model of introspection givenness enters the picture in this way: the disposition-governed conceptual response to sensation in perception is an inference from effect to cause. Sensations are not, on this account, *given* to us. They merely function as *causes* of perceptual representation. But this is *not* the way that internal states can function in self-observation. When we observe a tree, the tree is the epistemically direct object of the observation — the sensations are not observed. They are merely causally intermediate between the distal tree and the representation of it delivered by the perceptual system. But in self-observation, the putative *causes* of the perceptual state under consideration — the introspective representation of our current mental state — are themselves also the epistemic direct objects of (self-) observation. They hence must be represented to us — as is the tree in perceptual observation — as the kinds of things they are. But this would be for the *causes* of self-perception to be *given* as the *objects* of self-perception, which would be to require a dramatic version of immediate access to one's internal states, one surely unpalatable to the austere empiricism of eliminativism. The Feyerabend-Churchland-Rorty line thus reintroduces the Myth of the Given — albeit in its empiricist guise — putatively as a consequence of the very argument that they hold to be decisive against that very myth.

Before proceeding to a non-Lockean reinterpretation of the Myth of Jones, and a consideration of the consequences of that reinterpretation for an account of introspection and for a realism with respect to representa-

tions, it is worth considering just why the Myth is so often misread in this Lockean fashion. I think the answer is simple, but it is useful to note, so as to facilitate our attempts to understand the argument more thoroughly. Recall the denouement of the myth: *"What began as a purely theoretical use has gained a reporting role."* [189] When one puts together this claim for a reporting role for the vocabulary of thoughts with the claim that the states that are ostensibly reported are *inner*, one gets the claim that the vocabulary of thoughts has as its primary use the report of inner states. Put this together with the Sellarsian analysis of perception, endorsed by Feyerabend, Churchland, and Rorty, whereupon perception is, *inter alia*, a conceptual response to sensory stimulation. It is then easy to conclude that the consequence of the adoption of Jonesean psychology is that post-Joneseans conceptually respond to the sensory deliverances of an inner sense, one whose immediate contents are independent of the theories used in forming judgments regarding their causes or natures.

When the point is put this way, it is easy to see at the same time why it is tempting to interpret the account of introspection in this way, and to see what the problem is with this interpretation. It presupposes that belief report is a species of *observation*, distinguished only by the fact that it is the observer and his/her inner states that are the objects of observation. If we are to un-Locke the Myth then, it will be essential to develop a non-observational account of first-person belief reports, and to place this account in the context of the Myth of Jones.

6. Unlocking the Myth

I alluded above to a distinction between *professing* belief and *introspecting* belief. It is now time to return to that distinction, to draw it more carefully, and to explore its consequences. Wittgenstein's [1980] allusion to Moore's paradox is helpful here:

> 280. But what does this mean: "It's raining and I don't believe it" makes sense if I *mean* it as a hypothesis, and does not make sense if I meant it as an assertion or a report? . . .
>
> 281. One would also like to say: The assumption that I believe something is the assumption that I am disposed in a *particular* way. Whereas I should not want to say of the report "I believe . . ." that it says something about my disposition. Rather it is an utterance of this disposition.
>
> 283. If we were to have an obligatory "I believe" at the beginning of every assertion, "I believe it is so" would mean the same thing as "It is so." [II, p. 54e]

This occurrence of "I believe . . . ," as an *actualization* of a disposition, rather than as a *report* of a disposition or other kind of state in normal discourse is contrasted with its introspective use:

692. It is of course the indisputability [of first person belief-reports] that suggests the picture of something's being described here, something that we see and the other does not, and that is near to us and always accessible, but for the other is hidden: something that exists *within* us and which we become aware of by looking into ourselves. And psychology is now the theory of this inner thing.

693. So if I want to say that our "utterances," with which psychology has to do, absolutely are not all descriptions of experience-contents, I must say that what are called descriptions of experience-contents are only a small group of these "indisputable" utterances. . . . [I, 127e]

700. One would like to say that "I believe" can't *properly* be the present of "I believed." Or: one ought to be able to use a verb in such a way that its past has the sense of "I believed," while its present has a sense different from that of our "I believe." Or again: There ought to be a verb, whose third person in the present tense has the sense "he believes," but whose first person has a sense different from that of "I believe." . . . [I, 128e]

It is this other unavailable verb that would capture what I want — somewhat paradoxically — to call the introspective use of the verb. This is the sense of "believe" we use on those rather rare occasions when we are really trying to *figure out* what we believe, as perhaps in psychotherapy, or in understanding our response to a deception, or when tracing out the non-obvious consequences of our explicit, acknowledged beliefs. In these cases — and it is important to note how rare they are; in a moment we will argue that they are in fact parasitic on the normal use — we adopt a third-person attitude towards ourselves. Only then are we able to *observe* ourselves, to *theorize* about ourselves. And it is crucial to note that it is in these cases that the normal deference to first-person authority which is the germ of truth in privileged access theories of the mental is in fact waived. (It is, I think, ironic that it is in the case that *is* genuinely *introspective* that the privileged access that is supposed by friends of the given to attach to first-person belief-reports *in virtue of their introspective character* is most obviously lacking.)

In the normal case, to report a belief is to make an assertion, to profess that the content of the belief clause is true, to endorse that content. Unless that content has specifically to do with the internal state of the believer, to report a belief is usually therefore to say nothing about any inner state, though such a report to be sure would constitute evidence regarding the believer's beliefs. Belief on this model is internal utterance; belief report is public utterance. The point of Jones' taking assertion as the paradigm for belief is that belief-profession and utterance are indistinguishable semantic events. Third person belief reports on the other hand, or past-tense first-person reports, are theoretical claims about the utterances of another, or of oneself at a stage towards which one stands epistemically as

another. These are claims about internal assertions. The point about a special sense of "I believe," one with a peculiar introspective character, is that this is a sense in which the phrase marks a peculiar attitude towards oneself — one in which one in one voice professes, and in another theorizes. But it is of the utmost importance both to distinguish these two psychological activities, and to note that one could not theorize about anyone's — including one's own — professions without presupposing the occurrence of profession. The self-observational sense of "I believe . . ." hence presupposes — but is not presupposed by — the primary professing sense. The reporting sense, as in "I was just thinking . . ." or in "I once believed . . ." is also parasitic on the professing use. For it notes a counterfactual tendency to profession, and not some non-assertionlike inner state mistaken for an inner assertion.

On this account, the representational character of language undergirds and explains the representational character of belief. The fact that the profession of belief is the primary function of "I believe . . ." means that belief reports are usually not introspective reports, but are rather, in the typical case, assertions of the content of the sentence embedded in the belief-clause. And in the case where introspection *does* issue in belief-reports, profession is first of all primary, and the evidential basis for the belief report is not a given inner sense datum, but rather the relevant behavior, dispositions, and other beliefs of the reporter. Hence not only is introspection not the primary ground of belief-report, but in those cases where it is the ground, the model is observational, but not Lockean. Moreover, as we have seen, while it may be observational, it is not perceptual in the sense that it involves a "disposition-governed conceptual response" to some *sensation*. Introspection in this sense is theoretical inference. Some attitude reports (typically not of beliefs) are, however, perceptual in the sense that they are "disposition-governed conceptual responses." But these represent more the *actualization* of dispositions to assert than observations of or *responses to* dispositions.

7. Jones Without Mirrors

How do these considerations bear on the Churchland-Feyerabend attack on the reality of belief? Well, central to this attack was the claim that self-ascriptions of beliefs were theoretical responses to Lockean introspective data, making use of the P-theory to describe and explain one's behavior and psychological processes. It was from the alleged falsity of this alleged theory allegedly employed in such alleged introspection that the unreality of belief was meant to follow. But I have argued that on a correct understanding of the Myth of Jones there is no such introspection involved in standard self-ascriptions, and that the introspection involved in non-stan-

dard self-ascriptions is not Lockean in the requisite sense. I have argued further that such self-ascriptions do not involve any especially theoretical activity (or observational, for that matter) but are rather to be understood as professions. Given these understandings, the eliminativism is ungrounded, and is demonstrably a consequence of resmuggling the Myth of the Given into the interpretation of the Myth of Jones.

Moreover, the analysis I have offered undermines Rorty's appropriation of the Myth in the service of an argument against the representational character of mind. For central to that use of the Myth was the supposition that the Myth, in virtue of undermining the notion that there are no *privileged* representations,[17] undermines the notion that there are *any* representations. But this inference succeeds only to the extent that the model of mental representation one uses is a mirror model. If the model is a linguistic model, as I have argued that it is for Sellars and for Jones, no such inference is possible. On this view, though mind does not mirror, it does represent, just because its inner states can be understood as sharing in the semantic character of language, whose representational character is presumably not problematic in the way that Cartesian Thoughts are.

It is important to note before closing, however, that despite this rejection of much of Rorty's appropriation of the Myth, there are two important senses in which Rorty gets the moral of the Myth right — senses which emerge more clearly from this analysis. First, there is an important sense in which knowledge, on the Sellarsian picture emerging from the Myth, is homogeneous: in this sense there is no special privilege attaching to first person introspective judgments. This is because in introspecting, as we have seen, there is an important sense in which one is taking a third-person attitude towards oneself, and responding theoretically to one's observations of oneself. The theories one uses may well be wrong. One's observations may be distorted, due to deficiencies of sense, non-optimality of observation conditions, or a radically false theory in the context of which observation occurs. These are, of course, just the sources of fallibility of perception generally. What is important to note, however, is that not all belief-reports are introspective reports.

Second, as Rorty points out, being situated in a community of language users is a necessary condition for believing. Believing is not a Cartesian inner process. This is true simply because belief on the Jonesian view is assertion, and assertion is linguistic, and language is a collective enterprise. Again, though, this does not entail the unreality of belief, or of representation. It only means that belief and representation are not Cartesian

---

[17] That is, representations *given* in the appropriately mythic sense to consciousness through which we have epistemic access to reality.

in character. But why in the world would anyone think they were, if not for the powerful grip of the Myth of the Given? Perhaps the greatest lesson in all of this is that the grip of that M*y*th of the Given is tighter on our collective imagination than any of us realize.

References

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

Dennett, D. (1987). *The Intentional Stance.* Cambridge, Massachusetts: The MIT Press/Bradford Books.

Feyerabend, P. K. (1979). *Realism, Rationalism & Scientific Method.* Cambridge: Cambridge University Press.

Garfield, J. L. (1988). *Belief in Psychology.* Cambridge, Massachusetts: The MIT Press/Bradford Books.

Locke, J. (1959). *An Essay Concerning the Human Understanding Volume I,* ed. by A. C. Fraser. New York: Dover.

Lycan, W. (1987). *Consciousness.* Cambridge, Massachusetts: The MIT Press/Bradford Books.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories.* Cambridge, Massachusetts: The MIT Press/Bradford Books.

Putnam, H. (1975). "The Meaning of 'Meaning'," in *Mind, Language and Reality: Philosophical Papers Volume II.* Cambridge: Cambridge University Press.

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

Sellars, W. (1963). *Science, Perception, and Reality.* London: Routledge and Kegan Paul.

Van Fraassen, B. (1982). *The Scientific Image.* Oxford: Oxford University Press.

Wittgenstein, L. (1953). *Philosophical Investigations.* New York: Macmillan.

Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology,* ed. G. E. M. Anscombe and G. H. Von Wright. Chicago: University of Chicago Press.

Appendix — The Myth of Jones

48. . . . Imagine a stage in prehistory in which humans are limited to what I shall call a Rylean language, a language of which the fundamental descriptive vocabulary speaks of public properties of public objects located in space and enduring through time. Let me hasten to add that it is also Rylean in that although its basic resources are limited, its total expressive power is very great. For it makes use not only of the elementary logical operators of conjunction, disjunction, negation, and quantification, but especially of the subjunctive conditional. . . . [Sellars, p. 178]

49. The questions I am, in effect, raising are, "What resources would have to be added to the Rylean language of these talking animals in order that they might come to recognize each other and themselves as animals that *think, observe*, and have *feelings* and *sensations*, as we use these terms?," and "How could the additions of these resources be construed as reasonable?" In the first place, the language would have to be enriched with the fundamental resources of semantical discourse — that is to say, the resources necessary for making such characteristically semantical statements as "'*Rot*' means 'red'" and "'Der Mond ist rund' is true *iff* the moon is round.". . . [179]

50. With the resources of semantical discourse, the language of our fictional ancestors has acquired a dimension which gives considerably more plausibility to the claim that they are in a position to talk about *thoughts* just as we are. For characteristic of thoughts is their *intentionality, reference*, or *aboutness*, and it is clear that semantical talk about the meaning or reference of verbal expressions has the same structure as mentalistic discourse concerning what thoughts are about . . . [180]

53. But we are approaching the time for the central episode in our myth. I want you to suppose that in this Neo-Rylean culture there now appears a genius — let us call him Jones . . . [183]

56. . . . Suppose that in the attempt to account for the fact that his fellow men behave intelligently not only when their conduct is threaded on a string of overt verbal episodes — that is to say, as *we* would put it, when they "think out loud" — but also when no detectable verbal output is present, Jones develops a theory according to which overt episodes are but the culmination of a process which begins with certain inner episodes. *And let us suppose that his model for these episodes* which initiate the events which culminate in overt verbal behavior *is that of overt verbal behavior itself. In other words, using the language of the model, the theory is to the effect that overt verbal behavior is the culmination of a process which begins with "inner speech."* [186]

57. The first thing to note about Jonesean theory is that, as built on the model of speech episodes, *it carries over to these inner episodes the applicability of semantical categories*. Thus, just as Jones has, like his fellows, been speaking of overt utterances as *meaning* this or that, or being *about* this or that, so he now speaks of these inner episodes as meaning this or that, or being *about* this or that. . . . [187]

58. . . . (5) It cannot be emphasized too much that although these theoretical discursive episodes or *thoughts* are interpreted as *inner* episodes — which is merely to repeat that they are introduced as *theoretical* episodes — they are not introduced as *immediate experiences*. Let me remind you that Jones, like his Neo-Rylean contemporaries, does not yet have this concept. . . . [188]

59. Here, then, is the denouement. I have suggested . . . that although it would be misleading to say that concepts pertaining to thinking are theoretical concepts, yet their status might be illuminated by means of the contrast between theoretical and non-theoretical discourse. We are now in a position to see exactly why this is so. For once our fictitious ancestor, Jones, has developed the theory that overt verbal behavior is the expression of thoughts, and taught his compatriots to make use of the theory in interpreting each other's behavior, it is but a short step to the use of this language in self-description. Thus when Tom, watching Dick, has behavioral evidence which warrants the use of the sentence "Dick is thinking **p**." Dick, using the same behavioral evidence, can say, in the language of the theory, "I am thinking **p**." And now it turns out . . . that Dick can be trained to give reasonably reliable self-descriptions, using the language of the theory, without having to observe his overt behavior. . . . Our ancestors begin to speak of the privileged access each of us have to his own thoughts. *What began as a purely theoretical use has gained a reporting role*. . . . [189]