INCORRIGIBILITY, THE MENTAL AND MATERIALISM

Gerald Doppelt

December 1976
Abstract:

This paper constitutes a thoroughgoing critique of Rorty's interesting attempt to characterize the mental and its elimination within materialism in terms of the incorrigibility of mental reports. I elucidate, criticize, and improve the concept of incorrigibility his position requires. Then I argue: (1) that although mental-state reports are as corrigible as physical reports, this reflects contingent matters which do not affect the boundary of the mental and the physical; (2) that even if the familiar paradigm mental-event reports ("I am in pain") are incorrigible, there are mental events for which our language does not provide descriptions plausibly considered as incorrigible; (3) even the familiar mental-event reports are not incorrigible which I show through examples that explain how and why persons maintain false beliefs about their most simple sensations, thoughts, indeed anything. I then suggest that Rorty's conception of the triumph of materialism is simplistic and inadequate in a number of respects. Finally, I attempt to show how difficult if not impossible it is to define or eliminate the mental without presupposing it; in trying to get the barest sense of Rorty's materialist world, the mental forces itself into our mind at every turn.

*This paper constitutes an expansion of a talk by the same title delivered before the APA (Pacific Division meetings, March 1973; abstract Journal of Philosophy, Feb. 1973). The author wishes to express his appreciation to Zeno Vendler for several valuable criticisms and suggestions concerning this paper which have been incorporated into its text.
In "Incorrigibility as the Mark of the Mental"\(^1\) and "Functionalism, Machines, and Incorrigibility", Richard Rorty argues that a certain concept of incorrigibility which he develops provides the most adequate way of distinguishing the mental from the physical. On this basis, he further suggests what sorts of developments within science would be sufficient to eliminate the mental and thus establish materialism. Sections I-IV (pp. 1-35), of this paper attempt to expose fatal inadequacies in Rorty's characterization of the mental in terms of incorrigibility. In this process, however, I am more broadly concerned to argue against widespread and well-entrenched assumptions concerning mentalistic reports common to most incorrigibility theorists as well as their critics; the view thus elaborated leads to the conclusion that our mental self-knowledge and knowledge of others is complicated and vulnerable to error in ways that have not previously been understood. In section V (pp. 35-45) I go on to argue that Rorty's conception of the relationship between scientific development and the elimination of the mental is simplistic and incorrect. Here, my concern is to criticize Rorty's conception of materialism in a way that raises very general problems for certain widespread assumptions underlying the contemporary materialism debate. The main thrust of my critique is to expose features of our existing paradigms for knowledge of the mental which are neglected and effectively distorted by the prevailing approach to incorrigibility, the mental, and its eliminability. Our discussion begins with an exposition of Rorty's position and some interpretive remarks concerning its distinctive features and unclarities.

I

On Rorty's concept of incorrigibility, a person R's belief that p at time t is incorrigible if and only if there is no accepted procedure (empirical test, factual investigation) whose outcome would render it rational to believe not-p at t, given that R believes p at t. Rorty rightly distinguishes this epistemic property of incorrigibility from two others with which it has been confused: (1) incorrigibility as the logical impossibility of falsification (on which a belief is incorrigible if and only if belief entails knowledge, hence truth); and (2) incorrigibility as the non-inferentially knowable or credible (on which a belief is incorrigible if and only if the belief may be credible, evident, or known independently of evidence or inference). Furthermore, for Rorty, the incorrigibility of certain mentalistic beliefs is not to be explained by, or associated with, the postulation

\(^1\) *Journal of Philosophy*, LXVII, 12 (June 25, 1970): 399-424

\(^2\) *Journal of Philosophy*, LXIX, 8 (April 20, 1972): 203-220
of special mental entities (e.g. a given), or modes of aware­ness (e.g. direct acquaintance). Moreover, this incorrigibility is not viewed as a consequence of the meaning or "logic" of our present mentalistic vocabulary; such a view is rejected by Rorty on grounds of a general Quinian skepticism concerning the analytic-synthetic distinction. Thus, for Rorty, the ascription of incorrigibility to a certain class of beliefs marks a contingent fact about our present epistemic attitudes and practices concerning these beliefs, and not something essential to the nature of reality, human awareness, empirical knowledge, language, or even our present language. This ascription of incorrigibility asserts only that our linguistic community adopts a certain epistemic attitude to the class of beliefs in question: namely the attitude that with respect to such beliefs, no evidence against their truth whatsoever, will be allowed to discredit them in favor of their negations.

How, then, does Rorty employ this concept of incorrigibility to distinguish the mental from the physical? He distinguishes mental events (e.g. pain) from mental states (e.g. anger) and argues as follows: (1) any event $x$ is mental if and only if there is some person who can have an incorrigible belief in some statement $p$ which is a report on $x$; and (2) any state $y$ is mental if and only if there is some person who can have a belief in some statement $p$ which is a report on $y$ and "almost" incorrigible. Roughly stated, the resulting view comes to this: an event or state is mental if and only if at least one person (i.e. the owner) is in a position to maintain some belief about it whose initial credibility has an absolute or "almost" absolute weight against any putative body of negative, or disconfirming, evidence whatsoever. So, for example, my sensation of pain is a mental event because I am in a position to maintain the belief that I am in pain, and this belief is credible (evident, justified, reasonable) irrespective of any other evidence concerning behavior, situation, personality, etc., which may be produced. Likewise, my state of anger is a mental state because I am in a position to maintain the belief that I am angry and this belief remains credible "almost" irrespective of other evidence that may be produced. Although, in the latter case (and the general case of mental states) other evidence can override the belief in exceptional circumstances: nevertheless, "we are far less likely to have a report about a mental state...overridden than to have a report about something physical overridden."3 Thus, for Rorty, the existence of the mental comes to the fact that our linguistic community accords a special epistemic privilege

3"Incorrigibility as the Mark of the Mental", op.cit., p. 420
to certain mentalistic propositions in resolving the conflicting demands which the totality of new propositions exerts on our system of knowledge.

II

Before proceeding to criticism, I will make some interpretive remarks: (a) First, although Rorty does not address himself to the issue, it is important to note why he attempts to capture the mental in terms of a person's capacity to have some incorrigible belief in a statement which is a report on it. Presumably, Rorty will want his criterion to allow that e.g. my pain is mental even though either (1) I do not have any belief about it or (2) I happen to have some corrigible belief(s) about it. Let us deal with the first situation: this situation would obtain if I totally lack the concept of pain and thus have no (corrigible or incorrigible) beliefs about my pain. Presumably, Rorty will even want to hold that babies who (in one sense) lack the capacity of conceptualization and adults who lose this capacity through brain damage nonetheless may have pain (and "the mental"). We certainly believe this to be the case.

If, as I assume here, in all of these cases Rorty will maintain that the individual in question "can" have the incorrigible belief required to ascribe the mental to him, it becomes clear that "the capacity" involved in his analysis requires a quite special construal. He needs a sense of capacity compatible with the modes of incapacity mentioned above for the crucial incorrigible beliefs (consequent upon conceptual incapacities). As I interpret Rorty's position, the existence of the capacity in question implies the existence of a "mentalistic" language in our linguistic community, i.e., a language which permits the formulation of "incorrigible" statements, statements which this community treats in the privileged manner (explained above). The existence of the capacity, paradoxically, does not imply that the individual to whom it is ascribed knows or even can learn this language; all it implies is that he is in a state such that were he to be capable of appropriate modes of conceptualization, the relevant incorrigible beliefs or reports would be available to him. If, however, Rorty wants to deny that animals possess mental events, he will require a notion of capacity for incorrigible beliefs which excludes animals but includes all humans, in particular those who "lose" this capacity, or are not capable of developing it from birth. (b) Secondly, as I suggest above, Rorty will want to allow that my pain is mental even though I in fact maintain corrigible (or false) beliefs about it. Examples are my belief that this is the fifth sensation I've had in my arm today or my belief that this sensation is caused by the hammer's falling on my arm. Both of these beliefs are inferentially credible for me, i.e., credible on the basis of empirical
evidence, other things I now know. All such inferentially credible beliefs are corrigible through the falsification of their empirical premises.

The corrigibility of one's actual beliefs about his mental events is of course compatible with his capacity to believe some incorrigible report(s) on it. All that Rorty's view requires is that for every mental event, there be some predicate in our language which permits one to make some report on it which is non-inferentially credible, and moreover incorrigible. Indeed, these predicates would be akin to an empiricist's basic vocabulary, terms which will permit justified applications without further justifying evidence. The claim that such a language exists need not involve the additional empiricist doctrines of meaning through ostensive definition, the given, the immutability of this language, its 'foundational' role in empirical knowledge, etc. However, a serious problem which does arise concerns the scope of such a basic vocabulary, whether or not for every mental event or state a person experiences, there is some predicate in the basic (i.e., non-inferentially credible) vocabulary available for its characterization. A closely related problem concerns the question of what statements or predicates constitute "a report" on my mental event or state. Suppose all characterizations of an event are corrigible except the one "I am experiencing a visual sensation of some sort" or "I feel something in my arm". Are these sorts of characterizations sufficient to generate the mental? This would appear to violate the spirit and examples in Rorty's work, but Rorty will need to clarify this point, as well as the more general one of which of the statements one can make are "incorrigible", given that many (like the above examples) are corrigible.

(c) Rorty's concept of incorrigibility and view that the existence of the mental amounts to the existence of a class of "incorrigible" first-person experiential reports is of general epistemological significance. For this view is interestingly located somewhere between traditional empiricist epistemology (e.g. C.I. Lewis) and the current wholistic

---

4For an elaboration of this point see my doctoral dissertation, Characterization of Phenomenalist Statements in Theories of Knowledge, Chapter V, esp. pp. 36-46, Johns Hopkins University, 1969, or University Microfilm, Ann Arbor, Michigan
pragmatism associated with Quine and Scheffler. Thus while Rorty rejects the traditional view that the incorrigibility of experiential reports is a necessary feature of any body of empirical knowledge, he is not content to maintain that these reports are always open to rational revision on pragmatic grounds and are thus in the same epistemic boat in which all empirical claims float. If I grasp Rorty's position, he wishes to relativise the existence of incorrigible beliefs, (thus, "the mental") to epistemic and linguistic frameworks (or, epochs) without dissolving it as radical pragmatism appears to.

In this perspective, we might characterize Rorty as a framework-pragmatist but not a case-pragmatist. He maintains that epistemic frameworks (sets of epistemic attitudes, canons of evidence) may come and go, for pragmatic reasons; nonetheless, with respect to particular epistemic frameworks, one can still capture epistemologically significant structures, edifices, or practices within them, like the bifurcation between physical and mental, corrigible and incorrigible within our own. Furthermore, case-pragmatism distorts the truth in suggesting that within our present framework certain epistemic moves (e.g. rejecting my sensation report to preserve a useful theory) are in order or reasonable; when, indeed, the rationality of such a move depends on the pervasive revision of our present epistemic framework. In particular, the illegitimacy of a class of such moves with respect to certain mentalistic reports is the only general feature which differentiates the mental from the physical.

III

In this section, I argue that even within our present epistemic framework, incorrigibility and almost-incorrigibility do not serve to distinguish the mental from the physical, as Rorty's view requires.

(a) Mental Events and Mental States: First, I will argue that Rorty cannot distinguish mental events from states employing his notion of "almost-incorrigibility". In Rorty's view,


6This suggestion is illustrated and argued in Scheffler's Conditions of Knowledge, p. 38.
mental-state reports unlike reports of mental events, are "almost" incorrigible, because there are "accepted procedures" for overriding them. These procedures exist for mental-state reports but not reports of mental events because the former "are implicit predictions of future behavior, predictions which may be falsified" whereas the latter "are compatible with any range of future behavior." This won't do, as the classic incorrigibility empiricists well understood. Even reports on mental events like "I am now in pain" or "I am experiencing a red sense-datum" enjoy an incorrigibility confined to the present; for these reports can be falsified in the future where they may come into conflict with memory or other indirectly disconfirming evidence. What I or you say, remember, or learn in the future may falsify my present report of pain. Thus if "almost-incorrigibility" means falsifiability in the future, both mental-state and event reports will be only "almost-incorrigible," not simply "incorrigible."

However, Rorty is mistaken in thinking that mental-state reports are only falsifiable through procedures which disconfirm their implications for behavior in the future. These reports can be overridden in the present without recourse to any "procedures" or "future behavior." For example, one basis for rejecting in the present a person's sincere insistence that he is not angry is an observational knowledge of his preceding and on-going behavior, facial expression, and situation. In more complex cases, this kind of observation plus independent knowledge of his personality will be required to justify a rejection of his report; e.g. the knowledge that this individual is prone to hide his anger from himself and others by simulating a cool detached business-like manner in the situation, or the knowledge that he has some weighty if unconscious psychic interest in maintaining a veneer of tranquility, come what may.

Thus, mental-state reports are justifiably rejected in the present on the basis of evidence concerning the past or present. If "incorrigibility" is defined in these terms (i.e., as immunity from present falsifications) it is a more plausible candidate for distinguishing mental-states from mental events. For though mental-event reports are falsifiable in the future, it is an open question whether they can be rejected in the present, like mental-state reports. I return to this issue later.

7 "Incorrigibility as the Mark of the Mental", p. 40.
8 This falsifiability of sensation reports in the future though not in the present is exemplified in C.I. Lewis' "Empirical Knowledge", An Analysis of Knowledge and Valuation, Book II, Open Court, 1946.
(b) The "almost-incorrigibility" of Mental States: Can mental states be characterized as "almost-incorrigible"? Is it true, or relevant that "we are far less likely to have a report about a mental state overridden than to have a report about something physical overridden"? I think not. For, as I will argue, (1) there are many reports of mental states where rejection is far more likely and reasonable than that of most ordinary physical reports; and (2) in any case, variations of these relative frequencies of falsifications is a contingent matter in ways that do not affect the existence or boundaries of the mental. Let us begin with (1).

First of all, Rorty will have to relativise (1) to those mental-state reports which are at least non-inferentially credible if (1) is to have any initial plausibility. There are reports on mental states which entail physical reports and are as likely to be overridden as the latter. For example, consider the report "I am experiencing a fear resulting from the fact that my car has stalled again" which is at least as vulnerable to falsification as the report "My car has stalled again". If mental-state reports are to be "almost-incorrigible" in Rorty's sense, we must limit these reports to formulation in a basic vocabulary consisting of non-inferentially credible predicates. It is plausible to maintain that reports couched in such a vocabulary, e.g. "I love Mary", "I fear this dog", "I prefer wine to liquor", etc. are credible for the reporter independently of any evidence he has; furthermore, to maintain that no third party is in this position. However, it is clear that the non-inferential credibility of certain class of mental-state reports for the reporter does not imply that they are "almost-incorrigible". For, it is possible that one be justified in believing something mental without evidence, which nevertheless loses this epistemic status when certain evidence emerges, and with the same or greater frequency than in the case of physicalist beliefs. On this note, let us turn to consider the question of whether non-inferentially credible mental-state reports are "almost-incorrigible".

When an individual reports on complex mental states like that of love, hostility, pride, respect, goals, fears, ambivalence, etc., frequently such a report conflicts with knowledge we have of his behavior, situation, past reports or avowals, and his personality, or psychological background. Moreover, it is a matter of common knowledge in such cases that an individual has a vested interest in reporting and believing certain things about himself, in sustaining a certain image of himself which functions to rationalize his ways of acting and being in the world. Thus one's conception of himself may prevent him from seeing that his situation, behavior, feelings, and experience all evidence an anger or hostility which he sincerely denies, but is thoroughly obvious to others. Such an individual may even
provide verbal evidence of the anger he denies; he may "angrily" deny his anger, or "angrily" indicate that he has been treated badly, that he will never again subject himself to such a situation, etc. Yet he claims with perfect sincerity that while he morally "disapproves" of this treatment, thinks it "regrettable", he is neither outraged, angry, nor resentful of it. In such cases, one's mental report will admit of a much greater likelihood of falsification, or corrigibility, than many pedestrian physical reports one makes (e.g. concerning the weather).

Furthermore, in such cases we often reject an individual's reports not because we believe the reporter is lying, joking, being insincere, or verbally derelict. We take the reporter at his word; and assume he believes what he says, given evidence of his sincerity, verbal competence, etc. We reject his report, given its discrepancy with the other information we have about his state, on the assumption or knowledge that he is deceiving himself or is blind to the character of his own emotions, attitudes, beliefs, etc. -- a predicament not uncommon among human beings. Indeed, the person who reports moral attitudes which are incompatible with his behavior, or emotional responses, is often not simply a liar, hypocrite, or dishonest. If this were the typical case, the problem of moral education would be much simpler than it in fact is; the basic problem would focus on the virtue of honesty, whereas in fact, it turns on the broader more complex virtue of moral integration, or integrity.

Likewise, the complexity of the psychotherapist's task largely derives from the fact that bringing his patient to recognize, accept, and see that his state is one of hate, not love, is hardly a matter of getting him to stop lying, joking, or being insincere.9 It is noteworthy that within the context of dealing with neurotic and psychotic patients, many mentalistic beliefs and reports are treated by the psychoanalyst as, prima facie symptoms of their psychological condition and not simply as corrigible or false beliefs about it, although they are this as well. In any case, such contexts take us a long way from Rorty's model in which mental-state reports

are rarely falsified, and then only through their consequences for future behavior. In some cases of psychotherapy as well as profound moral change, a person may have his entire image of himself, a whole set of interrelated beliefs about his mental states, falsified. Such a process need not involve any references to future behavior but may revolve entirely upon the subject's own reinterpretation of his condition through the mediation of the psychotherapeutic transaction itself.

Consider this example of typical neurotic behavior. An undergraduate pre-medical student honestly reports that he wants to pursue a pre-medical course of study and enjoys it. Yet, he comes for psychotherapy with an array of symptoms and problems which evidence an extreme resistance and aversion to pre-medical study. He "can't" seem to get himself to study his pre-med courses, avoids them whenever he can, gets severe headaches when he tries to study pre-med material which subside when he studies other things, finds it difficult to awake for his pre-med courses and to concentrate in them. The student expresses confusion about his reasons for wanting to be a pre-med, which reduce to an "ambivalent" wish to please his parents. He also expresses some desire to rebel against his parents and act on reasons he has for pursuing a different course of study. Yet up to a certain point in his psychotherapy, despite overwhelming evidence to the contrary, the patient insists that he wants to continue as a pre-med and actually enjoys it.

In such conditions, the psychotherapist can and typically does justifiably conclude that his patient does not really want what he claims to want and enjoy; furthermore, the psychotherapist in certain cases can knowingly predict that the patient will come to see this for himself when and if certain fears, confusion, emotional conflicts, and experiential data are exposed and fully acknowledged by the patient. In my view, in such cases, the psychotherapist justifiably overrides the patient's mental reports, although they are crucial as evidence of the latter's real psychological condition.

While the existence and facts of such cases are not controversial, their philosophical and psychological interpretation is. One may argue in such a case either: (1) that the patient's report of his wants is correct (he wants what he says he wants e.g. to be a pre-med); though as he comes to see that this want is irrational, compulsive, and unhealthy in the context of his personality, he gives it up; (2) that the patient doesn't know what he wants, and in fact doesn't either want the one thing or the other, but is rather trapped by conflict and ambivalence between them (i.e., 'he does and does not want to be a pre-med'; 'part of
him wants to be a pre-med; part of him doesn't').

While the two accounts are accurate for some cases, to argue that they must be adequate for all such cases reflects an aprioristic philosophical dogmatism more concerned to preserve the incorrigibility of the mental, than to capture the realities of the cases themselves. In the case I have given, a reasonable interpretation of the evidence seems to exclude (1) the view that the patient does want to be a pre-med. The position formulated as (2) that he both 'does and doesn't' is more plausible. Indeed, even this position has the result that the patient's mental-reports are set aside, modified, or corrected by evidence, if not simply falsified. However, in such cases, the most adequate account seems to me to be one which allows the justifiable conclusion that the patient does not want or enjoy what he believes he does. This conclusion makes more sense of the totality of the patient's behavior, mental-reports, emotion, etc., than the alternatives.

In any case, it would seem that many a mental-state report is more likely to be falsified than any number of ordinary physical reports. But this brings us to my second argument: the frequency of falsification of mental-state reports or physicalist ones seems to depend on contingent and pragmatic considerations quite irrelevant to the existence of the mental. Let us suppose that aside from special psychotherapeutic contexts, mental-state reports are overridden "only rarely and with trepidation". Yet, the explanation seems to be that the sort of evidence about a person that is required is frequently not at hand, that correcting someone's mental-state reports can jeopardize friendships, social interests, etc., and that we typically find it easier to let such reports pass uncritically. In short, such de facto incorrigibility turns on our ignorance, naivete, sloth and practical interests, and not any epistemic privilege accorded the mental-state reports themselves. It is hard to believe that this de facto incorrigibility and pragmatism is what Rorty has in mind. Indeed, most ordinary physical reports enjoy the same degree of de facto incorrigibility. Likewise, in special context, e.g. psychotherapy and chemical experimentation, both kinds of reports are held up to a more critical scrutiny and more frequently rejected.

Rorty needs to distinguish "almost-incorrigibility" as a practical attitude towards falsifying mental-state reports, from this concept as an epistemic attitude towards prima facie disconfirming evidence against mental-state reports. He needs to say something like this: all pragmatic considerations aside, we less frequently consider ourselves justified in allowing prima facie disconfirming evidence to override mental-state reports than in the case of physical reports. I doubt this is true, but even if it were, I
cannot see that it affects the existence of the mental. Whether or not it is true depends on contingent facts like what mental-state reports are actually made, how strong the disconfirming evidence happen to be, and more generally, on the degree of inductive incoherencies between people's reports and other facts about them at any given historical point. The same holds for physical reports; the frequency with which negative evidence is allowed to rebut their initial credibility depends on what reports are made, the strength of the bodies of negative evidence, and the degree of inductive incoherencies between physical reports themselves, or between the reports and memory, empirical knowledge, etc.

It is hard to believe that the distinction between the mental and the physical could depend on what reports people actually make, how often there is strong evidence against these reports, and the degree of inductive incoherencies. These contingencies may and indeed do fluctuate without affecting our notion of what is mental and what is physical. Rorty's characterization of mental states as "almost-incorrigible" thus seems entirely inadequate.

(c) The Incorrigibility of Mental Events: Are mental events (thoughts and sensations) incorrigible, in Rorty's sense? Are there reports on these events which we do not allow to be falsified by any evidence whatsoever? Are there such reports for every "mental" event? My answer is negative.

As I have already suggested, incorrigible reports concerning one's own states depend on the availability of what I will call "non-inferential" predicates (e.g. "am in pain"), predicates which one can justifiably ascribe to oneself without a basis in evidence or further empirical knowledge; in short, incorrigible reports require non-inferentially credible predicates. The first problem for Rorty can be posed as a question: Is there any reason to think that for every mental event one wishes to report, there is available in our language some non-inferential predicate to characterize it (incorrigibly). To be sure, familiar mental events, sensations like pain, heat, nausea, red appearances, etc.--which are the stock examples of philosophers of mind--are characterizable in our language by non-inferential predicates (am in pain, am hot, etc.). But consider unfamiliar sensations which are not readily characterized by familiar predicates in our language; all predicates or descriptive terms available in our language for characterizing these sensations may exclusively exhibit an inferential, hence unequivocally corrigeble, credibility. Let us examine some examples.

Some bodily sensations appear to require a characterization which makes an essential reference to their typical causal conditions, e.g., 'the peculiar twinge I get in my arm when and only when the weather suddenly turns damp and cold' or
'the sensation I get in my nose when and only when I've just tasted hot mustard'. It appears that there may be no way of justifiably making an identifying report of such sensations without making empirical assumptions about the physical world (e.g. that the weather is and has been damp and cold when I believe I have the sensation which I report in these terms). In such cases, my report has inferential and corrigible credibility; e.g. it could be overridden by evidence which shows that the weather (or physical situation) through which I identify my sensation has not and does not obtain. The incorrigibility theorist may argue that in this case I could incorrigibly report such a sensation as 'the twinge I get in my arm when and only when I believe that the weather has suddenly turned cold and damp'. However, this characterization is not equivalent to the original one and I might well reject it; e.g. if you produce evidence that the weather is not cold and damp, I will retract my sensation report, be puzzled, and not recast it in the suggested way. There is no guarantee in a case like this that there is or must be some other term(s) available for characterizing my sensation which would be less vulnerable to the possibility of falsification than the initial one.

Consider the same kind of problem for visual sensations. Imagine a case in which one identifies the shape of a visual image or sensation as the distinctive shape produced when two spheres of clay collide at a certain speed; it is easy to imagine that a person involved in experimental work would identify many shapes of objects resulting from scientifically significant collisions in such a way. In this case, his reports will have a credibility based on empirical knowledge and thus exhibit corrigibility. If evidence were produced showing that objects colliding at this speed do not even stick together, let alone produce a new distinctively shaped object, then his report is undermined. Sense-datum philosophers assumed that for every sensation, there is a simple shape-color-texture description in the sense-datum language—which would yield a basic report plausibly considered incorrigible. But this assumption is dubious. If there are sensations whose identification is complex, and irreducible to the sense-datum vocabulary, then it is implausible to assume that in all such cases incorrigible reports are forthcoming.10 Philosophers have overlooked

10 For an elaboration of this point see my doctoral dissertation, Characterization of Phenomenalist Statements in Theories of Knowledge, Chapter V, esp. pp. 389-413, Johns Hopkins University (1969), or University Microfilm, Ann Arbor, Michigan.
this problem by concentrating on the cases of familiar sensations whose identification is easily and simply handled in our language. Every language favors the identification and ready characterization of some sensations over others; but, I shouldn't think that the mental status of the former is more secure than that of the latter.

The second problem arises specifically for the case of visual sensations. In the preceding paragraph, I have suggested that there are visual sensations which resist simple characterization by means of ordinary and philosophically familiar predicates. In some cases, one may have to characterize one's visual sensations in theoretical or semi-theoretical terms. For example, I may have to characterize what I sense or imagine as a Spyrogyrum undergoing mitosis, or a Hydrogen atom undergoing fission, etc. In such cases, the credibility of my report will depend on my knowing what a Spyrogyrum is, and what mitosis is. The incorrigibility theorist will maintain that this knowledge is conceptual or analytical, a knowledge of meanings or concepts (understanding "Spyrogyrum", "mitosis", "Hydrogen atom", etc.). Although Rorty wants to dissociate himself from an analytical-synthetic distinction on Quinian grounds, he will need it— as all incorrigibility theorists do; for, all reports are corrigible on the basis of evidence that the reporter doesn't understand the terms he is employing, lacks the conceptual knowledge in virtue of which we assume his report expresses what he believes.

In the case where my report employs theoretical terms, there is the additional complexity that knowing what some of these terms (cluster concepts) mean entails a knowledge of some empirical laws involving them; e.g. knowing what Spyrogyrum means or what a Spyrogyrum is entails knowing certain empirical laws concerning their properties. While this is a problem Rorty will have to confront, it is not the main problem I want to raise in this context. The complication of concern here arises from the fact that in the example before us there is a source of corrigibility absent in the familiar paradigm sensation reports. The credibility of my report that I sense a Spyrogyrum depends in part on my ability to visually identify Spyrogyra, on my knowing how they look. In my estimation, this visual capacity and knowledge is neither (1) knowledge of concepts (e.g. knowing the meaning of "Spyrogyrum") nor (2) propositional empirical knowledge (e.g. knowing that all Spyrogyra look x, y, z; this looks x, y, z; therefore,...). Rather, knowing how Spyrogyra look consists in the capacity to identify Spyrogyra on the basis of visual cues which need not be formulable as propositional evidence. Furthermore, having this capacity is not entailed by the mere possession of the concept of Spyrogyra; one might fully possess this concept...
and not know how to visually identify Spyrogyra (e.g. because one has never seen them or because one's capacity for visual discrimination of this sort is weak). This kind of possibility has been explicitly denied for concepts like "red", "triangular", etc. by incorrigibility theorists who argued that in these cases knowledge of the concept and the visual capacity in question are identical. Whether this is or is not true for the so-called sense-datum concepts, it seems clearly false for a concept like Spyrogyra.

The significance of this is that for sensation reports like "I sense a Spyrogyra", there is a source of corrigibility overlooked by incorrigibility theorists and their critics alike. In such a case even though my report is (1) prima facie non-inferentially credible and (2) does not involve any conceptual inadequacy on my part, nevertheless it can be overridden if others gain evidence that I cannot visually identify Spyrogyra, or that I do not know how they look. Such a possibility provides another reason why the non-inferential credibility of sensation reports does not entail their incorrigibility: namely, while such reports possess a credibility which doesn't presuppose that the reporter possess further empirical knowledge (one source of corrigibility), its credibility does depend on the reporter's visual capacities, an independent source of corrigibility. This problem will arise for all sensations which resist a ready characterization by means of sense-datum predicates (names of colors, shapes, or textures).

To give another example, suppose the only way to identify my visual sensation is to report sensing Nixon, or the present President of the U.S., or Johnson's successor, etc. In such a case, I may understand who Nixon is, or understand the meaning of "Johnson's successor" and yet not know how he looks, though I believe I do. Again, the ways that language permits me to identify my sensations will be corrigible through a knowledge that I lack a visual capacity not guaranteed by the mastery of language itself. If Rorty is to provide an adequate defense of his characterization of mental events in terms of incorrigibility, he will need to show that for every sensation (including visual sensations), there is some predicate in our language available for its identification invulnerable to the sorts of corrigibility argued here.

The third problem directly confronts the central question of incorrigibility head on. Are the traditional paradigms for reporting mental events, e.g. "I am in pain" or "I'm not in pain" incorrigible, even in the attenuated sense on which Rorty focuses limiting the question to the context of our present epistemic practices. My answer is no. I hold that every statement which reports a mental event is corrigible, or i.e., rationally rejectable if the
circumstances have a certain character, even within our present framework. In section III (a) above, I have argued that even the paradigmatically "incorrigible" reports of mental events are corrigible in the future, through memory and other indirect evidence. Though my report that I am in pain may possess an incorrigibility in the present, at some point in the future I may think I remember that I wasn't in pain, others may recall that I appeared quite all right, etc. Or, remembering full well what I felt and reported, I may later conclude that it was not pain but much more like a nervous giddy quality which I misinterpreted as pain. Such changes in what one believes he is feeling may come about when one repeatedly experiences the same feeling but with a clearer awareness of the situation in which it occurs, other aspects of its quality, and its immediate implications for one's behavior and mood.

Notice that other things being equal, the future falsification of one's initial sensation-reports depends on the subsequent course of his own avowals, and reinterpretation of what he believes. Thus, one might limit the incorrigibility of mental-events to their immunity from rational rejection by evidence other than that generated by the subject himself. The point I wish to argue here is this: if future avowals and behavior can override a mental-event report, then mental-event reports can also be overridden in the present. The simplest case is one in which the reporter himself overrides his own report in the present. As I suggest above, he may suddenly recognize that his sensation is identical to the giddy nervous feeling he typically gets in this sort of situation, and not pain at all. It is not clear whether an individual's present falsification of his own judgment concerning sensation is compatible with Rorty's incorrigibility, which appears to focus on third-person falsification.11

However, if an individual can override a present sensation-report on the basis of a knowledge of his past sensation and avowals, such knowledge may also justify others in rejecting his sensation-report in the present. Let us construct an example. We may know that R typically changes his mind about the nature of sensations he initially reports and believes to be pain. By itself, this knowledge will justify caution or skepticism but not a firm prediction he will change his mind in the present case. However, further suppose that we also know in which special situations he issues the reports

11 That is Rorty never discusses the rejection or falsification of a mental report (or belief) by the reporter himself; he assumes that it is always others who falsify one's mental report.
which he later changes, and we even understand why. The situations are ones in which (1) his behavior gives some evidence of twinges of nervousness and (2) nervousness is associated with a kind of weakness or vice he especially fears and others in this context especially demean. We also know that when in other situations later on which are less "threatening", R reflects back and decides that it was nervousness and not pain after all, though he is struck by the fact that he was convinced it was pain at the time. In this case, we can claim to know that under extreme pressures to appear calm to himself and others, R will falsely but sincerely interpret flashes of nervousness in his stomach as pain. When he makes his sensation report in this context, we can in the present justifiably override it. In these examples, the reporter's own disavowals play a central role in establishing the corrigibility of his avowals.

I now want to argue that there are other more radical examples in which we justifiably override a person's sensation reports without relying on his disavowals; these examples also show that Rorty is mistaken in his claim that behavior is irrelevant to the falsification of mental-event reports. Consider the following three situations: (1) we see a person R jerk his hand back quickly from a red hot burner, wince, groan, and shake his hand in the air; (2) we observe R jump back and gasp as a dog suddenly comes up behind him and barks; (3) we see R blink and move his arm up in front of his eyes, as someone shines a bright light in his eyes. Suppose R reports (1) that he feels no pain, (2) that he experiences no sensation of fright, and (3) that he does not sense a bright light. Unless we have evidence of an unusual sort that R is pretending, we will without a reasonable doubt assume that we know that R has experienced pain, fright, and a sensation of bright light.

Faced with such examples, incorrigibility theorists argue that the reporter has either lied, misspoken, or exhibited conceptual inadequacy; in effect, they grant the corrigibility of the sensation report while shielding the incorrigibility of sensation beliefs. However, in my view, in such cases

---

12 In "Mind-Body Identity, Privacy, and Categories" (Review of Metaphysics, XIX, 1965), Rorty adopts such a position. There he argues that in the case of sensation reports, "we do not have a genuine contrast between misnaming and misjudging". He concludes by adopting the standard position that overriding a person's sensation report is tantamount to the assertion that he lacks a grasp of the concepts involved, that his report does not correctly express what he believes in such cases.
we override R's report before and independently of knowing why he has issued the report, i.e., without knowing or assuming that he has lied or misspoken or exhibited a conceptual inadequacy. Indeed we may have every reason to believe that R knows what pain is, hasn't misspoken, and does not lie. We will be puzzled and wonder what R really believes, and if he really believes he felt no pain, why he does and how he can believe this, given that it is evident that he did feel pain. In short, the rejection of R's report is not necessarily premised on the assumption that his report does not express what he believes.

But, the incorrigibility theorist will argue, how can a man who is in pain believe that he isn't; what explanation can we give of his report other than explaining it as a lie, verbal error, or conceptual confusion? The basis for an answer to this question is suggested by some of my earlier examples: it becomes the question of what motivates someone to rigidly maintain false beliefs about his own experience. Let us further consider the above example in which R exhibits characteristic pain-behavior in a pain-situation but denies his pain. Suppose we learn that R intentionally thrusts his hand onto hot objects and instructs others to burn parts of his body, because he wishes to prove that he is invulnerable to pain; he is a Christian Scientist and believes that the authentic virtuous Christian will not suffer physical pain. We also learn that R believes that lying and pretense in general is a cardinal sin, and indeed is known for his extreme honesty and forthrightness. On the other hand, a knowledge of R's personality also reveals a generalized religious psychosis involving hallucinations, self-destructive behavior, and a pattern of denying certain sensations and emotions so clearly evident in his behavior; fear, anger, pain, melancholia, self-doubt--states which R sincerely believes he should not have and does not have, as the Christian Scientist he is. One may be tempted to conclude that R lacks our concept of pain, but this would be incorrect. For, apart from the fact that he grasps the grammar of pain in the normal way, we could not explain his personality and behavior as a whole without assuming that he has beliefs about pain and knows what it is; his struggle to be a person who is invulnerable to pain presupposes that he knows what it is, what induces it, how people respond to it, and what it feels like. We can even imagine that he offers honest if fantastic explanations of his own pain behavior compatible with an absence of pain. Indeed, suppose he groans continually until we administer a pain relieving medication; nonetheless, he sincerely and honestly insists that these are responses of his body, like the pain behavior, which are unconnected to what he feels and thinks. He is a dualist for certain sensations in his own case.
Philosophers and psychologists have long recognized that human beings maintain false and fantastic beliefs about the world in accordance with motivations which belie the image of man as rational animal. Yet, at least philosophers have failed to realize that practically any belief, including beliefs about the character of one's "immediate" experience and mental states, may be implicated in this process of motivated distortion. The example I have given involves the kind of radical alienation of a person from his sensations that we associate with madness, the madness of the witch, the extreme sadomasochist, the religious fanatic. Yet, there are more normal and mundane examples of the same phenomenon. For example, ordinary people honestly deny sensations of fear in situations where fear evidences a kind of cowardice, professional insecurity, or character defect all of us are anxious to avoid. We often say that one "convinces himself" that he does not feel fear, to indicate that this is no simple case of lying, or pretense that is involved. Some have described this situation as one of "self-deception" or "inauthenticity" or "repression" to indicate that at some level or in some sense, one knows precisely what one cannot and does not admit at the level of conscious, public belief. It is not my purpose here to review and discuss various psychological and philosophical conceptions of belief. But, my examples exhibit the fact that in certain cases, the questions of what a person believes, even about his most immediate experiences (pain, or fear), requires a body of empirical knowledge about him, as well as theoretical assumptions about the structure of personality. In these cases, a person's situation and behavior conclusively establish that he experiences a sensation, even though his report denies it; yet, the facts of the case do not permit the easy explanation that his report embodies a lie, a slip of the tongue, or a conceptual incapacity. Such cases establish that sensation beliefs, as well as sensation reports, are corrigible; although, in the difficult cases, we override the report quite independently of overriding the belief. For, while we can know from the agent's situation and behavior that his report is false, knowing what he believes requires much more background information and theoretical assumption.

This kind of case constitutes a decisive refutation of Rorty's position on the incorrigibility of sensation reports.13

13 It should be noted that other philosophers who have argued for the corrigibility of sensation reports and beliefs have relied on the sort of neurological or materialist developments which Rorty employs to eliminate the mental. For examples, see G. Sheridan, "The Electroencephalogram Argument against Incorrigibility" (American Philosophical Quarterly, 6, 1969). For Rorty, such arguments depend upon developments which initiate the materialist epoch and eliminate mental reports. In this context, the significance of my arguments emerge: the situations which I construct in which sensation reports and beliefs are overridden occur within our present epistemic framework and do not presuppose further scientific developments.
For, consider an account which attempts to conclude in this sort of case where R reports an absence of pain, there really is no pain, despite the evidence of situation, behavior, response to medication, etc. We may even consider cases in which we legitimately conclude that there is no pain, and that R's report does express what be believes. Surely, in the given circumstances, the report by itself would not be sufficient to justify this conclusion. At the very least, we will need to produce evidence which establishes that R's report expresses what he really believes, evidence which must also explain how he can be in a pain-inducing situation, exhibit pain-behavior, and show positive responses to medication if he has no sensation of pain. In short, in such a situation we may conclude that R is not in pain in accordance with his report, but only after we possess an empirical argument with empirical assumptions which rule out lying, conceptual confusion, etc. and which explain his situation, behavior, etc. in a way compatible with an absence of pain. Notice that the necessity of this procedure already implies that neither R's report, nor the unproblematic ad hoc assumption of what R believes, will suffice to establish the character of R's sensations. We will require an empirical argument with empirical assumptions to establish what R believes and what he really senses; in precisely the same way that we require such an argument with such assumptions to establish what R believes and what is true of the physical world in a similar situation where his report on the physical world flagrantly contradicts what we see right before our eyes. The only difference between these two cases is that in the physical case, there are ways of explaining false perceptual beliefs (e.g. illusion, hallucination, faulty vision) that involve less complex psychological assumptions than the ways of explaining false sensation beliefs. But, both are corrigible; whether we justifiably accept or falsify a report in these hard cases, our knowledge will depend on a corrigible empirical argument. Thus, in one respect a person's empirically privileged access to the contents of his mind is similar to an empirically privileged access to the physical contents in one's perceptual environment; both are contingent on certain though different psychological capacities of the person. In their absence, this privileged access is cancelled and we explain the resulting errors in terms of distortion, whether it is a consequence of the faculties of perception (as in errors about physical world), or of deeper dimensions of personality and character (as in errors about one's mind).

This line of argument does not simply establish that first person sensation beliefs are corrigible. It also suggests that the prima facie plausibility of holding that sensations are incorrigible objects (in Rorty's sense) is ultimately undermined by the fact that beliefs are corrigible objects, a
fact not disputed by Rorty. For, it is only plausible to accord incorrigibility to one's sensation report on the knowledge that he maintains a corresponding belief, a knowledge that is clearly corrigible; putting the matter differently, in effect our knowledge of an individual's sensations is mediated by our knowledge of his beliefs. Thus, in the hard cases I have developed, though an individual assures us that his sensation report expresses what he believes and believes it himself, any doubts the situation forces on us concerning what he really believes will necessarily affect our doubts concerning what he really feels or senses; the corrigibility of belief implicates sensations, robbing them of any incorrigibility we might have thought they have. Indeed, once the situation forces us to deal with the problem of what the reporter really believes, we need to learn much more about him, his character traits (honesty, capacity for self-knowledge, etc.), his structure of motivation, etc. Thus, coming to know whether to accept or reject the most simple sensation report may require that we gain a knowledge of the most complex aspects of the reporter's character and personality.14

At this point, I shall note that in my estimation, the same argument can be mounted against the incorrigibility of reports of thoughts. I will not develop the argument but simply suggest an example which will provide an indication of its structure. One may sincerely deny one has a thought(s) of Mary, who one is infatuated with, much to one's chagrin, embarrassment, shame, etc. Yet, suppose we see this individual copying Mary's name or writing her phone number over and over in one's notebook, as he gazes out the window with that forlorn look common among adolescents. Suppose, we have evidence that Mary has been on the individual's mind, though we also know that his shame, guilt, fear, etc. provide powerful motivation for him to deny these thoughts in ways that are thoroughly sincere, honest, and expressive of what be believes. He may be generally aware that he is perturbed, that something is happening, that Mary has something to do with it, etc. Yet, at any given moment, he is unable to recognize or acknowledge the character of his thoughts, to believe that they are thoughts of Mary; he is quite certain that Mary is the furthest thing from his mind. Such a case is quite similar to those presented above and generates the same sort of argument against the incorrigibility of mental entities.

14 Incorrigibility theorists and Rorty himself thus mask the problem of mentalist knowledge when they argue for the incorrigibility of the mental report, on the assumption that it expresses what the reporter believes. For, in the hard cases, it is precisely this assumption that is problematic and whose validation or falsification requires a corrigible empirical argument.
I do not intend to suggest that the typical mental report of pain or its absence generates these problems, any more than the typical report on the physical world in one's field of vision generates these problems. Both enjoy a de facto incorrigibility which stems from the fact that in such cases there is no counter-evidence, no reason to question either kind of report. Nonetheless the troublesome conflict cases reveal an epistemological dimension implicit in the ordinary cases as well: namely, that our epistemic attitude toward the most simple, least problematic sensation (or thought)—reports depends on rather complex empirical assumptions concerning the reporter, and general personality traits. We assume not simply that the reporter knows what pain is, is honest, sincere, etc.; but moreover, that he is free from the structures of motivation and personality which lead people to maintain false or unreliable beliefs about their sensation or thoughts, and beliefs contradicted by their situation, behavior, and responses. Although it is normally reasonable to make this assumption about a person even in the absence of specific knowledge about him, the fact that this assumption is always open to empirical disconfirmation makes all reports on mental entities corrigible, in Rorty's sense. All we need to learn is that the reporter exhibits a certain psychological background and history of radical alienation of belief from sensations, and the prima facie non-inferential credibility of his report becomes its opposite, a report that is falsified and overridden.15

There is one final point to be made concerning the epistemological and perhaps logical connection between many mental events and mental states. Many mental entities are identified and known by a person as parts or aspects of a mental state he knows himself to be in. To return to an example given earlier, if a person R knows that he is in a nervous, jumpy mental state, then he will justifiably identify a twinge in the stomach as butterflies or fluttering, other things being equal (e.g. he isn't suddenly struck in the stomach); on the other hand if he identifies his mental state

15Thus, in effect, my argument against the incorrigibility of sensation reports (and reports on thoughts) does not simply apply to some reports in unusual cases, but affects all sensation reports in all circumstances. To my knowledge, no philosopher has ever suggested that every sensation report is corrigible because it depends on certain contingent empirical assumptions concerning the reporter's personality and character. Notice that the ordinary case in which we routinely assume that a person is not blind to his own sensations only differs in degree from the ordinary case in which we routinely assume that a person is probably biased in judgments concerning his motives, real reasons, or intentions in performing a self-interested but morally dubious action.
as a severe stomach ache, he will justifiably identify a
twinge as one of pain, other things being equal. To take
another example: I can identify the orange that pops into my
head as the thought 'would that I had an orange' because it
occurs in the context of a self-conscious desire all day for
a piece of fruit; on the other hand, if I have had my fill of
fruit but am consciously thinking all day about getting away
to the tropics, then I identify an orange which pops into my
head as 'would that I were in a place where oranges grow'.
Finally, I can identify a sensation as a blush because I know
that I am embarrassed by the clothing I have on in the
presence of these people.

In these cases, one identifies a particular sensation
through a mental state one knows one is in. Consequently
there is at least this epistemological connection between such
sensations and mental states: if others can establish that
the one is not in the mental state one thought one was in,
this undermines one's sensation report as well.16 Thus, we
may establish that one's state was not embarrassment but
rather fear, and the sensation was not a blush, but rather a
surge of fright. The corrigibility of mental states and
mental events is essentially interrelated in such cases.
Rorty is thus mistaken in his empiricist assumption that mental
events are isolated atomic units whose character is prior to
and independent of the mental states and conditions in which
the former are typically embedded.

It is possible that Rorty will argue that in my examples
the mental states provide causal evidence of the sensations in
question, but that correcting R's belief about his mental
state leaves his sensation report unaffected.17 I of course
disagree with such an argument. Furthermore, it seems to me
that in the examples I have presented there is a logical
connection between the sensation and mental states, which
underlies the epistemological connection. In my example, R's
sensation is a blush only in so far as it is a constituent
of a larger whole, the state of embarrassment, without which

16 This directly contradicts Rorty's view that "as such mental
features (i.e. states) or beliefs and desires become more
particular and limited and thus approach the status of episodes
(i.e. events) rather than dispositions, they become more
incorrifiable" ("Incorrigibility as the Mark of the Mental;
p. 420). In a sense, my claim is that he puts the cart before
the horse.

17 I.e. Rorty might argue that a mental state causes a given
sensation but that nonetheless the sensation is logically and
epistemologically independent of the sensation.
it would lack its identity. Thus, when we or R come to believe that his state is definitely not one of embarrassment, contrary to R's initial belief, on this basis neither of us continues to identify the sensation as a blush. Indeed, it appears that one cannot experience a sensation (e.g., in the face or ears) as blushing unless he is embarrassed at something (a mental state). I am not arguing that this is true of all sensations; but it is true of many and for these, we have located an additional source of the corrigibility of mental-event reports.

IV

I have argued that the mental within our existing epistemic framework cannot be characterized in terms of incorrigibility. However, from the fact that mental reports and beliefs are neither incorrigible nor almost so, it does not follow that they have no special or essential role within our established ways of knowing persons. In all of the examples I have presented, while particular mental reports and beliefs are overridden, the knowledge (evidence) which is required in order that this move be justified, depends on a knowledge, or acceptance, of other mental reports of various sorts: (1) in some cases (cases of self-correction), the agent's own later disavowals; (2) in other cases which require knowledge that the agent is honest, or involved in self-deception, or mad, those of the agent's mental reports necessary to establish such judgments of personality and character; and (3) in those cases which require a general knowledge of human personality and motivation, the mental reports which ground this general knowledge.

This suggests that the mark of the mental may be the following: not that we never or rarely override certain reports (and) beliefs about the mental, but rather that whenever we justifiably override such reports, we do so only by accepting some other such "mental" reports at face value. However, when materialism triumphs, we will justifiably override such reports without any dependence on mental reports at all. This formulation does not appear to yield a "mark" of the mental for it requires a reference to the mental which seems ineliminable. Let us see.

Consider first Rorty's definition of the mental and secondly, the one suggested by my remarks:

(1) If there is some person who can have an incorrigible belief in some statement P which is a report on X, then X is a mental event.

(2) If there is some person who can have a belief in some statement P which is a report on X and which can only be falsified if certain other of his beliefs in statements reporting events or states are accepted, then X is a mental
Both of these formulations define the mental in a way that presupposes it by making an essential reference to belief, a mental notion. Apart from this appearance of circularity, belief causes real trouble for (1) as my earlier arguments suggest. Let us consider the trouble it causes for (2). We may begin by asking whether (2) also holds for physical events and is thus incorrect. It appears that it doesn't; for, one reason, a person R's belief that it is snowing can be falsified without requiring evidence that depends on accepting other beliefs of his as true. However, matters are not so simple, as can be seen by comparing this to the "mental" case. Can we falsify S's belief that he is not in pain simply on the basis of a pain-situation, behavior, etc. or need we rely as well on accepting some of S's report-beliefs? Recall my earlier argument. We can falsify his report without falling back on an acceptance of any other of his reports—simply on the basis of situation, behavior, etc.; but, on this basis, we cannot assume that his report expresses what he believes, that we have falsified his belief. Furthermore, to establish that his denial of pain expresses what he believes (not a lie, joke, conceptual error, etc.), we will need to gain a knowledge of him which will require that we accept some of his reports as expressing what he believes.

Now, I think the same can be said for the physical case. We can falsify R's report that it is snowing without accepting other of his reports-beliefs, but not his belief that it is snowing; as in the "mental" case, establishing that he believes this will itself require an acceptance of and appeal to other report-beliefs. Thus belief causes trouble, both for Rorty and for the present suggested definition. We can establish that what a person says is false, whether it is about the mental or the physical, without relying on the kind of knowledge about him which itself presupposes an acceptance of some of his reports; but we cannot in this way establish that he stands in a certain relation to the things or noises he utters, i.e. that it is assertion, belief, a lie, or whatever. Thus, finding a mark of the mental flounders on the necessity for presupposing belief, itself mental and troublesome.

V

The fact that mental reports, though neither incorrigible nor almost so, play an essential role in our way of knowing persons does not yield a mark of the mental but it does suggest serious problems for Rorty's view of the scientifically sufficient conditions for the elimination of the mental, and the triumph of materialism. My examples have emphasized the fact that even when we override a person's
mental reports and/or beliefs, they pose a problem for us, a
data to be explained; we ask, given that R is in pain, why
does he say he is not in pain? Is he joking, lying, fooling
himself, blind to the nature of his own sensations, and if
the latter, why--how come? Furthermore, in providing an
empirical explanation to answer such questions, we cite other
mental states-beliefs, desires, fears, compulsions--of R.
What this suggests is that at present the reports, events, and
states of persons characterized in mentalistic terms are
essentially involved in the explicanda and explicans of
empirical knowledge. Let us see how this affects Rorty's
view of materialism which eschews reductionism and stresses
the elimination of the mental as a function of the elimination
of mentalistic language. Rorty presents a conception of the
developments within science which would be sufficient for the
rationality of accepting materialism, if not its de facto
acceptance and consequent truth. In my estimation, there
are two serious lacuna in his conception, two gaps in his
argument which I will refer to as (1) the problem of the
ineliminable explicanda and (2) the problem of the ineliminable
explicans of scientific explanation.

(1) Some segment of our vocabulary or conceptual scheme,
e.g. mentalistic language, is presently ineliminable if it
is required to characterize what there is to be explained
by empirical inquiry. Rorty fails to deal with the problem
of how we determine, radically alter, or eliminate what there
is to be explained, the 'recalcitrant' descriptions of data
in whose terms we feel obliged to evaluate the scope, accuracy,
fruitfulness, predictive power, simplicity, practical value,
and explanatory force of scientific theories. In particular,
Rorty fails to consider the fact that at present part of what
there is to be explained are (1) mental events and states,
and (2) our true reports, lies, self-deceptive reports, and
contradictory reports about (1); to put it differently, the
present agenda of scientific inquiry is firmly constituted
by certain events, states, and verbal behavior mentalistically
characterized.

This failure undermines Rorty's argument that materialism
would become rational if and when mentalistic language and
entities are no longer required to explain behavior, because
physicalist (e.g. brain or neurological) entities and language
will accomplish this task in a superior and exhaustively
adequate way.18 Yet, if all behavior were explained

18 For this argument, see "Incorrigibility as the Mark of the
Mental", pp. 421-424 and "Mind-Body Identity, Privacy, and
Categories". Indeed, Rorty's one-sided emphasis on behavior
as the crucial explicanda of science and the fact that he
doesn't worry about the relation of activity and action to
behavior, both already reflect a materialist bias.
physicalistically, this would still leave science with the need to explain mental entities, i.e. thoughts, sensations, intentions, desires, etc. which are not (as Rorty correctly holds) eliminable to behavior. At present, mental entities (or, entities characterized in a mentalistic language) are not solely "theoretical" in the sense that the exclusive or major justification for assuming their existence resides in their explanatory value, viz. behavior. Entities characterized within a mentalistic vocabulary constitute a prima facie ineliminable explicanda of scientific explanations. Indeed, there is also the question of whether the explanation of behavior, in Rorty's sense, is also the explanation of action and activity (e.g. telling a lie), or whether the latter also remains as a prima facie ineliminable mentalistic explicanda.

These questions I raise also undermine Rorty's other major materialist argument to the effect that materialism would become rational if and when mental reports are no longer incorrigible but rather mental entities are routinely and conclusively known physicalistically, i.e. by observation of the physical (e.g. through a neurological or brain state meter). Suppose the neurological meter informs us that Jones has a thought of his mother, even though Jones reports that he isn't thinking of anything. There are at least two facts here to be explained which appear to require a mentalistic vocabulary; we want to understand why Jones is thinking of his mother and why he has lied about it to us. There is no reason to believe that the corrigibility of mental reports, any more than the dispensibility of mentalistic assumptions in explaining behavior, should eliminate the need to explain facts such as these which prima facie require a mentalistic characterization.

It is clear that Rorty will have to argue that these facts (e.g. the facts that Jones is thinking of his mother and lying about it) need not be characterized mentalistically in order to be explained. This kind of argument is in keeping with the spirit of materialism; however, such an argument seriously complicates the simple picture Rorty and other identity theorists give us of the sorts of scientific developments sufficient for the rationality of materialism. Let us consider this complication.

(2) The problem of the prima facie ineliminable mentalistic explicanda is closely associated with the problem of the prima facie ineliminable mentalistic explicans. Consider the above example. The neurological meter has told us that Jones is thinking of his mother. Or, to put it differently, that Jones is in brain state n. Materialism would require that this event can be explained without mental language, or
explained under a physicalist description. However, given our present paradigms of explaining such facts, explicans which explain the fact that Jones is in brain state \( n \) are not taken to explain the fact \( m \) that he is thinking of his mother. In the normal case, we require a rational or teleological explanation which itself requires a prima facie ineliminable mentalistic explicans; e.g. Jones believes his mother is about to die, he desires to see her once more, etc. Granting Rorty's premises about scientific progress, our neurological meter could conclusively inform us as to whether the premises of this explicans are true, whether Jones indeed possesses the alleged desires, beliefs, etc. Nonetheless, given our present explanatory paradigms, a complete neurological description of the thought event, its "causes-reasons", and some covering law will not constitute the basis for explaining the thought.

As long as we need a mentalistic vocabulary to characterize what there is to be explained and what there is which will explain it, the scientific developments which Rorty envisages will not be sufficient for the rationality of materialism. In place of Rorty's over-simplified picture of a one-dimensional scientific progress involving mental-physical correlations, he will need to discuss the conditions under which our present teleological paradigms of explaining many human phenomena would be replaced or replaceable by one compatible with materialism.\(^\text{19}\)

In this context, the debate concerning materialism awaits instruction from Kuhn's remarkable conception of scientific revolution in his *The Structure of Scientific Revolutions* where he suggests the complexity of the causes and reasons which are involved in the transition from one paradigm, one set of fundamental categories of thought, to another. Some of his observations are relevant here. For one thing, his account suggests that the transition to a neurological paradigm would grow out of serious crises ("anomalies") within the psychological-mentalist paradigm as judged in its own terms; crises, for example, which would effectively thwart the use of psychological paradigms in doing the kinds of research definitive of the disciplines as already constituted. Secondly, Kuhn stresses that paradigm shifts involve dramatic changes in theoretical interests and purposes, criteria of adequacy, the importance assigned to particular sorts of problems and specific domains of nature.

It is thus remarkable that philosophers have talked about the shift to materialism without dealing with it as a profoundly revolutionary shift in paradigms involving precisely the kinds of elements Kuhn has isolated in the history of science. On the other hand, perhaps it isn't so remarkable,

\(^{19}\text{One implication is that the materialist debate should deal with the literature on the explanation of action and action-theory in general; in particular, the arguments concerning the difference/...}

F-6
given that Kuhn's perspective tends to undermine and complicate the unproblematic concept of explanation required for the coherence of the materialism discussion among philosophers up to this point. One sometimes suspects that it is a positivist model of explanation that underlies the materialism discussion: explanation as prediction, as nomological subsumption.

In any case, within a Kuhnian perspective, is there anything we can say about the existing mentalist paradigm of explanation, the purposes, interests, priorities, which it embodies which would cause and-or justify a resistance to a shift to a materialist paradigm? I will attempt to locate one kind of consideration which can be usefully introduced into the materialism debate. At present, one of the ruling purposes of personality psychology (as one mentalist paradigm) is a productive psychotherapeutic practice, and more generally, productive child-rearing, educational practice, socialization processes, etc. That is, our explanatory paradigms in psychology embody a commitment to solve certain kinds of problems--theoretical and practical--which occur in the growth and development of human beings. While personality psychology, like all science, is thus concerned with control of nature--in this case human nature--like any particular science, it is concerned with certain kinds of control for certain ends, not with control per se as an abstraction. In the case of the theories of personality, social character, abnormal psychology, the kinds of control and ends which are central involve an implicit commitment to self-development whenever possible through deliberation, self-understanding, criticism, interpersonal communication, social transactions, etc.

Certain philosophers and psychologists have argued that these forms of human control require a kind of understanding and vocabulary which presuppose a mentalist paradigm and is incompatible with a materialist one. This may well be true, but it is hard to prove that it is or to understand why it should be so. Part of the problem arises from the difficulty in concretely envisaging a totally materialist science of man and the sorts of interests, purposes, control, and ends it would distinctively involve. Thus, it seems dogmatic to insist a priori that problems of human growth, development, and change could not be handled with the same emphasis on autonomy, deliberation, communication, etc. as that at the heart of existing mentalist paradigms. Nonetheless, this dimension places a burden of proof on the materialist to suggest how human interests, purposes, and ends central to mentalist paradigms can be expected to fare within a materialist scientific framework.
This line of discussion suggests another criticism of Rorty's conception of materialism. Rorty maintains that if mentalist language becomes unnecessary for "the" purposes of scientific explanation, materialism becomes rational to accept. In his view, the further condition required for its truth is that ordinary people in fact adopt this materialistic language of science. Whether or not this development occurs turns on "a preference for Occam's razor", it is "close to being a matter of taste". Elsewhere Rorty and other eliminative materialists have suggested that once mentalist language is dispensable for the purposes of science, the only obstacle to materialism is the "inconvenience" of speaking the language of materialism, a practical obstacle with no theoretical significance.

I find that these views express a naive scientism which assumes without argument that apart from scientific explanation, moral, aesthetic, educational, and everyday discourse and the activities which require it, do not place any distinctive demands on language other than those which reduce to matters of "taste" or "convenience". Is this the old positivism of scientific vs. emotive discourse rearing its ugly head? What reason is there to think that literature, philosophy lectures, news broadcasts, moral argument, etc., could fulfill their distinctive purposes employing their distinctive methods without mentalist language just because science could? Are these activities beyond the sphere of the rational; if the way they are carried on is an issue of "convenience" and "taste", why can't we say the same for scientific activity? Perhaps Rorty and materialists assume that science is the exclusive province of truth, that all other discourse does not really refer; but, then this metaphysic needs to be explicitly endorsed and defended.

In summary, the rather simple and straightforward scientific developments Rorty and other materialists envisage fall far short on a number of counts from establishing the rationality of materialism--either as a scientific outlook or a worldview. On the one hand, eliminating the mental (and establishing materialism) cannot be a matter of discarding incorrigibility or privileged access in Rorty's sense, because it does not really exist. On the other hand, it cannot simply be a matter of the one-dimensional discovery of mental-physical correlations however uniform, predictively adequate, etc., because this might occur and leave our commitment to a mentalist-psychological paradigm of explanation thoroughly intact. Furthermore, a need for mentalist language is implicit in forms of human life and activity which involve their own ontological commitments and whose peculiar resources

---

20 "Incorrigibility as the Mark of the Mental", p. 423.
are not reducible to the resources of science.

Finally, it is not remarkable that there appear to be no identifiable, non-circular, independent "mark" of the mental; for mental categories are so endemic to our established forms of thought and life that as soon as we attempt to define or eliminate them they reappear in our presuppositions and inferences. Rorty tells us of a thoroughly materialist world where we "correct" one's "beliefs" about himself through our "perception" of cerebroscope readings. Yet, we must envisage this process without beliefs and without perceptions, two entities whose present character requires "mentalese". And, we must be careful not to take our talk of "correction" here too seriously. But even if we succeed in achieving a semblance of the required stringency of language in characterizing what we do in the world of materialism, puzzles will arise which evoke the phantom of the mental.

To begin with, if a given state is physical—a state of my body (e.g. the physical correlate of pain or sensing red)—why should I have a non-inferential knowledge of it while others need to infer my state from the cerebroscope? One tends to think in a real materialist world, I would learn about my bodily states like everyone else—by reading the meter! Yet, it is clear that Rorty assumes people would stand in a special relationship to their bodily states; one will issue non-inferential reports on one's bodily states. But of course not all bodily states; I will be granted a non-inferential access to my pain-correlative but not to my physical state of stomach acidity. When we ask ourselves why within a materialist world people will still have non-inferential access to their own "bodily" states (inner states?), to some rather than others, we think it is because they continue to experience or feel these states. We fall back on a picture of the privileged physical states as mental, objects of sensation and experience.

Furthermore, it is clear that there must be a special mode of access to certain privileged "bodily" states if there is to be any empirical knowledge at all in this materialist world; for, our knowledge of the physical states of persons and the external world will depend on perception, but perception will require that the perceiver have non-inferential access to his "perceptual experience-correlative" physical states. As in our present body of empirical knowledge, in the materialist framework the buck will stop with the fact that certain people know without evidence that they are in certain "perceptual-experience" physical states. Of course as in our body of knowledge, people can always override any particular person's non-inferential claim to be in "perceptual-experience" physical state S, but only because they non-inferentially
accept their own "perceptual-experience" states as evidence of cerebroscope readings.

Thus, the materialist framework will require a special relation between people and certain physical states; not just their bodily states, but certain bodily states. People will be said to just know when these special "bodily" states obtain. Has the mental disappeared? Suppose a person R non-inferentially reports on one of these special states, say "pain-correlative" state p telling us he is in it. Then he non-inferentially reports on another one, a "perceptual-correlative" state c which informs him that the cerebroscope says he is not in state p, a conclusion which also squares with our judgments as well. We conclude he is not in p. But R insists he is in "pain-correlative" state p, a conclusion whose non-inferential credibility he prefers to others' inferential claim that he is not in p and his own indirect evidence against p from the cerebroscope. Perhaps, he is wrong, but his position is reasonable; given his situation, its asymmetry to others' situation, why should he give up a report which is as firm from his point of view as those on which all knowledge is based—even supposing that is the reasonable move for others to make. Though we are materialists, deep down we understand his point of view. Even if the scientific community legitimately overrides his report, we understand that for him in such a case it is a different story. This is "his" state—it exercises constraints on what he can do which the rest of us are free of. Etc. Has the "mental" really disappeared? One can say what one pleases, but in such a case who can resist the thought that our man R has experienced pain, felt it—and can't get it out of his mind, any more than we can get the mental out of ours.

Gerald Doppelt
Department of Philosophy
University of California,
San Diego
La Jolla, CA  92037