

Journal of Philosophy, Inc.

Incorrigibility as the Mark of the Mental

Author(s): Richard Rorty

Reviewed work(s):

Source: *The Journal of Philosophy*, Vol. 67, No. 12 (Jun. 25, 1970), pp. 399-424

Published by: [Journal of Philosophy, Inc.](#)

Stable URL: <http://www.jstor.org/stable/2024002>

Accessed: 06/01/2013 20:02

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Journal of Philosophy, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy*.

<http://www.jstor.org>

THE JOURNAL OF PHILOSOPHY

VOLUME LXVII, NO. 12, JUNE 25, 1970

INCORRIGIBILITY AS THE MARK OF THE MENTAL

IN this paper I argue, first, that various “topic-neutral” translations of mentalistic statements propounded by materialists are unsatisfactory in that they do not catch the specifically “mentalistic” element in these statements. I then go on to argue that to isolate this element one needs to insist on the incorrigibility of first-person reports of mental states. Finally, I consider whether this insistence is an obstacle to materialism.

We may begin by recalling that the origin of the attempt at “topic-neutral” translations of mentalistic statements was an attempt to avoid what we may call the “irreducible-properties objection” to the thesis that mental states are identical with brain states. This objection says that, even if the identity thesis frees us from nomologically dangling entities, it cannot free us from nomologically dangling properties—viz., those properties by which we originally identified the mental entities as such. Thus, for example, a sensation of yellow has the property “of yellow,” and the thought that *p* has the property “that *p*”; but it seems to make no sense for any brain process to have either sort of property. So these properties seem irreducible. J. J. C. Smart originally tried to get around this objection for the case of sensations by saying that “I am having a sensation of yellow” was equivalent to (or could roughly be paraphrased as) “Something is going on in me like what is going on when I see something yellow.”¹ More recently, D. M. Armstrong has employed the same technique in a program of translating (or paraphrasing) all statements ascribing mental states as statements containing the subject term “a state apt for the production of the following sorts of behaviour”.²

¹ “Sensations and Brain Processes,” *Philosophical Review*, LXVIII, 2 (April 1959): 141–156; reprinted in V. C. Chappell, ed., *The Philosophy of Mind* (Englewood Cliffs, N.J.: Prentice-Hall, 1962), pp. 166–167.

² *A Materialist Theory of the Mind* (London: Routledge & Kegan Paul; New York: Humanities, 1968), chap. 6. See pp. 116–117 for the use of this analysis in replying to the “irreducible-properties objection.”

It has been pointed out by many commentators that such "translations" do not succeed if construed as translations or meaning analyses in the strict senses of these terms. Armstrong has admitted the point (84–85), saying that all that is offered is an "account," just as Smart suggested that he was merely giving the "general purport" of mentalistic statements. But it is important to see that nothing less than a translation will do, if we hold the view that two properties can be identified one with another only if we show synonymy of the expressions signifying those properties. If, in other words, we want to show that all properties of mental states are properties of brain processes and if we believe that only showing that two terms mean the same thing can show that they signify the same property, our topic-neutral translations will have to be translations in the fullest sense of the term. To see the importance of this point, note that Smart himself has said that, in the light of criticism of his "translations," he feels compelled to say not that sensations and brain processes have the same properties, but that the sensations of common sense simply do not exist and that the explanatory function fulfilled by reference to these pseudo-entities is better fulfilled by reference to brain processes.³

By way of mapping the strategies available to materialists, we can say that if the irreducible-properties objection is to be overcome, materialists must either (a) improve topic-neutral translations so that genuine synonymy results, (b) drop the principle that properties are identical only if the terms referring to them have the same meaning (i.e., assert that there are contingent identifications of properties as well as necessary ones), or (c) adopt the principle that two things can be identical in a philosophically interesting sense even if they do not share all and only the same properties. The second alternative is adopted by Max Deutscher⁴ and by Wilfrid Sellars.⁵ The third

³ See Smart, "Comments on the Papers," in C. F. Presley, ed., *The Identity Theory of Mind* (St. Lucia, Brisbane: Univ. of Queensland Press, 1967), p. 91f.: "I am even doubtful now whether it is necessary to give a physicalist analysis of sensation reports. Paul Feyerabend may be right in his contention that common sense is invincibly dualistic, and that common sense introspective reports are couched in the framework of a dualistic conceptual scheme. . . . In view of Bradley's criticism of my translation form of the identity thesis, I suspect that I shall have to go over to a more Feyerabendian position."

⁴ "Mental and Physical Properties," in Presley, *op. cit.*, p. 75: "a distinction in meaning is not in itself sufficient reason to claim distinctness of properties."

⁵ "The Identity Approach to the Mind-Body Problem," in *Philosophical Perspectives* (Springfield, Ill.: Charles C Thomas, 1967). Sellars here says that "the fundamental strategy of the identity theorist" must be "an appeal to a supposed analogy between the speculatively entertained identity of raw-feel universals with brain-state universals, and the once speculative but now established identity of chemical universals with certain micro-physical universals" (382–383).

alternative is the "Feyerabend" alternative—the adoption of (in Cornman's⁶ phrase) "eliminative" rather than "reductive" materialism, according to which the sense of identity in question is the sense in which phlogiston is identical with (is replaced by, is eliminated in favor of) the kinetic motion of molecules.

Toward the end of this paper, I shall make some remarks about the choice between the second and third of these strategies, arguing for the third. For the present, however, I want to argue simply that the first should be abandoned, not simply because of the detailed criticisms of the particular topic-neutral translations that have been offered (by, e.g., Cornman⁷ and Bradley⁸) but for a more general reason. Briefly, the reason is that if Armstrong were right in saying that

The concept of a mental state is the concept of that, whatever it may turn out to be, which is brought about in a man by certain stimuli, and the cause within a man of certain responses (*op. cit.*, 79).

then we should never have been able to make sense of the contrast between (a) dualism and materialism, or (b) between the mental and the physical, or (c) between materialism and behaviorism.

One form of this point—that involved in (a)—has already been made by Bradley, as follows:

Descartes does indeed speak of non-material Substances of which the items constituent of experience are attributes; but it is hard to believe that he would ever have done so had he not thought that introspectible and physical qualities were utterly disparate. If, so to speak, he had been persuaded by Smart's attempt at a topic-neutral quasi-reduction of sensation statements, it seems natural to suppose that he would have then seen no point in postulating *two* sorts of Substance. . . . Secondly, if Smart's offer of a choice between materialism and dualism, once "topic-neutral" translations have been adopted, is the merest gesture towards the possibility of a dualism, it is also a barely intelligible gesture, or perhaps not even that. For the 'non-physical ghost stuff' will presumably not have introspectible (phenomenal) qualities, for if it did, they would, in consistency, have to be dealt with in a topic-neutral way. Its qualities must therefore be non-physicalism and non-phenomenal. What *they* might be baffles this reader at least.⁹

⁶ James W. Cornman, "On the Elimination of 'Sensations' and Sensations," *Review of Metaphysics*, xxii: 1 (September 1968): 15–35, p. 16.

⁷ "The Identity of Mind and Body," this JOURNAL, LIX, 18 (Aug. 30, 1962): 486–492.

⁸ M. C. Bradley, "Sensations, Brain-Processes, and Colors," *Australasian Journal of Philosophy*, 385–393.

⁹ "Critical Notice" of Smart's *Philosophy and Scientific Realism*, *Australasian Journal of Philosophy*, XLII, 2 (August 1964): 262–283, p. 278.

This point can be underlined and reinforced by noting that Smart and Armstrong think that topic-neutral versions of mentalistic statements can be reconciled with either immaterialism or materialism. They think, in other words, that we are being fair to Descartes as long as we give an analysis of the mental that leaves it open that mental events are taking place in an immaterial stuff. But this neglects the point that 'immaterial' gets its sense from its connection with 'mental'. If the mental is merely the unknown cause of certain behavior or the unknown effect of certain stimuli, then no sense is given to 'immaterial' because no example of the "nonextended" is available to us. The notions of "ghostly stuff" and of "immaterial substance" would never have become current if Descartes had not been able to use *cogitationes* as an illustration of what he intended. Even Aristotle, in looking for examples of form without matter, had to fall back on thought (the agent intellect of the *De Anima* and the "thought thinking itself" which is the "pure actuality" of *Metaphysics Lambda*) in order to find examples. "Immaterial" is not a notion we can hang on to once we have enfeebled our notion of mental in the way described by Armstrong.

Proceeding now to (b), the basic reason why dualism, and *a fortiori* the contrast between dualism and materialism, becomes unintelligible on Armstrong's view is that, if we have a contrast between two categories *X* and *Y*, which are supposed to form an exhaustive and mutually exclusive division of the universe, we cannot mean by '*X*' something that might turn out to be either *X* or *Y*. We cannot define 'mental' as something that might turn out to be either mental or physical, because we cannot define any term as something that might turn out to refer to what is denoted by a contrary term. It is part of the sense of 'mental' that being mental is incompatible with being physical, and no explication of this sense which denies this incompatibility can be satisfactory.

This point—that topic-neutral construals of what it is to be mental lose the mental-physical contrast—may also be brought out by noting that Armstrong's definition of a mental state covers many things that would normally be classified as physical states. Armstrong recognizes this difficulty and replies as follows:

A certain state of the liver, for instance, may be apt for the production of ill-tempered behaviour. Yet it is not a mental state. This objection forces us to say that not all states of the person apt for the production of certain sorts of behaviour are mental states. What marks off the mental states from others? If we consider the secretions of the liver it is clear that, considered as causes, they lack the complexity to bring

about such complexities of behaviour as are involved even in ill-tempered behaviour. It is not until the chain of causes reaches the brain that processes of a sufficient complexity occur . . .

. . . I think it can be replied that our concept of a *mental* state is the concept of a cause whose complexity mirrors the complexity of the behaviour it is apt for bringing about (118–119).

This reply is inadequate. It confuses the question “What is the measure of complexity of a mental state?” with the question “What is the measure of complexity of a physiological state?” We know some rough answers to the latter question; but do we have any idea what it would be for a mere “state apt for the production . . .” to be simple or complex? Armstrong is here assuming that it is already part of our concept of a mental state that it is to be identified with some physiological process or other. But the task he has posed himself is to give a concept of a mental state that makes no reference to such identification. Without such identification, the opposition between simple and complex makes no sense; for to say that a person is in a state apt for the production of certain sorts of behavior is merely to say that such behavior will, *ceteris paribus*, appear (where the *ceteris* may or may not be specified). This dispositional state cannot intelligibly be described as either simple or complex, except in so far as the *ceteris paribus* clause is filled in by spelling out the circumstances in which the behavior will be expected—in which case the length and complexity of the resulting subjunctive conditionals might be said to measure the complexity of the state. But then what we are measuring is the complexity of the behavior expected itself, not “the complexity of its cause.” It is only the (physiological or “immaterial”) state, which lies behind and explains the mental state (and with which the mental state may, on empirical grounds, be identified), that is a *cause*, and only it may be simple or complex. Before the discovery of such states-to-be-identified-with-mental-states we cannot use this contrast to characterize the mental states themselves.

To develop the point that, on Armstrong’s analysis, mental states are mere shorthand for subjective conditionals, it will be useful to go on to (c) and to take up Armstrong’s claim to have set out a genuine alternative to behaviorism. By refuting this claim, I wish to show that the materialism-behaviorism contrast itself makes no sense when ‘mental’ is interpreted in Armstrong’s way. Armstrong admits that his “talk about tendencies to initiate, and capacities for, behaviour” is “perilously close to the Behaviourist’s dispositions,” but insists that “Behaviourism and the Central-state theory still

remain deeply at odds about the way dispositions *are to be conceived*" (85). The difference, he says, is that the behaviorist holds a "Phenomenalist or Operationalist account of dispositions" according to which "to possess a dispositional property is not to be in a particular state," whereas the Central-state materialist holds a "Realist" view, described as follows:

According to the Realist view, to speak of an object's having a dispositional property entails that the object is in some non-dispositional state or that it has some property (there exists a 'categorical basis') which is responsible for the object manifesting certain behaviour in certain circumstances, manifestations whose nature makes the dispositional property the dispositional property it is. It is true that we may not know anything of the nature of the dispositional state (86).

To take a Realist view of dispositions, according to this account, is simply to be willing to say that there is some explanation for the existence of a given disposition, even if this explanation is entirely unknown, and that this explanation is not itself to be given in terms of dispositional properties. But what is the force of this last restriction? To attribute a dispositional property, after all, is merely to say that a given subjunctive conditional is true. But subjunctive conditionals are derivative from nomological generalizations. How do we tell which new nomological statements attribute mere dispositions to the entity in question and which attribute new "categorical" features? I think the only answer to this last question is that Armstrong has in mind *micro-structural* explanations as the paradigm of the case in which new categorical features of the entity are found. (This is suggested by the examples he uses, e.g., the brittleness of the glass being explained by a molecular pattern.) When we merely find a new law regulating the behavior of an entity, without finding or postulating any new entities, we have merely explained one disposition with another. But when we find or postulate new entities, we are explaining a disposition by reference to a categorical state.

If this analysis of the Realist view of dispositions is correct, an odd consequence follows: physicalism *must* be true. Once we accept Armstrong's account of the mental together with the Realist view, it is no longer a scientific, but an a priori, truth, that there are unknown physical entities that explain our being in mental states and are the "categorical bases" of those states. For mental states cannot have their categorical bases in other mental states—the mental cannot be a self-sustaining realm—since mental states are dispositions to

behave and cannot be explained (for a Realist) merely by other dispositions to behave. Further, unless we fall back on the dodge that these nonmental categorical bases might be states of a nonphysical substance, only physical states will do. But this dodge will not do, for the reason given above: the notion of a “nonphysical and nonmental immaterial substance” is a notion without content. “Material” and “physical” would be vacuous notions without the contrast with “mental.” “Immaterial” and “nonphysical” are notions that have sense only if the mental is given as an instance of them. So to adopt Armstrong’s position is to be committed, on a *a priori* grounds, to postulating physical entities “whose nature makes the dispositional properties the particular dispositional properties they are.”

This result, though unwelcome to Armstrong, is to be expected, given his *prima facie* resemblance to the Behaviorist. Both behaviorism and the “topic-neutral” analysis assign mental entities a character that one might call “explanation-hungry”; both dispositions and “states apt for . . .” cry out for something behind them that accounts for them. (Note, incidentally, that there is no reason why Ryle himself should not be a Realist about dispositions.) By making the realm of the mental a realm that contains only relations among physical entities and by accepting the common paradigm of explanation of modern physical science according to which the best explanations of relations among particulars are those which discover new (“micro-”) particulars, one naturally smooths the way for materialism. Unfortunately, however, the way is too smooth. By making materialism an *a priori* truth, we deprive the identity theory of any interest. The interest of the identity theory consists in saying that what used to be thought to be entities that had a nature incompatible with being physical, now turn out to be physical. But Armstrong’s topic-neutral explication of mentality, by making mental entities mere stand-ins for physical entities, leaves nothing to *turn out* to be identical with physical particulars. The materialist who wishes to hold that it is an empirical question whether or not the realm of the mental is self-sustaining—i.e., whether the ideal scientific account of the world might include mental entities as well as physical ones—must insist, against *both* Ryle and Armstrong, on preserving mental entities that have characters incompatible with being physical.¹⁰

¹⁰ I pass over without detailed comment two further points that Armstrong makes in connection with his distinction between Realist and Phenomenalist views of dispositions. In what he calls an *a priori* argument for the former view (*op. cit.*, pp. 86–87) he argues that the Phenomenalist cannot explain why coun-

The upshot of my discussion of Armstrong's reply to the behaviorist is that the difference between Armstrong and Ryle is, at most, the difference between a behaviorist analysis with, and one without, a faith in the possibility of micro-structural explanations of the occurrence of those dispositions to behave which we call mental states. (I say "at most" because, as I have suggested, nothing in *The Concept of Mind* prohibits such faith.) Both types of analysis would, if accepted, reduce the notion of the mental to a notion of relations among (known and unknown) physical entities, and would thereby deprive the notion of "the physical" of sense by stripping it of its contrast with another realm of entities having properties incompatible with physicality. Whether or not Ryle is right in thinking that such an analysis gives us our common-sense notion of the mental as contrasted with the notion held by Cartesian philosophers,¹¹ it seems clear that it is the notion held by Cartesian philosophers that we must explicate if we are to make sense of materialism. This latter notion must contain properties incompatible with properties of physical entities. I now proceed to canvass alternative candidates for the position of being these key incompatible properties, and to argue that incorrigibility is the best candidate.

In settling upon a mark of the mental, it is important to begin by distinguishing between two different notions of what counts as mental. The distinction I have in mind is that between the sort

terfactual conditionals are true whereas the Realist can. As far as I can see, this argument presupposes that inductive arguments to the presence of dispositional properties are *ipso facto* weaker than inductive arguments to the presence of categorical properties. *Pace* Armstrong, the Phenomenalist can reply to the question "Why should a thing not change its dispositional properties?" *not* by an appeal to underlying categorical properties, but simply by an appeal to the constancy of the dispositional property in the past.

The second further point Armstrong makes is that on the Phenomenalist view "dispositions cannot be causes," whereas, since the Realist identifies dispositions with *states*, his view permits them to be causes. Once again, it is not clear that the distinction between dispositions-as-states and dispositions-as-non-states comes to more than the distinction between behavioral law backed up by reference to new particulars, and those not so backed up. But is not clear why explanation by reference to unbacked-up behavior laws should not count as causal explanation.

¹¹ I do not mean to suggest that I think Ryle *is* right about this. On the contrary, I should hold that common sense is irredeemably Cartesian on the point. I should argue that Ryle's purported distinction between the concept built into our language and the Cartesian concept is actually a distinction between the concept most congenial to a verificationist and operationalist philosopher and the concept actually built into our language. On the operationalist presuppositions of *The Concept of Mind*, see Albert Hofstadter, "Professor Ryle's Category-Mistake," this JOURNAL, XLVIII, 9 (Apr. 26, 1951): 257-270. On its verificationist presuppositions, see Stuart Hampshire's review in *Mind*, LIX, 234 (April 1950): 237-255.

of mental entity that is an *event* and the sort that is not. In the first class fall, paradigmatically and perhaps solely, thoughts and sensations. By "thoughts" here I mean not beliefs, but occurrent, datable, thoughts—e.g., the entity referred to when one says "The thought that *p* suddenly struck me." By "sensations" here I mean not perceivings—not acquisitions of beliefs—but simply the entities that are reported in such ways as "Then I had a sensation of red" or "Then I had a painful sensation in my leg." These two sorts of entities make up the content of the stream of consciousness—what one finds when one asks "What's going on in me now?" In the second class fall all those mental entities which are not events and which are only dubiously "entities" at all—beliefs, moods, emotions, desires, purposes, intentions, motives, etc., etc. These might better be called "mental features" than "mental entities." Not only are they not events, but it strikes one as an odd, peculiarly philosophical, hypostatization, to think of them as *particulars* of any sort.

Another way of contrasting these two classes is to note that they are recalcitrant to behaviorist "reduction" in different ways. To say that thoughts and sensations are dispositions to behave sounds counterintuitive—as counterintuitive as saying that molecules are dispositions of macroscopic objects to behave. A Rylean approach to thoughts and sensations runs into the obstacle that here ordinary language seems to steadfastly support the Cartesian notion of a double series of events—one mental and the other physical—which when put together make up the human being. Here, if anywhere, we genuinely have a ghost in the machine—a set of nonphysical occurrences. When we come to beliefs, emotions, desires, purposes, and the like, however, we are no longer tempted to count them as episodes rather than dispositions. Here Ryle's general approach—that talking about these things is a way of talking about what behavior may be expected—has great intuitive plausibility. The recalcitrance in these cases is rather that when we try to give equivalents in terms of bodily movements for such expressions as "He wants *X*" and "He intends to *A*" we seem to fail. We cannot break out of the circle of terms whose most prominent members are 'belief' and 'desire', because, roughly, the putatively equivalent hypothetical sentences about physical movements always seem to require in their protases such qualifying phrases as 'Provided he believes that . . .' and 'Provided he wants that . ..'. Whereas thoughts and sensations were recalcitrant to reduction because they did not seem like dispositions at all, beliefs and desires (and the rest) are recalcitrant to

reduction because, though they may be dispositions, they cannot be isolated without reference to other such dispositions.

I wish now to argue that only the former class of mental entities generate the opposition between the mental and the physical, where this opposition is considered as an opposition between two incompatible types of entity, rather than an opposition between two ways of talking about human beings. The former class of entities—the thoughts and the sensations—are the paradigm illustrations of what is meant by the Cartesian notion of the mental as a separate realm. The latter class of entities are entities which, *if we had never heard of thoughts and sensations*, would never have generated the notion of a separate “realm” at all. If we had no notion of a mental *event*, but merely the notion of men having beliefs and desires and, therefore, acting in such-and-such ways, we would not have had a mind-body problem at all, and Ryle would have had no motive for writing. Believing and desiring would have appeared simply as distinctively *human* activities, and our only dualism would have been one between human beings qua agents (i.e., qua moving in ways to be explained by reference to beliefs and desires) and as mere bodies (i.e., qua moving in ways that can be explained without reference to beliefs and desires). This dualism would have been a dualism not between mind and body, nor between the mental and the physical as distinct realms, but simply between ways of explaining the doings of human beings—psychological explanations and nonpsychological explanations.¹²

I have presented this distinction between mental events—the content of the stream of consciousness—and mental features in order to explain why I shall be concentrating on the former in looking for a mark of the mental. In what follows, I shall be asking the question: What features or feature do thoughts and sensations have in common with each other, and with nothing physical? It will turn out that their single common feature—incorrigibility—is only in a weak and diminished sense a mark of such things as beliefs, desires, purposes, emotions, etc. I shall be forced to conclude, therefore, that

¹² My distinction between the mental entities that are events and those which are not might thus be expressed as a distinction between the *mental* and the *psychological*. This way of putting the matter would have the merit of calling attention to the difference between Descartes's distinction between the mind and the body and Aristotle's distinction between the soul and the body. However, I do not wish to press this terminology, for I cannot, in the space of the present paper, offer a full-blown account of the “Aristotelian” as opposed to the “Cartesian” notions of where the interesting lines fall. Cf. Wallace Matson, “Why Isn't the Mind-Body Problem Ancient?” in Paul Feyerabend and Grover Maxwell, eds., *Mind, Matter, and Method* (Minneapolis: Univ. of Minnesota Press, 1966), pp. 92–102.

there is no single mark of all the entities customarily called mental. But I believe that isolating the features of the paradigmatically nonphysical, the mental *events*, serves two useful purposes. In the first place, it lets us see how the notion of mutually exclusive realms of being came to exist. In the second place, it lets us see that there are family resemblances between mental events and mental entities that are not events—resemblances which account for the tendency to use the term ‘mental’ of both, despite the differences I have mentioned and despite the fact that one can construct no set of necessary and sufficient conditions for mentality.

Proceeding now to candidates for marks of mental events, we may begin by noting that two familiar marks of the mental—intentionality and “purposiveness”—are excluded from consideration by our inclusion of *sensations* as mental. None of the marks of the intentional—e.g., those proposed by Chisholm—would make “I am having a sensation of red now” or “I am having a painful sensation now” an intentional sentence. To have a sensation, unlike having a thought, is not to be in a state which has “aboutness” or which can somehow refer to the inexistent. Nor does there seem to be anything distinctively “purposive” about sensations. We can explain what a sensation is without any reference to beliefs or desires. The notion of sensation is not a part of the circle of terms used to explain action as opposed to movement (although of course reference to sensations may enter into such explanations, just as reference to physical objects may enter). If we are to find something that sensations and thoughts have in common with each other and not with anything physical we must look away from intentionality and purposiveness to the following group of marks: *introspectibility*, *nonspatiality*, and *privacy*. These are the sorts of characteristics that distinguish the contents of the stream of consciousness from “the external world” and generate the notion of the physical and the mental as distinct realms.

To begin with introspectibility, although everything mental is introspectible and conversely, nevertheless it is unhelpful to cite this as a mark of the mental. The unhelpfulness comes out when we try to distinguish introspection from such borderline cases as sensing that one’s stomach is fluttering or that a vein in one’s leg is throbbing. These latter cases do not count as cases of introspection simply because the object reported on is physical. In short, we cannot explain what introspection is except by reference to an antecedently understood notion of what is mental. To say that all and only mental events are introspectible is no more informative than

saying that all and only these are knowable in that unique way in which we know our own mental events.

Nor is nonspatiality a satisfactory mark of the mental. The difficulty here is that it makes excellent sense to give thoughts and sensations a location, though a vague one—namely, to say that they are located where the person doing the thinking or the sensing is located. From the point of view of the identity theory, this position has the advantage that, given reasons for identifying thoughts and sensations with brain processes, it will make sense to make the location of the former more precise than it was previously. From the point of view of our search for marks of the mental, it should be noted that we cannot make “vague spatiality” as opposed to “precisely locatable spatial position” a mark of the mental, because the same vagueness applies to the location of my weight, my build, my health, and my behavior—all of which are located where I am, but are not more precisely locatable, and none of which are “mental.” Nor can we back up the notion that the mental is unextended and the physical extended by claiming that the *shape* or *size* of thoughts and sensations is a contentless notion, whereas all physical things have shape and size. The mass or the weight of physical objects does not have shape or size, and indeed no *state* of an object, as opposed to the object itself, does. To insist that mental events are shapeless and sizeless is merely to remind us that they are states of persons.

The temptation to explicate ‘mental’ as ‘unextended’ comes, I think, from taking a special case—images of physical objects had in dreams or hallucinations—as paradigmatic of the mental. It is easy to say that Macbeth’s “dagger of the mind” does not occupy space, or at least not “physical” space, and then extrapolate from there. But it is a bad example to extrapolate from, since it is an example of something that does not exist. What does exist are certain sensations (sense impressions of something daggerlike) and thoughts (that there is a dagger there) in Macbeth. And these are not on the table where the dagger seems to be, but where Macbeth is. The temptation here is to think that mental things are objects rather than states, that all objects must have features homogeneous with those of real physical objects (e.g., color, shape, and size), that “mental objects” have such features only in some Pickwickian sense (e.g., “phenomenal” color or size or location), and then to conclude that it is this Pickwickian possession of familiar features that characterizes the mental. But this is to confuse the mental with the in-existent or the imaginary, to confuse thoughts with their intentional (and possibly in-existent) objects (as if to think about unicorns

was to have a nonexistent unicorn in our minds), and sensations with the objects that the presence of certain sensations may lead us (mistakenly) to believe exist. Descartes's preoccupation with dreaming, and the habit of treating objects dreamt of as "mental objects," led to this confusion and thus to much of the obscurity surrounding the notion of the mental.

To supplement these last remarks, and also as a way of introducing the topic of incorrigibility, it will be useful to digress for a moment to a view about marks of the mental which is suggested by Sellars. Sellars emphasizes the point we have just made—that sensations and thoughts are states of persons rather than quasi-substances—and adds the further point that their intrinsic features are features that are not shared by physical objects, real or imagined. On his "mythical" account, thoughts were originally theoretical entities, postulated as "inner" states that explained certain sorts of behavior. But they were not merely Rylean dispositions nor Armstrongian "states apt . . ."; for they had certain intrinsic features. For example, they were true or false, and were *about* things, in the way in which sentences are. They shared, in other words, the "semantical" features of sentences—the features sentences possessed not qua physical objects (inscriptions) but qua types (as opposed to tokens)—but had no other features. Sensations, in turn, were also originally theoretical entities—"inner" states postulated to explain the occurrence of certain thoughts (e.g., the thought that there is a red triangle before me, when there isn't). They too had certain intrinsic features, but, again, features not shared by any physical objects qua physical objects. Their intrinsic features were, e.g., being "of red" and "of a triangle." ['Of' here is not a relational expression, but is a device for introducing such new theoretical predicates as (using hyphens to mark unanalyzability) "of-red"—a predicate which applies *only* to sensations and gains application through such "correlation rules" as that which says that red triangles perceived in standard conditions give rise to sensations that are "of-red" and "of-a-triangle."] When originally proposed as theoretical entities (by Jones, the man who, in Sellars's myth, invented the concept of mind) sensations and thoughts were not conceived of as immediate experiences—they were not the objects of noninferential introspective reports, much less of incorrigible reports. Instead, they were inferred entities—known to exist in the way in which positrons are known to exist, by inference from the behavior they cause. It is only after Jones has instructed others in his theory and subjected them to a prolonged training process

that it turns out they can make noninferential reports of their own inner states.¹³

What, we may now ask, is the mark of the mental on Sellars's account? What intrinsic features do sensations and thoughts have in common? Oddly enough, the answer is that they have *no* features in common save the Armstrongian one of being "inner" states apt for the production of certain behavior. Though they both have intrinsic features, and not merely relational features, they have no *common* intrinsic features save "innerness." But what does being "inner" come to? I suggest that all the term can mean (before the day when Jones trains his fellows to make not merely noninferential, but incorrigible reports, of their thoughts and sensations) is "beneath the skin." To postulate such states, like postulating Rylean dispositions or Armstrongian states, is not to give a basis for the notion of the "nonphysical"—or, to put it more accurately, does not provide a means for giving the notion of "physical" a sense by contrasting it with something else. Rather, the natural thing for Jones's pupils to think is that he is telling them that something happens somewhere in their bodies which accounts for their behavior, something on all fours with internal secretions or muscle movements.

To see this point, it helps to notice that Jones might just as well have introduced the notion of "brain-process-about-*p*" or "brain-process-of-a-red-triangle" as have invented the neologisms 'thought about *p*' and 'sensation of a red triangle'. What counts is not whether a new word or an old is used, but merely the theoretically postulated intrinsic features of the entities in question and the relevant "correlation rules." It would have been just as good an explanation of intelligent behavior to say that some brain processes had, like sentences, the special feature of being "about" things, as to say that an invented state called a "thought" did. These new properties of brains would have been, if Jones had phrased his theory in this way, "unobservable" properties, but they would not have been nonphysical properties, any more than the spin of an electron is a nonphysical property.

Coming now to the point, I want to say that Jones did not invent the concept of *mind* by inventing the notions of unobservable inner states with certain intrinsic features. Given Sellars's description of his theory, all that Jones did was to propose a micro-structural account of the causes of human behavior, but not an account in

¹³ This paragraph summarizes pp. 186–196 of "Empiricism and the Philosophy of Mind" in Sellars's *Science, Perception, and Reality* (New York: Humanities, 1963).

terms of specifically mental events. We cannot make Armstrongian “states apt . . .” into *mental* states just by adding an assortment of intrinsic features to them unless there is among those features one which separates off all such states from any other states we know of and, thereby, establishes a new category of existence.

This seems a strong requirement, but it is exactly what is supplied by the *privacy* of mental events. We must be careful, however, to isolate the right sense of ‘private’. As A. J. Ayer has pointed out,¹⁴ mental events have been said to be private in at least the following four senses: incommunicability, special access, unsharability, and incorrigibility. In the first sense, things are private to a person if only he can know of their existence or some of their features. Mental events are clearly not private in this sense, unless one believes that thoughts or sensations have special felt qualities that are not signified by any term in a public language. But the latter view is hardly part of common sense or of our normal conception of the mental. In the second sense, things are private to a given person if he can know about them in ways different from those in which anyone else can know about them. But in this sense my stomach is private to me, for I can know that it is fluttering by feeling that it is, and no one else can do that. So this sense will not give us what we want. In the third sense—“unsharable”—things are private to a given person if it is impossible for anyone else to have them. But this again extends too far, for no one else can have my state of health or my behavior. Nor is it clear, because of the possibility of telepathy and of interlocked brains, that thoughts and sensations *are* unsharable. If we want to say they are, we have to rule that in telepathic communication we only have the same *kind* of thought and not the same thought, but this ruling seems arbitrary and ad hoc. The fourth sense of privacy, however—incorrigibility—does give us what we want. Mental events are unlike any other events in that certain knowledge claims about them cannot be overridden. We have no criteria for setting aside as mistaken first-person contemporaneous reports of thoughts and sensations, whereas we do have criteria for setting aside all reports about everything else.

We may accept Sellars’s “myth” as a reasonable account of how terms that were eventually to refer to the mental entered the language, but we must guard against thinking that the notions of inner states “about *p*” or “of-red” give us the notion of something mental, something categorically distinct from everything else.

¹⁴ Cf. *The Concept of a Person* (New York: St. Martin’s, 1963), p. 79.

Only after the emergence of the convention, the linguistic practice, which dictates that first-person contemporaneous reports of such states are the last word on their existence and features, do we have a notion of the mental as incompatible with the physical (and thus a way of making sense of such positions as parallelism and epiphenomenalism). For only this practice gives us a rationale for saying that thoughts and sensations must be *sui generis*—the rationale being that any proposed entity with which they could be identified would be such that reports about its features were capable of being overruled by further inquiry. Before this practice arose, it would have made no sense to ask whether Jones was giving us a theory about mental or about physical entities.

The force of this point may be brought out by noting that if, as we suggested above, Jones had produced a theory of “brain processes about *p*” and “brain processes of red” he would still, *if the same linguistic practice had arisen*, have invented something that turned out to be mental. Instead of states of a person that were incorrigibly reportable, there would have been states of brain processes that were incorrigibly reportable. There would, so to speak, have been no mental *entities*, but brain processes would have had mental *properties*. What makes an entity mental is not whether or not it is something that explains behavior, and what makes a property mental is not whether or not it is a property of a physical entity. The only thing that can make either an entity or a property mental is that certain reports of its existence or occurrence have the special status that is accorded to, e.g., reports of thoughts and sensations—the status of incorrigibility.

In what precedes, I have given reasons for denying to the following the title of the mark of mental events: intentionality, purposiveness, nonspatiality, introspectibility, privacy as incommunicability, privacy as special access, and privacy as unsharability. I have also urged that Sellars’s account of thoughts and sensations in terms of certain special features (being “about *X*” or “of red”) will not do the job. I have emerged with the conclusion that only incorrigibility marks off a common feature of our paradigms of mental events—thoughts and sensations—which distinguishes mental events from anything physical. I now turn to making this notion of “incorrigibility” more precise and to defending against objections the claim that we have incorrigible knowledge.

It is customary to define incorrigibility in terms of the notions of entailment or logical possibility. Thus Armstrong (101) offers the

following definition of “ p is logically indubitable for A ”:

- (i) A believes p
- (ii) (A 's belief that p) logically implies (p)

and George Nakhnikian¹⁵ gives the following definition of “it is incorrigible for S at t that p ”:

- (i) It is logically possible that at t S believes attentively that p , and
- (ii) “At t S believes attentively that p ” entails “At t S knows that p ”

I wish, however, to eschew reference to logical modalities, both because of general Quinean doubts about the existence of necessities other than “natural” ones and because of a particular difficulty that arises when we try to spell out ‘logically possible’ in this context. Suppose that, in the familiar manner, we try to spell out the force of this term in such sentences as “It is logically impossible that I believe that I am thinking that p , and not be” by the notion “impossible by virtue of the meaning of terms.” We shall then arrive at the conclusion that the meaning of the terms ‘thinking’ and ‘thought’ is such that it is impossible to have incorrect contemporaneous beliefs about what one is thinking. But now let us recur to Jones, who uses the word ‘thought’ before people have learned to make introspective reports of their thoughts, much less come to view such reports as incorrigible. Must we say that when Jones first invented the notion of “thought,” meaning by it “inner state that can be about X , be true or false . . . , etc.,” he did not mean by the word what we do? Did the meaning of ‘thought’ change when people came to make noninferential reports of their own thoughts? Did it change when these reports came to be regarded as the last word? Would it change if cerebroscopes came to be regarded as offering better evidence for what someone was thinking than his own introspective reports?

I regard these questions as unanswerable, and affirmative answers to them as dogmatic pieces of what Hilary Putnam calls “unreasonable linguistics.”¹⁶ We have here a case where Quine’s thesis of the indeterminacy of translation—the point that we may regard either *meanings* or *beliefs* as having changed, with no clear reason for choosing one alternative over the other save elegance or sim-

¹⁵ “Incorrigibility,” *Philosophical Quarterly*, xviii, 72 (July 1968): 207.

¹⁶ “Brains and Behaviour,” in R. J. Butler, ed., *Analytical Philosophy: Second Series* (New York: Barnes & Noble, 1963), p. 19, where Putnam is arguing against the view that the ascription of pain to beings who never wince nor admit to being in pain *purely* on the basis of brain waves involved a change in the meaning of ‘pain’.

plicity—is directly relevant to philosophical issues. My own preference would be to say that in none of the imagined cases does the meaning of ‘thought’ change and to defend this claim by invoking Putnam’s notion of a “cluster-concept” (*ibid.*, 5). I do not wish to defend this piece of impromptu linguistics, however, but merely to urge that we should not let a definition of incorrigibility in terms of logical modalities drive us to such conclusions as that Jones did not mean by ‘thought’ what we do. Whether the myth of Jones be true or not, we should at least have the ability to tell a coherent story along the lines of the Jonesian myth. We can do this if we define incorrigibility “naturalistically,” so to speak, in terms of the linguistic practices adopted by Jones’s successors.

What, then, did these successors do when they “made” reports of sensations and thoughts incorrigible (and thus, on my view, made sensations and thoughts mental)? Well, something like this. They found that, when the behavioral evidence for what Smith was thinking about conflicted with Smith’s own report of what he was thinking about, a more adequate account of the sum of Smith’s behavior could be obtained by relying on Smith’s report than by relying on the behavioral evidence. Thus, for example, if Smith’s cave-reentering and ax-grasping behavior seemed to point to his just having had the thought that he had left his ax in the cave, his subsequent use of the ax nevertheless confirmed the truth of his report that what he had actually thought at the moment in question was that he might have broken the ax-handle yesterday. The growing conviction that the best explanation in terms of thoughts for Smith’s behavior would always be found by taking Smith’s word for what he was thinking found expression in the convention that what Smith said went. The same discovery occurred, *mutatis mutandis*, for sensations. It became a regulative principle of behavioral science that first-person contemporaneous reports of these postulated inner states were never to be thrown out on the ground that the behavior or the environment of the person doing the reporting would lead one to suspect that they were having a different thought or sensation from the one reported. In other words, it became a constraint on explanations of behavior that they should fit all reported thoughts or sensations into the over-all account being offered. This constraint came to be reflected in linguistic practice, so that the expression ‘You must be mistaken about what you’re thinking’, which had had an established use in the past (*viz.*, to reflect apparent conflicts between behavior or environment and reports), fell into desuetude.

If this is a plausible myth, it can be described either as the history of how the meanings of 'thought' and 'sensation' changed or as the history of how people came to acquire new beliefs about thoughts and sensations (viz., that certain reports about them could not be mistaken). For the reason given above, I prefer to describe it in the second way. This choice means that I must define incorrigibility in terms not of logical possibility, but of the procedures for resolving doubts accepted at a given era. Thus I submit the following¹⁷:

S believes incorrigibly that *p* at *t* if and only if

- (i) *S* believes that *p* at *t*
- (ii) There are no accepted procedures by applying which it would be rational to come to believe that not-*p*, given *S*'s belief that *p* at *t*

As an initial comment on this definition of incorrigible belief, let me note that it is immune from certain familiar objections which Armstrong and others have brought against incorrigible belief as belief that implies its own truth.¹⁸ Armstrong points out that there is a prima facie incompatibility between the materialist claim that knowledge of one's own mental states is a result of self-scanning by the brain and the claim that we possess logically incorrigible knowledge of such states. For how could it be logically impossible for the scanning process to go wrong? On our definition, this is not a problem. All we are asserting, when we say that contemporaneous beliefs about our own mental states are incorrigible, is that there is no assured way to go about correcting them if they should be in error. Viewing the matter in this way reduces incorrigibility to what Armstrong refers to as "empirically privileged access"¹⁹—an epistemological status relative to the state of empirical inquiry, and one capable of being lost if, for example, cerebroscopes should come to overrule first-spoken reports. Against such privilege, Armstrong

¹⁷ Cf. my article "Intuition" in *The Encyclopedia of Philosophy*, edited by Paul Edwards (New York: Macmillan, 1967), vol. iv, pp. 204–212, where I present a variant of this definition in the context of an account of intuitive knowledge.

¹⁸ Cf. "Is Introspective Knowledge Incorrigible?", *Philosophical Review*, lxxii, 4 (October 1963): 417–432, and *A Materialist Theory of the Mind*, pp. 100–113. See also Smart, *Philosophy and Scientific Realism*, p. 100, and the references given there.

I shall not take up here the argument against incorrigible knowledge that Malcolm, in his review of *Philosophical Investigations*, imputes to Wittgenstein—viz., that where we cannot make a mistake we cannot make a knowledge claim, so that it is nonsense to say "I know that I am in pain." I have tried to rebut this argument in another article—"Wittgenstein, Privileged Access, and Incommunicability," *American Philosophical Quarterly*, forthcoming.

¹⁹ Cf. *A Materialist Theory of the Mind*, p. 108.

has nothing to say. In particular, his argument that there can be no logical connection between distinct existences—e.g., between our mental state and our awareness of our mental state—is irrelevant to a sense of ‘in corrigible’ that eschews reference to logical connection.

If our definition has the advantage of circumventing familiar objections to incorrigible knowledge, it has the disadvantage of including more than just knowledge of the mental. For there are three varieties of statements that may be believed incorrigibly in the sense just defined: (a) statements knowable a priori; (b) statements reporting mental events; and (c) statements about how something appears, looks, or seems to someone. In the case of a priori knowable propositions, the phrase ‘given *S*’s belief that *p*’ in the above definition can be ignored, since our belief in such statements as “ $2 + 2 = 4$ ” and “Every event has a cause” is guaranteed simply by the absence (at the moment) of accepted procedures for overthrowing them. In the case of (b) and (c), however, the phrase is essential. It serves, roughly, to summarize the fact that present procedures for adjudicating belief claims are such that the fact of *S*’s belief at *t* that *p* is at least as strong evidence for *p* as any imaginable state of affairs could be for not-*p*. This is the situation that obtains for statements like “It looks brown to me now” and like “I’m in pain now.” In both cases we may doubt the fact expressed, wondering whether it really looked brown or really *was* pain, and others may follow us in these doubts, but we have no procedures available for resolving such doubts. Granted that, as Austin says, we may come to suspect that it just wasn’t *brown* that it looked to be, there is no way we can rationally decide that it *didn’t* look brown in the face of the contemporaneous belief. We are condemned to hesitation. (This is why I prefer to speak of *incorrigible* rather than *indubitable* belief; any belief can be doubted, but not all such doubts are rationally resolvable.)

Given this definition of ‘in corrigible belief’, how now may we put it to use to mark off the mental? I propose the following strategy. The thesis presented is that all and only mental events are the sorts of entities certain reports about which are incorrigible. We may get rid of a priori statements by noting that they are not *reports*—they are not descriptions of particular states of affairs.²⁰ What about

²⁰ It might be urged here that “This bachelor is unmarried” is a description of a particular state of affairs, though knowable a priori. To exclude such cases, we can include an additional restriction on the notion of “report”—viz., that reports are not known to be true by virtue of knowing that universally quantified statements are true.

“appears” statements? These do seem to be reports of particular states of affairs. Are they, and are they then reports of mental states? This is a subtle question. One move would be to say “yes” to both questions—to say that they are reports of thoughts, “It looks brown to me now” being equivalent to “I am thinking that it may be brown.” I do not wish to make this move, however, because “appears” statements seem to me, in their primary meaning, statements one could use without ever having heard of thoughts or sensations—statements which, so to speak, have a pre-Jonesian use. In this use, they simply mark refusals to commit oneself to making a report of a certain sort. To say that “X looks brown” is, at the least, to express hesitation about saying that X is brown. Is it also to make a report—a report that one is in a state of hesitation about saying that X is brown, that one is tempted to do so but not quite willing to do so? I would claim that (post-Jones) it may be and that, when it is, it is a report of a mental state, a thought of a given sort. But I would urge that it may not be, and may be simply an *expression* of hesitation, rather than a *report* of a hesitation. So I wish to use the notion of “report” to mark off among incorrigibly believable statements those which are about mental states. Only the reports are about mental states; the others are not. By “reports,” once again, I mean descriptions of particular states of affairs. An “appears” statement may be a description of a particular state of affairs, and in that case it is a description of the mental. But it may be simply a refusal to make a description of a particular state of affairs, and this is its primary, pre-Jonesian use.

We can sum up the results of this strategy by noting that we now have a set of necessary and sufficient conditions for something being a mental *event*, namely

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.

We now, however, have to face up to the question of whether we have necessary and sufficient conditions for something being a mental *entity*—whether, in other words, our criterion can be made to apply to beliefs, desires, moods, emotions, intentions, etc., as well as to thoughts and sensations. Here the answer, unfortunately, is “no.” Those mental entities which I have contrasted with mental events as mental *features* are such that our subsequent behavior may provide sufficient evidence for overriding contemporaneous reports of them. If I say that I believe that *p*, or desire X, or am afraid, or

am intending to do *A*, what I go on to do may lead others to say that I couldn't *really* have believed *p*, or desired *X*, or been afraid, or intended to do *A*. This fact is what we should expect, given the nonepisodic, dispositional, character of these entities. Statements about beliefs, desires, emotions, and intentions are implicit predictions of future behavior, predictions which may be falsified. Such falsification provides an accepted procedure for overriding reports. In this they are distinct from reports of thoughts and sensations, which are compatible with any range of future behavior.

But the fact that we are not incorrigible in our reports of mental features as we are about mental events should not blind us to the fact that we are *almost* incorrigible. The possibility of overriding reports about such features is real, but it is actualized only rarely and with trepidation. We are far less likely to have a report about a mental state, even one that is not an event, overridden than to have a report about something physical overridden. Further, as such mental features as beliefs and desires become more particular and limited and, thus, approach the status of episodes rather than dispositions, they become *more* incorrigible. It is not clear that there *are* accepted procedures for overriding someone's sincere report that he believes there is a table before him or that he desires a peach now. This may be explained by noting that there is no clear distinction between saying I believe that there is a table before me and saying that the thought has struck me that there is a table before me, nor is there a clear distinction between saying I want a peach now and saying that the thought "Would that I had a peach!" has just struck me. Again, there is no clear distinction between saying I am afraid of the tiger I just encountered and saying I had a sensation of fright when I encountered him. Beliefs and desires about momentary matters tend to collapse into thoughts, and momentary emotions tend to collapse into sensations. Short-run beliefs, desires, emotions, and intentions are less like predictions of future behavior than like avowals of contemporaneous thoughts or sensations. That is why they are more like episodes than like dispositions.

The two factors we have just mentioned—the near-incorrigibility of reports of mental features, and their tendency to become strictly incorrigible as they become more particular and limited—account, I believe, for the term 'mental' having been stretched from the paradigm cases of the nonphysical—thoughts and sensations—to such things as beliefs, desires, emotions, and intentions. If I am right in saying that strict incorrigibility is the mark of mental

events and if I was right in saying above that it was mental events, as opposed to mental features, which engendered the Cartesian notion of the mental and the physical as separate realms, then it is appropriate that near-incorrigibility should be the basis for widening the realm of the mental. The likeness of near-incorrigibility to strict incorrigibility is the family resemblance that ties the various things called "mental" together and makes it possible to contrast them all with the physical. But the distinctness of near- from strict incorrigibility is what makes it impossible to find any interesting set of necessary and sufficient conditions for mentality.

I have now completed my search for marks of the mental. I shall end by turning to the relevance of my results to materialism. I began by arguing that the attempt to avoid the "irreducible-properties objection" to mind-brain identity foundered on the incompatibility between the mental and the physical. I have now isolated that incompatibility as the incompatibility between what we are strictly or nearly incorrigible about and what we are straightforwardly corrigible about. What can the materialist do in the face of this incompatibility?

I suggest that he can say, simply, that it might turn out that there are no entities about which we are incorrigible, nearly or strictly. This discovery would be made if the use of cerebroscopes (or some similar mechanism) led to a practice of overriding reports about mental entities on the basis of knowledge of brain states. If we should, as a result of correlations between neurological and mental states, begin taking a discovery of a neurological state as better evidence about a subject's mental state than his own report, mental states would lose their incorrigible status and, thus, their status as *mental*. This possibility is a result of the way in which we defined 'incorrigible belief'. By phrasing our definition in terms of accepted procedures, rather than in terms of the logical impossibility of error, we leave room for the sort of change that would confirm "eliminative" materialism.

There is, however, another way in which materialism could be vindicated, but this too involves a shift in linguistic practices on the part of our descendants. If it came to pass that people found that they could explain behavior at least as well by reference to brain states as by reference to beliefs, desires, thoughts, and sensations, then reference to the latter might simply disappear from the language. Reports of thoughts and sensations, e.g., might be replaced by reports of brain processes. To invoke a possibility I have ex-

plored in another article,²¹ reference to mental states might become as outdated as reference to demons, and it would become natural to say that, although people had once believed that there were mental states, we had now discovered that there were no such things. Instead of our continuing, as in the first alternative suggested above, to speak about thoughts, desires, and the like but ceasing to let ourselves be incorrigible about them, we might simply cease to talk about them at all (except for antiquarian purposes). Either of these changes would give the "eliminative" materialist the right to say that it had been discovered that there were no mental entities.

This conclusion amounts to saying that only the third of the three strategies I described at the outset as available to the materialist is viable. The first strategy, involving "topic-neutral" translations, has already been discussed. The second involves circumventing the "irreducible-properties objection" by making contingent identifications of mentalistic properties with neurological properties. But the second strategy, like the first, founders on the incompatibility of the mental and the physical. Even if we could identify "being about *p*" or "being of red" with neurological universals, we are never going to identify the property of being the subject of an incorrigible or near-incorrigible report with any neurological property. For this is not a feature which mental states have, so to speak, by themselves and which might be found mirrored in neurology—it is a feature attached to them by the linguistic practices of a community. This property may cease to hold of thoughts, beliefs, sensations, desires, etc.—in which case these things would cease to be *mental* entities. Or these entities might be rejected altogether. But nothing would count as finding a neurological property that *was* the property of being the subject of incorrigible reports.

Only the third strategy, therefore—the one which admits that there is an incompatibility between being mental and being physical, but suggests that there may be no mental entities—will do as an explication of the materialist thesis. But to say that it might turn that there are no mental entities is to say something not merely about the relative explanatory powers of psychological and physiological accounts of behavior, but about possible changes in people's ways of speaking. For as long as people continue to report, incorrigibly, on such things as thoughts and sensations, it will seem

²¹ Cf. "Mind-Body Identity, Privacy, and Categories," *Review of Metaphysics*, xix, 1 (September 1965): 24–54.

silly to say that mental entities do not exist—no matter what science may do. The eliminative materialist cannot rest his case solely on the practices of scientists, but must say something about the ontology of the man in the street.

Yet it may seem outrageously paradoxical to say that the truth of an ontological thesis depends in part upon what linguistic practices are adopted by the community. One's feeling is that it should be the other way around—that such practices should shift as a result of the discovery of ontological truths. Perhaps the paradoxical flavor may be diminished, however, by noting the near-invisibility of the difference between the identity thesis and a certain form of parallelism. On this form of parallelism, there are neural-mental correlations of such a sort that every "natural kind" of mental state is constantly correlated with a "natural kind" of neural state.²² If such correlations occurred, every explanation of behavior in terms of mental states would be isomorphic to an explanation of behavior in terms of neural states, neither mode of explanation being simpler or more elegant or more fruitful than the other. The discovery of this form of parallelism would, it seems clear, be a necessary condition for either of the changes in linguistic practices I have described. (No one should be inclined to let cerebroscopes correct introspection, nor to stop talking about mental states altogether, unless this degree of "interchangeability" of the mental and the neural ways of speaking had been discovered.) But it seems equally clear that it is not a sufficient condition. The further condition necessary would be, roughly, a preference for Occam's razor over old ways of speaking.

Whether this preference is felt by the community as a whole in the way the materialist would like it to be felt is something very close to being a matter of taste, not to be decided by either empirical discoveries or philosophical argumentation. But even if the general will goes against him, the proponent of the identity thesis should not be abashed. He can still say that it would be *rational* to go

²² The possibility I am suggesting here is one envisaged by Charles Taylor in "Mind-Body Identity: A Side Issue?," *Philosophical Review*, LXXVI, 2 (April 1967): 201–213—that not only will every mental state be correlated with a brain state, but that regularities among brain states adequate to the explanation of behavior will appear not only when, in Taylor's phrase, we "characterize these events as *embodiments* of the corresponding thoughts and feelings" but also when we consider them purely in their own, neurological, terms. Taylor thinks that this "(ultimately empirical) question of the most fruitful forms of explanation of behavior" is "the major question in dispute between materialists and their opponents" and that the materialists win if such purely neurologically characterizable regularities appear. I am arguing that, for the materialist to win, a further step would be necessary—a change in linguistic practices.

beyond parallelism to identity. But would he bother? Would not the purported advantage of saying 'identical' rather than 'correlated' have begun to seem a mere shibboleth? Would the identity thesis still be an interesting point of controversy, once parallelism of the sort we have described is found to hold? I suspect not, and therefore I take what I have said about the need for changes in linguistic practices in order for the identity thesis to be affirmed, although formally correct, to be somewhat misleading. When ontological issues boil down to matters of taste, they cease to be ontological issues. If parallelism of the sort described were discovered, there would, I think, cease to be an issue about materialism. For the materialist would have succeeded in showing that all phenomena can be explained completely in physicalistic terms, and this would be enough to satisfy his ontological intuitions. Insistence on the "identity" of the mental and the physical would seem an unnecessary rhetorical flourish.

RICHARD RORTY

Princeton University

COMMENTS AND CRITICISM

THALBERG'S DEFENSE OF JUSTIFIED TRUE BELIEF

IRVING THALBERG'S defense of the proposition that knowledge is justified true belief cannot succeed.*

Edmund Gettier had attacked the proposition in *Analysis* in 1963.† His argument was this. Consider a proposition p and a logical consequence of it q . Suppose that A is justified in believing p , that he has inferred q from p , and that he consequently believes q . He is therefore justified in believing q . (This is the principle Thalberg calls "*a principle of deducibility for justification*, abbreviated '(PDJ)' " [796].) But q may be true and p false. A false belief may be justified. If, therefore, knowledge were justified true belief, A would not know that p was true but would know that q was. And this is absurd. A may have no other grounds for q .

Thalberg questions (PDJ). This is his defense of the proposition Gettier had questioned. Thalberg affirms that knowledge is justified true belief. And he attacks the principle that Gettier had used in questioning this affirmation. I believe that this leads to difficulty.

* "In Defense of Justified True Belief," this JOURNAL, LXVI, 22 (Nov. 20, 1969): 794-803.

† "Is Justified True Belief Knowledge?," *Analysis*, xxiii.6, ns 96 (June 1963): 121-123.