

the review of metaphysics

a philosophical quarterly

Mind-Body Identity, Privacy, and Categories

Author(s): Richard Rorty

Source: *The Review of Metaphysics*, Vol. 19, No. 1 (Sep., 1965), pp. 24-54

Published by: Philosophy Education Society Inc.

Stable URL: <http://www.jstor.org/stable/20124096>

Accessed: 09/09/2009 16:20

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=pes>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.



Philosophy Education Society Inc. is collaborating with JSTOR to digitize, preserve and extend access to *The Review of Metaphysics*.

<http://www.jstor.org>

MIND-BODY IDENTITY, PRIVACY, AND CATEGORIES

RICHARD RORTY

1. *Introductory*

CURRENT CONTROVERSIES about the Mind-Body Identity Theory form a case-study for the investigation of the methods practiced by linguistic philosophers. Recent criticisms of these methods question that philosophers can discern lines of demarcation between "categories" of entities, and thereby diagnose "conceptual confusions" in "reductionist" philosophical theories. Such doubts arise once we see that it is very difficult, and perhaps impossible, to draw a firm line between the "conceptual" and the "empirical," and thus to differentiate between a statement embodying a conceptual confusion and one that expresses a surprising empirical result. The proponent of the Identity Theory (by which I mean one who thinks it sensible to assert that empirical inquiry will discover that *sensations* (not thoughts) are identical with certain brain-processes¹) holds that his opponents' arguments to the effect that empirical inquiry *could* not identify brain-processes and sensations are admirable illustrations of this difficulty. For, he argues, the classifications of linguistic expressions that are the ground of his opponents' criticism are classifications of a language which is as it is because it is the language spoken at a given stage of empirical inquiry. But the sort of empirical results that would show brain processes and sensations to be identical would also

¹ A proponent of the Identity Theory is usually thought of as one who predicts that empirical inquiry *will* reach this result—but few philosophers in fact stick their necks out in this way. The issue is not the truth of the prediction, but whether such a prediction makes sense. Consequently, by "Identity Theory" I shall mean the assertion that it does make sense.

I include only sensations within the scope of the theory because the inclusion of thoughts would raise a host of separate problems (about the reducibility of intentional and semantic discourse to statements about linguistic behavior), and because the form of the Identity Theory which has been most discussed in the recent literature restricts itself to a consideration of sensations.

bring about changes in our ways of speaking. These changes would make these classifications out of date. To argue against the Identity Theory on the basis of the way we talk now is like arguing against an assertion that supernatural phenomena are identical with certain natural phenomena on the basis of the way in which superstitious people talk. There is simply no such thing as a method of classifying linguistic expressions that has results guaranteed to remain intact despite the results of future empirical inquiry. Thus in this area (and perhaps in all areas) there is no method which will have the sort of magisterial neutrality of which linguistic philosophers fondly dream.

In this paper I wish to support this general line of argument. I shall begin by pressing the claims of the analogy between mental events and supernatural events. Then I shall try to rebut the objection which seems generally regarded as fatal to the claims of the Identity Theory—the objection that “privacy” is of the essence of mental events, and thus that a theory which holds that mental events might *not* be “private” is *ipso facto* confused. I shall conclude with some brief remarks on the implications of my arguments for the more general metaphilosophical issues at stake.

2. *The Two Forms of the Identity Theory*

The obvious objection to the Identity Theory is that “identical” either means a relation such that

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

(the relation of “strict identity”) or it does not. If it does, then we find ourselves forced into

saying truthfully that physical processes such as brain processes are dim or fading or nagging or false, and that mental phenomena such as after-images are publicly observable or physical or spatially located or swift,²

and thus using meaningless expressions, for

we may say that the above expressions are meaningless in the sense

² James Cornman, “The Identity of Mind and Body,” *Journal of Philosophy*, 59 (1962), p. 490.

that they commit a category mistake; i.e., in forming these expressions we have predicated predicates, appropriate to one logical category, of expressions that belong to a different logical category. This is surely a conceptual mistake.³

But if by "identical" the Identity Theory does *not* mean a relation of strict identity, then what relation *is* intended? How does it differ from the mere relation of "correlation" which, it is admitted on all sides, might without confusion be said to hold between sensations and brain-processes?

Given this dilemma, two forms of the identity theory may be distinguished. The first, which I shall call the *translation* form, grasps the first horn, and attempts to show that the odd-sounding expressions mentioned above do not involve category-mistakes, and that this can be shown by suitable translations into "topic neutral" language of the sentences in which these terms are originally used.⁴ The second, which I shall call the *disappearance* form, grasps the second horn, and holds that the relation in question is not strict identity, but rather the sort of relation which obtains between, to put it crudely, existent entities and non-existent entities when reference to the latter once served (some of) the purposes presently served by reference to the former—the sort of relation that holds, e.g., between "quantity of caloric fluid" and "mean kinetic energy of molecules." There is an obvious sense of "same" in which what used to be called "a quantity of caloric fluid" is *the same thing* as what is now called a certain mean kinetic energy of molecules, but there is no reason to think that all features truly predicated of the

³ Cornman, p. 491.

⁴ Cf. J. J. C. Smart, "Sensations and Brain Processes," reprinted in *The Philosophy of Mind*, ed. by V. C. Chappell (Englewood Cliffs, 1962), pp. 160-172, esp. pp. 166-68, and especially the claim that "When a person says 'I see a yellowish-orange after-image' he is saying something like this: 'There is something going on which is like what is going on when I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange'" (p. 167). For criticisms of Smart's program of translation, see Cornman, *op. cit.*; Jerome Shaffer, "Could Mental States Be Brain Processes?," *Journal of Philosophy*, 58 (1961), pp. 812-822; Shaffer, "Mental Events and the Brain," *Journal of Philosophy*, 60 (1963), pp. 160-166. See also the articles cited in the first footnote to Smart's own article.

one may be sensibly predicated of the other.⁵ The translation term of the theory holds that if we really understood what we were saying when we said things like "I am having a stabbing pain" we should see that since we are talking about "topic-neutral" matters, we might, for all we know, be talking about brain-processes. The disappearance form holds that it is unnecessary to show that suitable translations (into "topic-neutral" language) of our talk about sensations can be given—as unnecessary as to show that statements about quantities of caloric fluid, when properly understood, may be seen to be topic-neutral statements.⁶

From the point of view of this second form of the theory, it is a mistake to assume that "X's are nothing but Y's" entails "All attributes meaningfully predicable of X's are meaningfully predicated of Y's," for this assumption would forbid us ever to express the results of scientific inquiry in terms of (in Cornman's useful phrase) "cross-category identity."⁷ It would seem that the verb in such statements as "Zeus's thunderbolts are discharges of static electricity" and "Demonic possession is a form of hallucinatory psychosis" is the "is" of identity, yet it can hardly express *strict* identity. The disappearance form of the Identity Theory suggests that we view such statements as elliptical for e.g., "What people used to call 'demonic possession' is a form of hallucinatory psychosis," where the relation in question is strict identity. Since there is no reason why "what people call 'X'" should be in the

⁵ No statement of the disappearance form of the theory with which I am acquainted is as clear and explicit as Smart's statement of the translation form. See, however, Feyerabend, "Mental Events and the Brain," *Journal of Philosophy*, 60 (1963), pp. 295-296, and "Materialism and the Mind-Body Problem," *The Review of Metaphysics*, 17 (1963), pp. 49-67. See also Wilfrid Sellars, "The Identity Approach to the Mind-Body Problem," *ibid.*, 18 (1965). My indebtedness to this and other writings of Sellars will be obvious in what follows.

⁶ Both forms agree, however, on the requirements which would have to be satisfied if we are to claim that the empirical discovery in question has been made. Roughly, they are (1) that one-one or one-many correlations could be established between every type of sensation and some clearly demarcated kind(s) of brain-processes; (2) that every known law which refers to sensations would be subsumed under laws about brain-processes; (3) that new laws about sensations be discovered by deduction from laws about brain-processes.

⁷ Cornman, p. 492.

same "category" (in the Rylean sense) as "X," there is no need to claim, as the translation form of the theory must, that topic-neutral translations of statements using "X" are possible.

In what follows, I shall confine myself to a discussion and defense of the disappearance form of the theory. My first reason for this is that I believe that the analysis of "Sensations are identical with certain brain-processes" proposed by the disappearance form (viz., "What people now call 'sensations' are identical with certain brain-processes") accomplishes the same end as the translation form's program of topic-neutral translation—namely, avoiding the charge of "category-mistake," while preserving the full force of the traditional materialist position. My second reason is that I believe that an attempt to defend the translation form will inevitably get bogged down in controversy about the adequacy of proposed topic-neutral translations of statements about sensations. There is obviously a sense of "adequate translation" in which the topic-neutrality of the purported translations *ipso facto* makes them inadequate. So the proponent of the translation form of the theory will have to fall back on a weaker sense of "adequate translation." But the weaker this sense becomes, the less impressive is the claim being made, and the less difference between the Identity Theory and the non-controversial thesis that certain brain-processes may be constantly correlated with certain sensations.

3. *The Analogy between Demons and Sensations*

At first glance, there seems to be a fatal weakness in the disappearance form of the Identity Theory. For normally when we say "What people call 'X's' are nothing but Y's" we are prepared to add that "There are no X's." Thus when, e.g., we say that "What people call 'caloric fluid' is nothing but the motion of molecules" or "What people call 'witches' are nothing but psychotic women" we are prepared to say that there are no witches, and no such thing as caloric fluid. But it seems absurd to say that there might turn out to be no such things as sensations.

To see that this disanalogy is not fatal to the Identity Theory, let us consider the following situation. A certain primitive tribe holds the view that illnesses are caused by demons—a different

demon for each sort of illness. When asked what more is known about these demons than that they cause illness, they reply that certain members of the tribe—the witch-doctors—can see, after a meal of sacred mushrooms, various (intangible) humanoid forms on or near the bodies of patients. The witch-doctors have noted, for example, that a blue demon with a long nose accompanies epileptics, a fat red one accompanies sufferers from pneumonia, etc., etc. They know such further facts as that the fat red demon dislikes a certain sort of mold which the witch-doctors give people who have pneumonia. (There are various competing theories about what demons do when not causing diseases, but serious witch-doctors regard such speculations as unverifiable and profitless.)

If we encountered such a tribe, we would be inclined to tell them that there are no demons. We would tell them that diseases were caused by germs, viruses, and the like. We would add that the witch-doctors were not seeing demons, but merely having hallucinations. We would be quite right, but would we be right on *empirical* grounds? What empirical criteria, built into the demon-talk of the tribe, go unsatisfied? What predictions which the tribesmen make fail to come true? If there are none, a sophisticated witch-doctor may reply that all modern science can do is to show (1) that the presence of demons is constantly correlated with that of germs, viruses, and the like, and (2) that eating certain mushrooms sometimes makes people think they see things that aren't really there. This is hardly sufficient to show that there are no demons. At best, it shows that if we forget about demons, then (a) a simpler account of the cause and cure of disease and (b) a simpler account of why people make the perceptual reports they do, may be given.

What do we reply to such a sophisticated witch-doctor? I think that all that we would have left to say is that the simplicity of the accounts which can be offered if we forget about demons is an excellent reason for saying that there are no demons. Demon-discourse is one way of describing and predicting phenomena, but there are better ways. We *could* (as the witch-doctor urges) tack demon-discourse on to modern science by saying, first, that diseases are caused by the compresence of demons and germs (each

being a necessary, but neither a sufficient, condition) and, second, that the witch-doctors (unlike drunkards and psychotics) really do see intangible beings (about whom, alas, nothing is known save their visual appearances). If we did so, we would retain all the predictive and explanatory advantages of modern science. We would know as much about the cause and cure of disease, and about hallucinations, as we did before. We would, however, be burdened with problems which we did not have before: the problem of why demons are visible only to witch-doctors, and the problem of why germs cannot cause diseases all by themselves. We avoid both problems by saying that demons do not exist. The witch-doctor may remark that this use of Occam's Razor has the same advantage as that of theft over honest toil. To such a remark, the only reply could be an account of the practical advantages gained by the use of the Razor in the past.

Now the Identity Theorist's claim is that sensations may be to the future progress of psycho-physiology as demons are to modern science. Just as we now want to deny that there are demons, future science may want to deny that there are sensations. The only obstacle to replacing sensation-discourse with brain-discourse seems to be that sensation-statements have a reporting as well as an explanatory function. But the demon case makes clear that the discovery of a new way of explaining the phenomena previously explained by reference to a certain sort of entity, *combined with a new account of what is being reported by observation-statements about that sort of entity*, may give good reason for saying that there are no entities of that sort. The absurdity of saying "Nobody has ever felt a pain" is no greater than that of saying "Nobody has ever seen a demon," *if* we have a suitable answer to the question "What *was* I reporting when I said I felt a pain?" To this question, the science of the future may reply "You were reporting the occurrence of a certain brain-process, and it would make life simpler for us if you would, in the future, say 'My C-fibers are firing' instead of saying 'I'm in pain'." In so saying, he has as good a *prima facie* case as the scientist who answers the witch-doctor's question "What *was* I reporting when I reported a demon?" by saying "You were reporting the content of your

hallucination, and it would make life simpler if, in the future, you would describe your experiences in those terms."

Given this *prima facie* analogy between demons and sensations, we can now attend to some disanalogies. We may note, first, that there is no simple way of filling in the blank in "What people called 'demons' are nothing but —." For neither "hallucinatory contents" nor "germs" will do. The observational and the explanatory roles of "demon" must be distinguished. We need to say something like "What people who reported seeing demons were reporting was simply the content of their hallucinations," and *also* something like "What people explained by reference to demons can be explained better by reference to germs, viruses, etc." Because of the need for a relatively complex account of how we are to get along without reference to demons, we cannot *identify* "What we called 'demons'" with anything. So, instead, we simply deny their existence. In the case of sensations, however, we can give a relatively simple account of how to get along in the future. Both the explanatory *and* the reporting functions of statements about sensations can be taken over by statements about brain-processes. Therefore we are prepared to identify "What we called 'sensations'" with brain-processes, and to say "What we called 'sensations' turn out to be nothing but brain-processes."

Thus this disanalogy does not have the importance which it appears to have at first. In both the demon case and the sensation case, the proposed reduction has the same pragmatic consequences: namely, that we should stop asking questions about the causal and/or spatio-temporal relationships holding between the "reduced" entities (demons, sensations) and the rest of the universe, and replace these with questions about the relationships holding between certain other entities (germs, hallucinatory experiences, brain-processes) and the rest of the universe. It happens, for the reasons just sketched, that the proposed reduction is put in the form of a denial of existence in one case, and of an identification in another case. But "There are no demons" and "What people call 'sensations' are nothing but brain processes" can both equally well be paraphrased as "Elimination of the referring use of the expres-

sion in question ('demon,' 'sensation') from our language would leave our ability to describe and predict undiminished."

Nevertheless, the claim that there might turn out to be no such thing as a "sensation" seems scandalous. The fact that a witch-doctor might be scandalized by a similar claim about demons does not, in itself, do much to diminish our sense of shock. In what follows, I wish to account for this intuitive implausibility. I shall argue that it rests *solely* upon the fact that elimination of the referring use of "sensation" from our language would be in the highest degree *impractical*. If this can be shown, then I think that the Identity Theorist will be cleared of the charge of "conceptual confusion" usually leveled against him. Rather than proceeding directly to this argument, however, I shall first consider a line of argument which has often been used to show that he is guilty of this charge. Examining this line of argument will permit me to sketch in greater detail what the Identity Theorist is and is not saying.

4. *The Eliminability of Observation Terms*

The usual move made by the opponents of the Identity Theory is to compare suggested reduction of sensations to brain-processes to certain other cases in which we say that "X's turn out to be nothing but Y's." There are two significantly different classes of cases and it might seem that the Identity Theorist confuses them. First, there is the sort of case in which both "X" and "Y" are used to refer to observable entities, and the claim that "What people called 'X's' are nothing but Y's" backed up by pointing out that the statement "This is an X" commits one to an empirically false proposition. For example, we say that "What people called 'unicorn horns' are nothing but narwhal horns," and urge that we cease to respond to a perceptual situation with "This is a unicorn horn." We do this because "This is a unicorn horn" commits one to the existence of unicorns, and there are, it turns out, no unicorns. Let us call this sort of case *identification of observables with other observables*. Second, there is the sort of case in which "X" is used to refer to an observable entity and "Y" is used to refer to an unobservable entity. Here we do not (typically) back up the

claim that "What people called 'X's' are nothing but Y's" by citing an empirically false proposition presupposed by "This is an X." For example, the statement that "What people call 'tables' are nothing but clouds of molecules" does not suggest, or require as a ground, that people who say "This is a table" hold false beliefs. Rather, we are suggesting that something *more* has been found out about the sort of situation reported by "This is a table." Let us call this second sort of case *identification of observables with theoretical entities*.

It seems that we cannot assimilate the identification of sensations with brain-processes to either of these cases. For, unlike the typical case of identification of observables with other observables, we do not wish to say that people who have reported sensations in the past have (necessarily) any empirically disconfirmed beliefs. People are not wrong about sensations in the way in which they were wrong about "unicorn horns." Again, unlike the typical case of the identification of observables with theoretical entities, we do not want to say that brain-processes are "theoretical" or unobservable. Furthermore, in cases in which we identify an observable X with an unobservable Y, we are usually willing to accept the remark that "That does not show that there are no X's." The existence of tables is not (it would seem) impugned by their identification with clouds of electrons, as the existence of unicorn horns is impugned by their identification with narwhal horns. But a defender of the disappearance form of the Identity Theory *does* want to impugn the existence of sensations.

Because the claim that "What people call 'sensations' may turn out to be nothing but brain-processes" cannot be assimilated to either of these cases, it has been attacked as trivial or incoherent. The following dilemma is posed by those who attack it: either the Identity Theorist claims that talk about sensations presupposes some empirically disconfirmed belief (and what could it be?) or the "identity" which he has in mind is the uninteresting sort of identity which holds between tables and clouds of molecules (mere "theoretical replacability").

The point at which the Identity Theorist should attack this dilemma is the premiss invoked in stating the second horn—the premiss that the identification of tables with clouds of molecules

does not permit us to infer to the non-existence of tables. This premiss is true, but *why* is it true? That there is room for reflection here is apparent when we place the case of tables side-by-side with the case of demons. If there is any point to saying that tables are nothing but clouds of molecules it is presumably to say that, in principle, we could stop making a referring use of "table," and of any extensionally equivalent term, and still leave our ability to describe and predict undiminished. But this would seem just the point of (and the justification for) saying that there are no demons. Why does the realization that nothing would be lost by the dropping of "table" from our vocabulary still leave us with the conviction that there are tables, whereas the same realization about demons leave us with the conviction that there are no demons? I suggest that the only answer to this question which will stand examination is that although we could *in principle* drop "table," it would be monstrously inconvenient to do so, whereas it is both possible in principle and convenient in practice to drop "demon." The reason "But there still are tables" sounds so plausible is that nobody would dream of suggesting that we stop reporting our experiences in table-talk and start reporting them in molecule-talk. The reason "There are no demons" sounds so plausible is that we are quite willing to suggest that the witch-doctors stop reporting their experiences in demon-talk and start reporting them in hallucination-talk.

A conclusive argument that this practical difference is the *only* relevant difference would, obviously, canvass all the other differences which might be noted. I shall not attempt this. Instead, I shall try to make my claim plausible by sketching a general theory of the conditions under which a term may cease to have a referring use without those who made such a use being convicted of having held false beliefs.

Given the same sorts of correlations between X's and Y's, we are more likely to say "X's are nothing but Y's" when reference to X's is habitually made in non-inferential reports, and more likely to say "There are no 'X's'" when such reference is never or rarely made. (By "non-inferential report" I mean a statement in response to which questions like "How did you know?" "On what evidence do you say . . . ?" and "What leads you to think . . . ?"

are normally considered misplaced and unanswerable, but which is nonetheless capable of empirical confirmation.) Thus we do not say that the identification of temperature with the kinetic energy of molecules shows that there is no such thing as temperature, since "temperature" originally (i.e., before the invention of thermometers) stood for something which was always reported non-inferentially, and still is frequently so reported. Similarly for all identifications of familiar macro-objects with unfamiliar micro-objects. But since in our culture-circle we do not *habitually* report non-inferentially the presence of caloric fluid, demons, etc., we do not feel unhappy at the bald suggestion that there are no such things.

Roughly speaking, then, the more accustomed we are to "X" serving as an observation-term (by which I mean a term habitually used in non-inferential reports) the more we prefer, when inquiry shows the possibility of accounting for the phenomena explained by reference to X's without such reference, to "identify" X's with some sort of Y's, rather than to deny existence to X's *tout court*. *But the more grounds we have for such identification, the more chance there is that we shall stop using "X" in non-inferential reports*, and thus the greater chance of our eventually coming to accept the claim that "there are no X's" with equanimity. This is why we find borderline cases, and gradual shifts from assimilations of X's to Y's to an assertion that X's do not exist. For example, most people do not report the presence of pink rats non-inferentially (nor inferentially, for that matter), but some do. The recognition that they are in the minority helps those who do so to admit that there are no pink rats. But suppose that the vast majority of us had always seen (intangible and uncachable) pink rats; would it not then be likely that we should resist the bald assertion that there are no pink rats and insist on something of the form "pink rats are nothing but . . ." ? It might be a very long time before we came to drop the habit of reporting pink rats and began reporting hallucinations instead.

The typical case-history of an observation-term ceasing to have a referring use runs the following course: (1) X's are the subjects

of both inferential and non-inferential reports;^{*} (2) empirical discoveries are made which enable us to subsume X-laws under Y-laws and to produce new X-laws by studying Y's; (3) inferential reports of X's cease to be made; (4) non-inferential reports of X's are reinterpreted either (4a) as reports of Y's, or (4b) as reports of mental entities (thoughts that one is seeing an X, hallucinatory images, etc.); (5) non-inferential reports of X's cease to be made (because their place is taken by non-inferential reports either of Y's or of thoughts, hallucinatory images, etc.); (6) we conclude that there simply are no such things as X's.

This breakdown of stages lets us pick out two crucial conditions that must be satisfied if we are to move from "X's are nothing but Y's" (stage 2) to "there are no X's" (stage 6). These conditions are;

- (A) The Y-laws must be *better* at explaining the kinds of phenomena explained by the X-laws (not just equally good). Indeed, they must be sufficiently better so that *the inconvenience of changing one's linguistic habits by ceasing to make reports about X's is less than the inconvenience of going through the routine of translating one's X-reports into Y-reports in order to get satisfactory explanations of the phenomena in question*. If this condition is not satisfied, the move from stage (2) to stage (3) will not be made, and thus no later move will be made.
- (B) Either Y-reports may themselves be made non-inferentially, or X-reports may be treated as reports of mental entities. For we must be able to have some answer to the question "What *am* I reporting when I non-inferentially report about an X?", and the only answers available are "you're reporting on a Y" or "you're reporting on some merely mental entity." If neither answer is available, we can move neither to (4a) nor to (4b), nor, therefore, on to (5) and (6).

Now the reason we move from stage (2) to stage (3) in the case of demons is that (A) is obviously satisfied. The phenomena which we explained by reference to the activity of demons are so much better explained in other ways that it is simpler to stop inferring to the existence of demons altogether than to continue making such

^{*} Note that if X's are *only* referred to in inferential reports—as in the case of "neutrons" and "epicycles," no philosophically interesting reduction takes place. For in such cases there is no hope of getting rid of an explanandum; all we get rid of is a putative explanation.

inferences, and then turning to laws about germs and the like for an explanation of the behavior of the demons. The reason why we do *not* move from (2) to (3)—much less to (6)—in the case of temperature or tables is that explanations formulated in terms of temperatures are so good, on the ground which they were originally intended to cover, that we feel no temptation to stop talking about temperatures and tables merely because we can, in some cases, get more precise predictions by going up a level to laws about molecules. The reason why we move on from (3) to (4) in the case of demons is that the alternative labeled (4b) is readily available—we can easily consign experiences of demons to that great dumping-ground of out-dated entities, the Mind. There were no experiences of demons, we say, but only experiences of mental images.

Now it seems obvious that, in the case of sensations, (A) will not be satisfied. The inconvenience of ceasing to talk about sensations would be so great that only a fanatical materialist would think it worth the trouble to cease referring to sensations. If the Identity Theorist is taken to be predicting that some day “sensation,” “pain,” “mental image,” and the like will drop out of our vocabulary, he is almost certainly wrong. But if he is saying simply that, at no greater cost than an inconvenient linguistic reform, we *could* drop such terms, he is entirely justified. And I take this latter claim to be all that traditional materialism has ever desired.

Before leaving the analogy between demons and sensations, I wish to note one further disanalogy which an opponent of the Identity Theory might pounce upon. Even if we set aside the fact that (A) would not be satisfied in the case of sensations, such an opponent might say, we should note the difficulty in satisfying (B). It would seem that there is no satisfactory answer to the question “What *was* I non-inferentially reporting when I reported on my sensations?” For neither (4a) nor (4b) seems an available option. The first does not seem to be available because it is counter-intuitive to think of, e.g., “I am having my C-fibers stimulated,” as capable of being used to make a non-inferential report. The second alternative is simply silly—there is no point in saying that when we report a sensation we are reporting some “merely mental” event.

For sensations are *already* mental events. The last point is important for an understanding of the *prima facie* absurdity of the disappearance form of the Identity Theory. The reason why most statements of the form "there might turn out to be no X's at all" can be accepted with more or less equanimity in the context of forecasts of scientific results is that we are confident we shall always be able to "save the phenomena" by answering the question "But what about all those X's we've been accustomed to observe?" with some reference to thoughts-of X's, images-of-X's, and the like. Reference to mental entities provides non-inferential reports of X's with something to have been about. But when we want to say "There might turn out to be no mental entities at all," we cannot use this device. This result makes clear that if the analogy between the past disappearance of supernatural beings and the possible future disappearance of sensations is to be pressed, we must claim that alternative (4a) is, appearances to the contrary, still open. That is, we must hold that the question "What *was* I non-inferentially reporting when I non-inferentially reported a stabbing pain?" can be sensibly answered "You were reporting a stimulation of your C-fibers."

Now why should this *not* be a sensible answer? Let us begin by getting a bad objection to it out of the way. One can imagine someone arguing that this answer can only be given if a stimulation of C-fibers is strictly identical with a stabbing pain, and that such strict identification involves category-mistakes. But this objection presupposes that "A report of an X is a report of a Y" entails that "X's are Y's." If we grant this presupposition we shall not be able to say that the question "What was I reporting when I reported a demon?" is properly answered by "You were reporting the content of an hallucination which you were having." However, if we ask why this objection is plausible, we can see the grain of truth which it embodies and conceals. We are usually unwilling to accept "You were reporting a Y" as an answer to the question "What *was* I non-inferentially reporting when I non-inferentially reported an X?" unless (a) Y's are themselves the kind of thing we habitually report on non-inferentially, and (b) there does not exist already an habitual practice of reporting Y's non-inferentially. Thus we accept "the content of an hallucination"

as a sensible answer because we know that such contents, being "mental images," are just the sort of thing which does get non-inferentially reported (once it is recognized for what it is) and because we are not accustomed to making non-inferential reports in the form "I am having an hallucinatory image of. . . ."⁹ To take an example of answers to this sort of question that are *not* sensible, we reject the claim that when we report on a table we are reporting on a mass of whirling particles, for either we think we know under what circumstances we should make such a report, and know that these circumstances do not obtain, or we believe that the presence of such particles can only be inferred and never observed.

The oddity of saying that when I think I am reporting on a stabbing pain I am actually reporting on a stimulation of my C-fibers is similar to these last two cases. We either imagine a situation in which we can envisage ourselves non-inferentially reporting such stimulation (periscope hitched up to a microscope so as to give us a view of our trepanned skull, overlying fibers folded out of the way, stimulation evident by change in color, etc., etc.), or else we regard "stimulation of C-fibers" as not the sort of thing which *could* be the subject of a non-inferential report (but inherently a "theoretical" state of affairs whose existence can only be inferred, and not observed). In either case, the assertion that we have been non-inferentially reporting on a brain-process all our lives seems absurd. So the proponent of the disappearance form of the Identity Theory must show that reports of brain-processes are neither incapable of being non-inferential nor, if non-inferential, necessarily made in the way just imagined (with the periscope-microscope gadget) or in some other peculiar way. But now we must ask who bears the burden of proof. Why, after all, should we think that brain-processes are *not* a fit subject-

⁹ Note that people who *become* accustomed to making the latter sort of reports may no longer accept explanations of their erroneous non-inferential reports by reference to hallucinations. For they know what mental images are like, and they know that *this* pink rat was not an hallucinatory content. The more frequent case, fortunately, is that they just cease to report pink rats and begin reporting hallucinations, for their hallucinations no longer deceive them.

matter for non-inferential reports? And why should it not be the case that the circumstances in which we make non-inferential reports about brain-processes are just those circumstances in which we make non-inferential reports about sensations? For this will in fact be the case if, when we were trained to say, e.g., "I'm in pain" we were in fact being trained to respond to the occurrence within ourselves of a stimulation of C-fibers. If this is the case, the situation will be perfectly parallel to the case of demons and hallucinations. We *will*, indeed, have been making non-inferential reports about brain-processes all our lives *sans le savoir*.

This latter suggestion can hardly be rejected a priori, unless we hold that we can only be taught to respond to the occurrence of A's with the utterance "A!" if we were able, prior to this teaching, to be aware, when an A was present, that it was present. But this latter claim is plausible only if we assume that there is an activity which can reasonably be called "awareness" prior to the learning of language. I do not wish to fight once again the battle which has been fought by Wittgenstein and many of his followers against such a notion of awareness. I wish rather to take it as having been won, and to take for granted that there is no a priori reason why a brain-process is inherently unsuited to be the subject of a non-inferential report. The distinction between observation terms and non-observation-terms is relative to linguistic practices (practices which may change as inquiry progresses), rather than capable of being marked out once and for all by distinguishing between the "found" and the "made" elements in our experience. I think that the recognition of this relativity is the first of the steps necessary for a proper appreciation of the claims of the Identity Theory. In what follows, I want to show that this first step leads naturally to a second: the recognition that the distinction between *private* and *public* subject-matters is as relative as that between items signified by observation-terms and items not so signified.

The importance of this second step is clear. For even if we grant that reports of brain-processes may be non-inferential, we still need to get around the facts that reports of sensations have an epistemological peculiarity that leads us to call them reports of *private* entities, and that brain-processes are intrinsically *public*

entities. Unless we can overcome our intuitive conviction that a report of a private matter (with its attendant infallibility) cannot be identified with a report of a public matter (with its attendant fallibility), we shall not be able to take seriously the claim of the proponents of the disappearance form of the Identity Theory that alternative (4a) is open, and hence that nothing prevents sensations from disappearing from discourse in the same manner, and for the same reasons, as supernatural beings have disappeared from discourse. So far in this paper I have deliberately avoided the problem of the "privacy" of sensations, because I wished to show that if this problem *can* be surmounted, the Identity Theorist may fairly throw the burden of proof onto his opponent by asking whether a criterion can be produced which would show that the identification of sensations and brain-processes involves a conceptual confusion, while absolving the claim that demons do not exist because of such a confusion. Since I doubt that such a criterion *can* be produced, I am inclined to say that if the problem about "privacy" is overcome, then the Identity Theorist has made out his case.

5. *The "Privacy" Objection*

The problem that the privacy of first-person sensation reports presents for the Identity Theory has recently been formulated in considerable detail by Baier.¹⁰ In this section, I shall confine myself to a discussion of his criticism of Smart's initial reply to this argument. Smart holds that the fact that "the language of introspective reports has a different logic from the logic of material processes" is no objection to the Identity Theory, since we may expect that empirical inquiry can and will change this logic:

It is obvious that until the brain-process theory is much improved and widely accepted there will be no *criteria* for saying 'Smith has an experience of such-and-such a sort' except Smith's introspective reports. So we have adopted a rule of language that (normally) what Smith says goes.¹¹

¹⁰ Kurt Baier, "Smart on Sensations," *Australasian Journal of Philosophy*, 40 (1962), pp. 57-68.

¹¹ Smart, "Sensations and Brain Processes," p. 169.

Baier thinks that this reply "is simply a confusion of the privacy of the subject-matter and the availability of external evidence."¹² Baier's intuition is that the difference between a language-stratum in which the fact that a report is sincerely made is sufficient warrant for its truth, and one in which this situation does not obtain, seems so great as to call for an explanation—and that the only explanation is that the two strata concern different subject-matters. Indeed Baier is content to let the mental-physical distinction stand or fall with the distinction between "private" subject-matters and "public" subject-matters, and he therefore assumes that to show that "introspective reports are necessarily about something private, and that being about something private is *incompatible with being* about something public"¹³ is to show, once and for all, that the Identity Theory involves a conceptual confusion. Baier, in short, is undertaking to show that "once private, always private."

He argues for his view as follows:

To say that one day our physiological knowledge will increase to such an extent that we shall be able to make absolutely reliable encephalograph-based claims about people's experiences, is only to say that, if carefully checked, our encephalograph-based claims about 'experiences' will always be *correct*, i.e. will make the *same claims* as a *truthful* introspective reports. If correct encephalograph-based claims about Smith's experiences contradict Smith's introspective reports, we shall be entitled to infer that he is *lying*. In that sense, what Smith says will no longer go. But we cannot of course infer that he is making a mistake, for that is nonsense. . . . *However good the evidence may be, such a physiological theory can never be used to show to the sufferer that he was mistaken in thinking that he had a pain, for such a mistake is inconceivable.* The sufferer's epistemological authority must therefore be better than the best physiological theory can ever be. Physiology can therefore never provide a person with more than *evidence* that someone else is having an experience of one sort or another. It can never lay down *criteria* for saying that someone is having an experience of a certain sort. Talk about brain-processes therefore must be about something other than talk about experiences. Hence, introspective reports and brain process talk cannot be merely different ways of talking about the same thing.¹⁴

Smart's own reply to this line of argument is to admit that

¹² Baier, p. 63.

¹³ Baier, p. 59.

¹⁴ Baier, pp. 64-65; italics added.

No physiological evidence, say from a gadget attached to my skull, could make me withdraw the statement that I have a pain when as a matter of fact I feel a pain. For example, the gadget might show no suitable similarities of cerebral processes on the various occasions on which I felt a pain. . . . I must, I think, agree with Baier that if the sort of situation which we have just envisaged did in fact come about, then I should have to reject the brain process thesis, and would perhaps espouse dualism.¹⁵

But this is not the interesting case. The interesting case is the one in which suitable similarities are in fact found to occur—the same similarities in all subjects—until one day (long after all empirical generalizations about sensations *qua* sensations have been subsumed under physiological laws, and long after direct manipulation of the brain has become the exclusive method of relieving pain) somebody (call him Jones) thinks he has no pain, but the encephalograph says that the brain-process correlated with pain did occur. (Let us imagine that Jones himself is observing the gadget, and that the problem about whether he might have made a mistake is a problem for Jones; this eliminates the possibility of lying.) Now in most cases in which one's observation throws doubt on a correlation which is so central to current scientific explanations, one tries to eliminate the possibility of observational error. But in Baier's view it would be absurd for Jones to do this, for "a mistake is inconceivable." Actually, however, it is fairly clear what Jones' first move would be—he will begin to suspect that he does not know what pain is—i.e., that he is not using the word "pain" in the way in which his fellows use it.¹⁶

So now Jones looks about for independent verification of the hypothesis that he does not use "I am in pain" incorrectly. But here he runs up against the familiar difficulty about the vocabulary used in making introspective reports—the difficulty of distinguishing between "misuse of language" and "mistake in judgment", between (a) recognizing the state of affairs which obtains for what it is, but describing it wrongly because the words used in the description are not the right words, and (b) being able to describe

¹⁵ Smart, "Brain Processes and Incorrigibility—a Reply to Professor Baier," *Australasian Journal of Philosophy*, 40 (1962), p. 68.

¹⁶ This problem will remain, of course, even if Jones merely *thinks* about whether he is in pain, but does not say anything.

it rightly once it is recognized for what it is, but not in fact recognizing it for what it is (in the way in which one deceived by an illusion does not recognize the situation for what it is). If we do not have a way of determining which of these situations obtains, we do not have a genuine contrast between misnaming and misjudging. To see that there is no genuine contrast in this case, suppose that Jones was not burned prior to the time that he hitches on the encephalograph, but now he is. When he is, the encephalograph says that the brain-process constantly correlated with pain-reports occurs in Jones' brain. However, although he exhibits pain-behavior, Jones thinks that he does not feel pain. (But, now as in the past, he both exhibits pain-behavior and thinks that he feels pain when he is frozen, stuck, struck, racked, etc.) Now is it that he does not know that *pain* covers what you feel when you are burned as well as what you feel when you are stuck, struck, etc.? Or is it that he really does not feel pain when he is burned? Suppose we tell Jones that what he feels when he is burned is *also* called "pain." Suppose he then admits that he does feel *something*, but insists that what he feels is quite *different* from what he feels when he is stuck, struck, etc. Where does Jones go from here? Has he failed to learn the language properly, or is he correctly (indeed infallibly) reporting that he has different sensations than those normally had in the situation in question? (Compare the parallel question in the case of a man who uses "blue" in all the usual ways except that he refuses to grant that blue is a color—on the ground that it is so different from red, yellow, orange, violet, etc.)

The only device which would decide this question would be to establish a convention that anyone who sincerely denied that he felt a pain while exhibiting pain-behavior and being burned ipso facto did not understand how to use "pain." This denial would *prove* that he lacked such an understanding. But this would be a dangerous path to follow. For not to understand when to use the word "pain" in non-inferential reports is presumably to be unable to know which of one's sensations to call a "pain." And the denial that one felt pain in the circumstances mentioned would only prove such inability if one indeed *had* the sensation normally called a pain. So now we would have a public criterion, satis-

faction of which would count as showing that the subject had such a sensation—i.e., that he felt a pain even though he did not think that he did. But if such a criterion exists, its application overrides any contradictory report that he may make—for such a report will be automatically disallowed by the fact that it constitutes a demonstration that he does not know what he is talking about. The dilemma is that either a report about one's sensations which violates a certain public criterion is a sufficient condition for saying that the reporter does not know how to use "pain" in the correct way, or there is no such criterion. If there is, the fact that one cannot be mistaken about pains does not entail that sincere reports of pain cannot be over-ridden. If there is not, then there is no way to answer the question formulated at the end of the last paragraph, and hence no way to eliminate the possibility that Jones may not know what pain is. Now since the a priori probability that he does not is a good deal higher than the a priori probability that the psycho-physiological theory of Jones' era is mistaken, this theory has little to fear from Jones. (Although it would have a great deal to fear from a sizable accumulation of cases like Jones'.)

To sum up this point, we may look back at the italicized sentence in the above quotation from Baier. We now see that the claim that "such a mistake is inconceivable" is an ellipsis for the claim that a mistake, made *by one who knows what pain is*, is inconceivable, for only this expanded form will entail that when Jones and the encephalograph disagree, Jones is always right. But when formulated in this way our infallibility about our pains can be seen to be empty. Being infallible about something would be useful only if we could draw the usual distinction between misnaming and misjudging, and, having ascertained that we were not misnaming, know that we were not misjudging. But where there are no criteria for misjudging (or to put it more accurately, where in the crucial cases the criteria for misjudging turn out to be the same as the criteria for misnaming) than to say that we are infallible is to pay ourselves an empty compliment. Our neighbors will not hesitate to ride roughshod over our reports of our sensations unless they are assured that we know our way around among them, and we cannot satisfy them on this point unless, up

to a certain point, we tell the same sort of story about them as they do. The limits of permissible stories are flexible enough for us to be able to convince them occasionally that we have odd sensations, but not flexible enough for us to use these surprising sensations to break down, at one blow, well-confirmed scientific theories. As in the case of other infallible pronouncements, the price of retaining one's epistemological authority is a decent respect for the opinions of mankind.

Thus the common-sense remark that first-person reports always will be a better source of information about the occurrence of pains than any other source borrows its plausibility from the fact that we normally do not raise questions about a man's ability to use the word "pain" correctly. Once we *do* raise such questions seriously (as in the case of Jones), we realize that the question (1) "Does he know which sensations are called 'pains'?" and (2) "Is he a good judge of whether he is in pain or not?" are simply two ways of asking the same question: viz., "Can we fit his pain-reports into our scheme for explaining and predicting pains?" or, more bluntly, "Shall we disregard his pain-reports or not?" And once we see this we realize that if "always be a better source of information" means "will never be over-ridden on the sort of grounds on which presumed observational errors are over-ridden elsewhere in science," then our common-sensical remark is probably false. If "always be a better source of information" means merely "can only be over-ridden on the basis of a charge of misnaming, and never on the basis of a charge of misjudging," then our common-sensical remark turns out to depend upon a distinction that is not there.

This Wittgensteinian point that sensation-reports must conform to public criteria or else be disallowed may also be brought out in the following way. We determine whether to take a surprising first-person report of pain or its absence seriously (that is, whether to say that the sensation reported is something that science must try to explain) by seeing whether the reporter's overall pattern of pain-reporting is, by the usual behavioral and environmental criteria, normal. Now suppose that these public criteria (for "knowing how to use 'pain'") change as physiology and technology progress. Suppose, in particular, that we find it con-

venient to speed up the learning of contrastive observation predicates (such as "painful," "tickling," etc.) by supplying children with portable encephalographs-cum-teaching-machines which, whenever the appropriate brain-process occurs, murmur the appropriate term in their ears. Now "appropriate brain-process" will start out by meaning "brain-process constantly correlated with sincere utterances of 'I'm in pain' by people taught the use of 'pain' in the old rough-and-ready way." But soon it will come to mean, "the brain-process which we have always programmed the machine to respond to with a murmur of 'pain.'" (A meter is [now, but was not always] what matches the Standard Meter; intelligence is [now, but was not always] what intelligence tests test; pains will be [but are not now] what the Standard "Pain"-Training Program calls "pain.") Given this situation, it would make sense to say things like "You say you are in pain, and I'm sure you are sincere, but you can see for yourself that your brain is not in the state to which you were trained to respond to with "Pain," so apparently the training did not work, and you do not yet understand what pain is." In such a situation, our "inability to be mistaken" about our pains would remain, but our "final epistemological authority" on the subject would be gone, for there would be a standard procedure for overriding our reports. Our inability to be mistaken is, after all, no more than our ability to have such hypothetical statements as "If you admit that I'm sincere and that I know the language, you have to accept what I say" accepted by our fellows. But this asset can only be converted into final epistemological authority if we can secure both admissions. Where a clear-cut public criterion *does* exist for "knowing the language," inability to be mistaken does not entail inability to be over-ridden.

Now Baier might say that if such criteria did exist, then we should no longer be talking about what we presently mean by "pains." I do not think that this needs to be conceded,¹⁷ but suppose that it is. Would this mean that there was now a subject-

¹⁷ My reasons for thinking this concession unnecessary are the same as those presented in some recent articles by Hilary Putnam: cf. "Minds and Machines," *Dimensions of Mind*, ed. by S. Hook (New York, 1961), pp. 138-161, esp. pp. 153-160; "The Analytic and the Synthetic," *Minnesota*

matter which was not being discussed—viz., the private subject-matter the existence of which Baier's argument was intended to demonstrate.⁹ That we once had contact with such a subject-matter, but lost it.⁹ These rhetorical questions are meant to suggest that Baier's explanation of the final epistemological authority of first-person reports of pains by the fact that this "logic" is "a function of this type of subject-matter" rather than, as Smart thinks, a convention—is an explanation of the obscure by the more obscure. More precisely, it will not be an explanation of the epistemological authority in question—but only an unenlightening redescription of it—unless Baier can give a meaning to the term "private subject-matter" other than "kind of thing which is reported in reports which cannot be over-ridden." These considerations show the need for stepping back from Baier's argument and considering the criteria which he is using to demarcate distinct subject-matters.

6. "Privacy" as a Criterion of Categorical Demarcation

The closest Baier comes to giving a definition of "private subject-matter" is to say that

We must say that 'I have a pain' is about 'something private,' because in making this remark we report something which is (1) *necessarily owned* . . . (2) *necessarily exclusive and unsharable* . . . (3) *necessarily imperceptible by the senses* . . . (4) *necessarily asymmetrical*, for whereas it makes no sense to say 'I could see (or hear) that I had a pain', it makes quite good sense to say 'I could see (or hear) that *he* had a pain'; (5) something about the possession of which the person who claims to possess it could not possibly examine, consider, or weigh any evidence, although other people could . . . and lastly (6) it is something about which the person whose private state it is has final epistemological authority, for it does not make sense to say 'I have a pain unless I am mistaken.' ¹⁸

Now this definition of "something private" entails that nothing could be private except a state of a person, and is constructed to

Studies in the Philosophy of Science, III, pp. 358-397; "Brains and Behavior," in *Analytic Philosophy*, II, ed. by R. J. Butler (Oxford, 1965).

¹⁸ Baier, "Smart on Sensations," p. 60; the numbers in parentheses have been added.

delimit all and only those states of a person which we call his "mental" states. To say that mental states are private is to say simply that mental states are described in the way in which mental states are described. But it is not hard to take *any* Rylean category of terms (call it *C*), list all the types of sentence-frames which do and do not make sense when their gaps are filled with terms belonging to this category, and say that "something *C*" is distinguished by the fact that it is "necessarily *X*," "necessarily *Y*," etc. where "*X*" and "*Y*" are labels for the fact that certain sentence-frames will or will not receive these terms as gap-fillers. For example, consider the thesis that:

We must say that 'The devil is in that corner' is about 'something supernatural' because in making this report we report something which is *necessarily intangible*, since it makes no sense to ask about the texture of his skin, not *necessarily simply-located*, since it does not follow from the fact that a supernatural being is in the corner that the same supernatural being is not simultaneously at the other side of the globe, *necessarily immortal*, since it does not make sense to say that a supernatural being has died, *necessarily perceptible to exorcists*, since it would not make sense to say that a man was an exorcist and did not perceive the devil when he was present. . . .

Are devils hallucinations? No, because when one reports an hallucination one reports something which, though intangible, is simply-located, is neither mortal nor immortal, and is not always perceptible to exorcists. Are reports of devils reports of hallucinations? No, because reports of devils are reports of something supernatural and reports of hallucinations are reports of something private. It is simply because we lack further information about devils that we take exorcists' sincere reports as the best possible source for information about them? No, for this suggestion confuses the supernatural character of the subject-matter with the availability of external evidence. Those without the supernatural powers with which the exorcist is gifted may find ways of gathering *evidence* for the presence of supernatural beings, but they can never formulate an overriding and independent *criterion* for saying that such a being is present. Their theories might become so good that we might sometimes say that a given exorcist was *lying*, but we could never say that he was *mistaken*.

If this pastiche of Baier's argument seems beside the point, it is presumably either (1) because the language-game I have described is not in fact played, or else (2) because "necessarily intangible, not necessarily simply-located, necessarily immortal, and necessarily perceptible to exorcists" it does not delimit a subject-matter in the way in which "necessarily owned, exclusive, imperceptible by the senses, asymmetrical, etc., etc." does. In (1) one has to ask "what if it *had* been played?" After all, if the technique of detecting distinct subject-matters which Baier uses is a generally applicable technique, and not just constructed *ad hoc* to suit our Cartesian intuitions, then it ought to work on imaginary as well as real language games. But if it is, we ought to be able to formulate rules for applying it which would tell us *why* (2) is the case. For if we cannot, and if the language-game described once was played, then Baier's objection to the Identity Theory is an objection to the theory that reports of visible supernatural beings are reports of hallucinations.

Baier gives no more help in seeing what these rules would be. But I think that the root of Baier's conviction that "something private" is a suitable candidate for being a "distinct subject matter" is the thesis that certain terms are *intrinsically* observation predicates, and signify, so to speak, "natural explananda." When in quest of such predicates we look to the "foundations" of empirical knowledge, we tend to rapidly identify "observation predicate" with "predicate occurring in report having final epistemological authority" with "predicate occurring in report about something private." This chain of identifications leaves us with the suspicion that if there were no longer a private subject-matter to be infallible about, the whole fabric of empirical inquiry about public matters would be left up in the air, unsupported by any absolute epistemological authority. The suggestion that the distinction between items reportable in infallible reports and items not so reportable is "ultimate," or "irreducible," or "categorical," owes its intuitive force to the difficulty of imagining a stage in the progress of inquiry in which there was not *some* situation in which absolute epistemological authority about *something* would be granted to *somebody*.

There probably could *not* be such a stage, for inquiry cannot proceed if everything is to be doubted at once, and if inquiry is even to get off the ground we need to get straight about what is to be questioned and what not. These practical dictates show the kernel of truth in the notion that inquiry cannot proceed without a foundation. Where we slide from truth into error is in assuming that certain items are *naturally* reportable in infallible reports, and thus assume that the items presently so reportable always were and always will be reportable (and conversely for items not presently so reportable). A pain looks like the paradigm of such an item, with the situation described by "seems to me as if I were seeing something red" almost as well-qualified. But in both cases, we can imagine situations in which we should feel justified in over-riding sincere reports using these predicates. More important, we see that the device which we should use to justify ourselves in such situations—viz., "The reporter may not know how to use the word . . ."—is one which can apply in *all* proposed cases. Because this escape-hatch is always available, and because the question of whether the reporter does know how to use the word or not is probably not itself a question which could ever be settled by recourse to any absolute epistemological authority, the situation envisaged by Baier—namely, the body of current scientific theory foundering upon the rock of a single over-riding report—can probably never arise. Baier sees a difference in kind between the weight of evidence produced by such a theory and the single, authoritative, *criterion* provided by such a report. But since there can be no over-riding report until the ability of the speaker to use the words used in the report is established, and since this is to be established only by the weight of the evidence and not by recourse to any single criterion, this difference in kind (even though it may indeed be "firmly embedded in the way we talk" for millennia) is always capable of being softened into a difference of degree by further empirical inquiry.

7. *Reductionist Philosophical Theories and Categorical Distinctions*

In the preceding sections of this paper I have constantly invoked the fact that language changes as empirical discoveries are made, in order to argue that the thesis that "What people now call 'sensations' might be discovered to be brain-processes" is sensible and unconfused. The "deviance" of a statement of this thesis should not, I have been urging, blind us to the facts that (a) entities referred to by expressions in one Rylean category may also be referred to by expressions in another, (b) expressions in the first category may drop out of the language once this identity of reference is realized, and (c) the thesis in question is a natural way of expressing the result of this realization in the case of "sensation" and "brain-process." Now a critic might object that this strategy is subject to a *reductio ad absurdum*. For the same fact about linguistic change would seem to justify the claim that *any* statement of the form (S) "What people call 'X's' may be discovered to be Y's" is *always* sensible and unconfused. Yet this seems paradoxical, for consider the result of substituting, say "neutrino" for "X" and "mushroom" for "Y." If the resulting statement is not conceptually confused, what statement is?

In answer to this objection, I should argue that it is a mistake to attribute "conceptual confusions" to *statements*. No statement can be known to express a conceptual confusion simply by virtue of an acquaintance with the meanings of its component terms. Confusion is a property of people. Deviance is a property of utterances. Deviant utterances made by using sentences of the form (S) *may* betoken confusion on the part of the speaker about the meanings of words, but it may simply indicate a vivid (but unconfused) imagination, or perhaps (as in the neutrino-mushroom case) merely idle fancy. Although the making of such statements may be *prima facie* evidence of conceptual confusion—i.e., of the fact that the speaker is insufficiently familiar with the language to find a non-deviant way of making his point—this evidence is only *prima facie*, and questioning may bring out evidence pointing the other way. Such questioning may show that the speaker actually has some detailed suggestions about possible empir-

ical results which would point to the discovery in question, or that he has no such suggestions, but is nevertheless not inclined to use the relevant words in any *other* deviant utterances, and to cheerfully admit the deviance of his original utterance. The possibility of such evidence, pointing to imagination or to fancy rather than to confusion, shows that from the fact that certain questions are typically asked, and certain statements typically made, by victims of conceptual confusion, it does not follow that all those who use the sentences used to ask these questions or to make these statements are thus victimized.

This confusion about confusion is due to the fact that philosophers who propound "reductionist" theories (such as "There is no insensate matter," "There are no minds," "There are no physical objects," etc.) often *have* been conceptually confused. Such theories are often advocated as solutions to pseudo-problems whose very formulation involves deviant uses of words—uses which in fact result from a confusion between the uses of two or more senses of the same term, or between two or more related terms (e.g., "name" and "word") or between the kind of questions appropriately asked of entities referred to by one set of terms and the kind appropriately asked of entities referred to by another. (That these deviant uses *are* the result of such confusion, it should be noticed, is only capable of being determined by questioning of those who use them—and we only feel *completely* safe in making this diagnosis when the original user has, in the light of the linguistic facts drawn to his attention, admitted that his putative "problem" has been dissolved.) Because reductionist theories may often be choked off at the source by an examination of uses of language, anti-reductionist philosophers have lately become prone to use "conceptual confusion" or "category-mistake" as an all-purpose diagnosis for any deviant utterance in the mouth of a philosopher. But this is a mistake. Predictions of the sort illustrated by (S) may be turned to confused purposes, and they may be made by confused people. But we could only infer with certainty from the deviance of the utterance of a sentence of the form (S) to the conceptual confusion of the speaker if we had a map of the categories which are exhibited in all possible languages, and were thus in a position to say that the cross-category identification envisaged

by the statement was eternally impossible. In other words, we should only be in a position to make this inference with certainty if we knew that empirical inquiry could *never* bring about the sort of linguistic change which permits the non-deviant use of "There are no X's" in the case of the "X's" to which the statement in question refers. But philosophers are in no position to say that such change is impossible. The hunt for categoreal confusions at the source of reductionist philosophical theories is an extremely valuable enterprise. But their successes in this enterprise should not lead linguistic philosophers to think that they can do better what metaphysicians did badly—namely, prove the irreducibility of entities. Traditional materialism embodied many confusions, but at its heart was the unconfused prediction about future empirical inquiry which is the Identity Theory. The confusions may be eradicated without affecting the plausibility or interest of the prediction.¹⁹

Princeton University.

¹⁹ I have been greatly helped in preparing this paper by the comments of Richard Bernstein, Keith Gunderson, Amélie Rorty, and Richard Schmitt.