Individualism and the Mental

TYLER BURGE

Since Hegel's *Phenomenology of Spirit*, a broad, inarticulate division of emphasis between the individual and his social environment has marked philosophical discussions of mind. On one hand, there is the traditional concern with the individual subject of mental states and events. In the elderly Cartesian tradition, the spotlight is on what exists or transpires "in" the individual—his secret cogitations, his innate cognitive structures, his private perceptions and introspections, his grasping of ideas, concepts, or forms. More evidentially oriented movements, such as behaviorism and its liberalized progeny, have highlighted the individual's publicly observable behavior—his input-output relations and the dispositions, states, or events that mediate them. But both Cartesian and behaviorist viewpoints tend to feature the individual subject. On the other hand, there is the Hegelian preoccupation with the role of social institutions in shaping the individual and the content of his thought. This tradition has dominated the continent since Hegel. But it has found echoes in English-speaking philosophy during this century in the form of a concentration on language. Much philosophical work on language and mind has been in the interests of Cartesian or behaviorist viewpoints that I shall term "individualistic." But many of Wittgenstein's remarks about mental representation point up a social orientation that is discernible from his flirtations with behaviorism. And more recent work on the theory of reference has provided glimpses of the role of social cooperation in determining what an individual thinks.

In many respects, of course, these emphases within philosophy—individualistic and social—are compatible. To an extent, they may be regarded simply as different currents in the turbulent stream of ideas that has washed the intellectual landscape during the last hundred and some odd years. But the role of the social environment has received considerably less clear-headed philosophical attention (though perhaps not less philosophical attention) than the role of the states, occurrences, or acts in, on, or by the individual. Philosophical discussions of social factors

have tended to be obscure, evocative, metaphorical, or platitudinous, or to be bent on establishing some large thesis about the course of history and the destiny of man. There remains much room for sharp delineation. I shall offer some considerations that stress social factors in descriptions of an individual's mental phenomena. These considerations call into question individualistic presuppositions of several traditional and modern treatments of mind. I shall conclude with some remarks about mental models.

I. TERMINOLOGICAL MATTERS

Our ordinary mentalistic discourse divides broadly into two sorts of idiom. One typically makes reference to mental states or events in terms of sentential expressions. The other does not. A clear case of the first kind of idiom is 'Alfred thinks that his friends' sofa is ugly'. A clear case of the second sort is 'Alfred is in pain'. Thoughts, beliefs, intentions, and so forth are typically specified in terms of subordinate sentential clauses, that-clauses, which may be judged as true or false. Pains, feels, tickles, and so forth have no special semantical relation to sentences or to truth or falsity. There are intentional idioms that fall in the second category on this characterization, but that share important semantical features with expressions in the first—idioms like 'Al worships Buicks'. But I shall not sort these out here. I shall discuss only the former kind of mentalistic idiom. The extension of the discussion to other intentional idioms will not be difficult.

In an ordinary sense, the noun phrases that embed sentential expressions in mentalistic idioms provide the *content* of the mental state or event. We shall call that-clauses and their grammatical variants "*content clauses.*" Thus the expression 'that sofas are more comfortable than pews' provides the content of Alfred's belief that sofas are more comfortable than pews. My phrase 'provides the content' represents an attempt at remaining neutral, at least for present purposes, among various semantical and metaphysical accounts of precisely how that-clauses function and precisely what, if anything, contents are.

Although the notion of content is, for present purposes, ontologically neutral, I do think of it as holding a place in a systematic *theory* of mentalistic language. The question of when to count contents different, and when the same, is answerable to theoretical restrictions. It is often remarked that in a given context we may ascribe to a person two that-clauses that are only loosely equivalent and count them as attributions of the "same attitude." We may say that Al's intention to climb Mt. McKinley and his intention to climb the highest mountain in the United States are the "same intention." (I intend the terms for the mountain to occur obliquely here. See later discussion.) This sort of point extends even to content clauses with extensionally non-equivalent counterpart notions. For contextually relevant purposes, we might count a thought that the glass contains some water as "the same thought" as a thought that the glass contains some thirst-quenching liquid, particularly if we have no reason to attribute either content as opposed to the other, and distinctions between them are contextually irrelevant. Nevertheless, in both these examples,

every systematic theory I know of would want to represent the semantical contribution of the content-clauses in distinguishable ways—as "providing different contents."

One reason for doing so is that the person himself is capable of having different attitudes described by the different content-clauses, even if these differences are irrelevant in a particular context. (Al might have developed the intention to climb the highest mountain before developing the intention to climb Mt. McKinley—regardless of whether he, in fact, did so.) A second reason is that the counterpart components of the that-clauses allude to distinguishable elements in people's cognitive lives. 'Mt. McKinley' and 'the highest mountain in the U.S.' serve, or might serve, to indicate cognitively different notions. This is a vague, informal way of generalizing Frege's point: the thought that Mt. McKinley is the highest mountain in the U.S. is potentially interesting or informative. The thought that Mt. McKinley is Mt. McKinley is not. Thus when we say in a given context that attribution of different contents is attribution of the "same attitude," we use 'same attitude' in a way similar to the way we use 'same car' when we say that people who drive Fords (or green 1970 Ford Mavericks) drive the "same car." For contextual purposes different cars are counted as "amounting to the same."

Although this use of 'content' is theoretical, it is not I think theoretically controversial. In cases where we shall be counting contents different, the cases will be uncontentious: On any systematic theory, differences in the *extension*—the actual denotation, referent, or application—of counterpart expressions in that-clauses will be semantically represented, and will, in our terms, make for differences in content. I shall be avoiding the more controversial, but interesting, questions about the general conditions under which sentences in that-clauses can be expected to provide the same content.

I should also warn of some subsidiary terms. I shall be (and have been) using the term '*notion*' to apply to components or elements of contents. Just as whole that-clauses provide the content of a person's attitude, semantically relevant components of that-clauses will be taken to indicate notions that enter into the attitude (or the attitude's content). This term is supposed to be just as ontologically neutral as its fellow. When I talk of understanding or mastering the notion of contract, I am not relying on any special epistemic or ontological theory, except insofar as the earlier-mentioned theoretical restrictions on the notion of content are inherited by the notion of notion. The expression, '*understanding (mastering) a notion*' is to be construed more or less intuitively. Understanding the notion of contract comes roughly to knowing what a contract is. One can master the notion of contract without mastering the term 'contract'—at the very least if one speaks some language other than English that has a term roughly synonymous with 'contract'. (An analogous point holds for my use of 'mastering a content'.) Talk of notions is roughly similar to talk of concepts in an informal sense. 'Notion' has the advantage of being easier to separate from traditional theoretical commitments.

I speak of *attributing* an attitude, content, or notion, and of *ascribing* a that-clause or other piece of language. Ascriptions are the linguistic analogs of attributions. This use of 'ascribe' is nonstandard, but convenient and easily assimilated.

There are semantical complexities involving the behavior of expressions in content clauses, most of which we can skirt. But some must be touched on. Basic to the subject is the observation that expressions in content clauses are often not intersubstitutable with extensionally equivalent expressions in such a way as to maintain the truth value of the containing sentence. Thus from the facts that water is H_2O and that Bertrand thought that water is not fit to drink, it does not follow that Bertrand thought that H_2O is not fit to drink. When an expression like 'water' functions in a content clause so that it is not freely exchangeable with all extensionally equivalent expressions, we shall say that it has *oblique occurrence*. Roughly speaking, the reason why 'water' and 'H_2O' are not interchangeable in our report of Bertrand's thought is that 'water' plays a role in characterizing a different mental act or state from that which 'H_2O' would play a role in characterizing. In this context at least, thinking that water is not fit to drink is different from thinking that H_2O is not fit to drink.

By contrast, there are non-oblique occurrences of expressions in content clauses. One might say that some water—say, the water in the glass over there—is thought by Bertrand to be impure; or that Bertrand thought that *that* water is impure. And one might intend to make no distinction that would be lost by replacing 'water' with 'H_2O'—or 'that water' with 'that H_2O' or 'that common liquid', or any other expression extensionally equivalent with 'that water'. We might allow these exchanges even though Bertrand had never heard of, say, H_2O. In such purely nonoblique occurrences, 'water' plays *no role* in providing the *content* of Bertrand's thought, *on our use of 'content'*, or (in any narrow sense) in characterizing Bertrand or his mental state. Nor is the water part of Bertrand's thought content. We speak of Bertrand *thinking his content of* the water. At its nonoblique occurrence, the term 'that water' simply isolates, in one of many equally good ways, a portion of wet stuff to which Bertrand or his thought is related or applied. In certain cases, it may also mark a context in which Bertrand's thought is applied. But it is expressions at oblique occurrences within content clauses that primarily do the job of providing the content of mental states or events, and in characterizing the person.

Mentalistic discourse containing obliquely occurring expressions has traditionally been called *intentional discourse*. The historical reasons for this nomenclature are complex and partly confused. But roughly speaking, grammatical contexts involving oblique occurrences have been fixed upon as specially relevant to the representational character (sometimes called "intentionality") of mental states and events. Clearly oblique occurrences in mentalistic discourse have something to do with characterizing a person's epistemic perspective—how things seem to him, or in an informal sense, how they are represented to him. So without endorsing all the commitments of this tradition, I shall take over its terminology.

The crucial point in the preceding discussion is the assumption that obliquely occurring expressions in content clauses are a primary means of identifying a person's intentional mental states or events. A further point is worth remarking here. It is normal to suppose that those content clauses correctly ascribable to a person that are not in general intersubstitutable *salva veritate*—and certainly those that involve

extensionally non-equivalent counterpart expressions—identify different mental states or events.

I have cited contextual exceptions to this normal supposition, at least in a manner of speaking. We sometimes count distinctions in content irrelevant for purposes of a given attribution, particularly where our evidence for the precise content of a person or animal's attitude is skimpy. Different contents may contextually identify (what amount to) the "same attitude." I have indicated that even in these contexts, I think it best, strictly speaking, to construe distinct contents as describing different mental states or events that are merely equivalent for the purposes at hand. I believe that this view is widely accepted. But nothing I say will depend on it. For any distinct contents, there will be imaginable contexts of attribution in which, even in the loosest, most informal ways of speaking, those contents would be said to describe different mental states or events. This is virtually a consequence of the theoretical role of contents, discussed earlier. Since our discussion will have an "in principle" character, I shall take these contexts to be the relevant ones. Most of the cases we discuss will involve *extensional* differences between obliquely occurring counterpart expressions in that-clauses. In such cases, it is particularly natural and normal to take different contents as identifying different mental states or events.

II. A THOUGHT EXPERIMENT

IIa. First Case

We now turn to a three-step thought experiment. Suppose first that:

> A given person has a large number of attitudes commonly attributed with content clauses containing 'arthritis' in oblique occurrence. For example, he thinks (correctly) that he has had arthritis for years, that his arthritis in his wrists and fingers is more painful than his arthritis in his ankles, that it is better to have arthritis than cancer of the liver, that stiffening joints is a symptom of arthritis, that certain sorts of aches are characteristic of arthritis, that there are various kinds of arthritis, and so forth. In short, he has a wide range of such attitudes. In addition to these unsurprising attitudes, he thinks falsely that he has developed arthritis in the thigh.

Generally competent in English, rational and intelligent, the patient reports to his doctor his fear that his arthritis has now lodged in his thigh. The doctor replies by telling him that this cannot be so, since arthritis is specifically an inflammation of joints. Any dictionary could have told him the same. The patient is surprised, but relinquishes his view and goes on to ask what might be wrong with his thigh.

The second step of the thought experiment consists of a counterfactual supposition. We are to conceive of a situation in which the patient proceeds from birth through the same course of physical events that he actually does, right to and including the time at which he first reports his fear to his doctor. Precisely the same

things (non-intentionally described) happen to him. He has the same physiological history, the same diseases, the same internal physical occurrences. He goes through the same motions, engages in the same behavior, has the same sensory intake (physiologically described). His dispositions to respond to stimuli are explained in physical theory as the effects of the same proximate causes. All of this extends to his interaction with linguistic expressions. He says and hears the same words (word forms) at the same times he actually does. He develops the disposition to assent to 'Arthritis can occur in the thigh' and 'I have arthritis in the thigh' as a result of the same physically described proximate causes. Such dispositions might have arisen in a number of ways. But we can suppose that in both actual and counterfactual situations, he acquires the word 'arthritis' from casual conversation or reading, and never hearing anything to prejudice him for or against applying it in the way that he does, he applies the word to an ailment in his thigh (or to ailments in the limbs of others) which seems to produce pains or other symptoms roughly similar to the disease in his hands and ankles. In both actual and counterfactual cases, the disposition is never reinforced or extinguished up until the time when he expresses himself to his doctor. We further imagine that the patient's non-intentional, phenomenal experience is the same. He has the same pains, visual fields, images, and internal verbal rehearsals. The *counterfactuality* in the supposition touches only the patient's social environment. In actual fact, 'arthritis', as used in his community, does not apply to ailments outside joints. Indeed, it fails to do so by a standard, non-technical dictionary definition. But in our imagined case, physicians, lexicographers, and informed laymen apply 'arthritis' not only to arthritis but to various other rheumatoid ailments. The standard use of the term is to be conceived to encompass the patient's actual misuse. We could imagine either that arthritis had not been singled out as a family of diseases, or that some other term besides 'arthritis' were applied, though not commonly by laymen, specifically to arthritis. We may also suppose that this difference and those necessarily associated with it are the only differences between the counterfactual situation and the actual one. (Other people besides the patient will, of course, behave differently.) To summarize the second step:

> The person might have had the same physical history and non-intentional mental phenomena while the word 'arthritis' was conventionally applied, and defined to apply, to various rheumatoid ailments, including the one in the person's thigh, as well as to arthritis.

The final step is an interpretation of the counterfactual case, or an addition to it as so far described. It is reasonable to suppose that:

> In the counterfactual situation, the patient lacks some—probably *all*—of the attitudes commonly attributed with content clauses containing 'arthritis' in oblique occurrence. He lacks the occurrent thoughts or beliefs that he has arthritis in the thigh, that he has had arthritis for years, that stiffening joints and various sorts of aches are symptoms of arthritis, that his father had arthritis, and so on.

We suppose that in the counterfactual case we cannot correctly ascribe any content clause containing an oblique occurrence of the term 'arthritis'. It is hard to see how the patient could have picked up the notion of arthritis. The word 'arthritis' in the counterfactual community does not mean *arthritis*. It does not apply only to inflammations of joints. We suppose that no other word in the patient's repertoire means *arthritis*. 'Arthritis', in the counterfactual situation, differs both in dictionary definition and in extension from 'arthritis' as we use it. Our ascriptions of content clauses to the patient (and ascriptions within his community) would not constitute attributions of the same contents we actually attribute. For counterpart expressions in the content clauses that are actually and counterfactually ascribable are not even extensionally equivalent. However we describe the patient's attitudes in the counterfactual situation, it will not be with a term or phrase extensionally equivalent with 'arthritis'. So the patient's counterfactual attitude contents differ from his actual ones.

The upshot of these reflections is that the patient's mental contents differ while his entire physical and non-intentional mental histories, considered in isolation from their social context, remain the same. (We could have supposed that he dropped dead at the time he first expressed his fear to the doctor.) The differences seem to stem from differences "outside" the patient considered as an isolated physical organism, causal mechanism, or seat of consciousness. The difference in his mental contents is attributable to differences in his social environment. In sum, the patient's internal qualitative experiences, his physiological states and events, his behaviorally described stimuli and responses, his dispositions to behave, and whatever sequences of states (non-intentionally described) mediated his input and output—all these remain constant, while his attitude contents differ, even in the extensions of counterpart notions. As we observed at the outset, such differences are ordinarily taken to spell differences in mental states and events.

IIb. Further Exemplifications

The argument has an extremely wide application. It does not depend, for example, on the kind of word 'arthritis' is. We could have used an artifact term, an ordinary natural kind word, a color adjective, a social role term, a term for a historical style, an abstract noun, an action verb, a physical movement verb, or any of various other sorts of words. I prefer to leave open precisely how far one can generalize the argument. But I think it has a very wide scope. The argument can get under way in any case where it is intuitively possible to attribute a mental state or event whose content involves a notion that the subject incompletely understands. As will become clear, this possibility is the key to the thought experiment. I want to give a more concrete sense of the possibility before going further.

It is useful to reflect on the number and variety of intuitively clear cases in which it is normal to attribute a content that the subject incompletely understands. One need only thumb through a dictionary for an hour or so to develop a sense of the extent to which one's beliefs are infected by incomplete understanding.[1] The phenomenon is rampant in our pluralistic age.

a. Most cases of incomplete understanding that support the thought experiment will be fairly idiosyncratic. There is a reason for this. Common linguistic errors, if entrenched, tend to become common usage. But a generally competent speaker is bound to have numerous words in his repertoire, possibly even common words, that he somewhat misconstrues. Many of these misconstruals will not be such as to deflect ordinary ascriptions of that-clauses involving the incompletely mastered term in oblique occurrence. For example, one can imagine a generally competent, rational adult having a large number of attitudes involving the notion of sofa—including beliefs that *those* (some sofas) are sofas, that some sofas are beige, that his neighbors have a new sofa, that he would rather sit in a sofa for an hour than on a church pew. In addition, he might think that sufficiently broad (but single-seat) overstuffed armchairs are sofas. With care, one can develop a thought experiment parallel to the one in section IIa, in which at least some of the person's attitude contents (particularly, in this case, contents of occurrent mental events) differ, while his physical history, dispositions to behavior, and phenomenal experience—non-intentionally and asocially described—remain the same.

b. Although most relevant misconstruals are fairly idiosyncratic, there do seem to be certain types of error which are relatively common—but not so common and uniform as to suggest that the relevant terms take on new sense. Much of our vocabulary is taken over from others who, being specialists, understand our terms better than we do.[2] The use of scientific terms by laymen is a rich source of cases. As the arthritis example illustrates, the thought experiment does not depend on specially technical terms. I shall leave it to the imagination of the reader to spin out further examples of this sort.

c. One need not look to the laymen's acquisitions from science for examples. People used to buying beef brisket in stores or ordering it in restaurants (and conversant with it in a general way) probably often develop mistaken beliefs (or uncertainties) about just what brisket is. For example, one might think that brisket is a cut from the flank or rump, or that it includes not only the lower part of the chest but also the upper part, or that it is specifically a cut of beef and not of, say, pork. No one hesitates to ascribe to such people content-clauses with 'brisket' in oblique occurrence. For example, a person may believe that he is eating brisket under these circumstances (where 'brisket' occurs in oblique position); or he may think that brisket tends to be tougher than loin. Some of these attitudes may be false; many will be true. We can imagine a counterfactual case in which the person's physical history, his dispositions, and his non-intentional mental life, are all the same, but in which 'brisket' is commonly applied in a different way—perhaps in precisely the way the person thinks it applies. For example, it might apply only to beef and to the upper and lower parts of the chest. In such a case, as in the sofa and arthritis cases, it would seem that the person would (or might) lack some or all of the propositional attitudes that are actually attributed with content clauses involving 'brisket' in oblique position.

d. Someone only generally versed in music history, or superficially acquainted with a few drawings of musical instruments, might naturally but mistakenly come to

think that clavichords included harpsichords without legs. He may have many other beliefs involving the notion of clavichord, and many of these may be true. Again, with some care, a relevant thought experiment can be generated.

e. A fairly common mistake among lawyers' clients is to think that one cannot have a contract with someone unless there has been a written agreement. The client might be clear in intending 'contract' (in the relevant sense) to apply to agreements, not to pieces of paper. Yet he may take it as part of the meaning of the word, or the essence of law, that a piece of formal writing is a necessary condition for establishing a contract. His only experiences with contracts might have involved formal documents, and he undergeneralizes. It is not terribly important here whether one says that the client misunderstands the term's meaning, or alternatively that the client makes a mistake about the essence of contracts. In either case, he misconceives what a contract is; yet ascriptions involving the term in oblique position are made anyway.

It is worth emphasizing here that I intend the misconception to involve the subject's attaching counterfactual consequences to his mistaken belief about contracts. Let me elaborate this a bit. A common dictionary definition of 'contract' is 'legally binding agreement'. As I am imagining the case, the client does not explicitly define 'contract' to himself in this way (though he might use this phrase in explicating the term). And he is not merely making a mistake about what the law happens to enforce. If asked why unwritten agreements are not contracts, he is likely to say something like, 'They just aren't' or 'It is part of the nature of the law and legal practice that they have no force'. He is not disposed without prodding to answer, 'It would be possible but impractical to give unwritten agreements legal force'. He might concede this. But he would add that such agreements would not be contracts. He regards a document as inseparable from contractual obligation, regardless of whether he takes this to be a matter of meaning or a metaphysical essentialist truth about contracts.

Needless to say, these niceties are philosopher's distinctions. They are not something an ordinary man is likely to have strong opinions about. My point is that the thought experiment is independent of these distinctions. It does not depend on misunderstandings of dictionary meaning. One might say that the client understood the term's dictionary meaning, but misunderstood its essential application in the law—misconceived the nature of contracts. The thought experiment still flies. In a counterfactual case in which the law enforces both written and unwritten agreements and in which the subject's behavior and so forth are the same, but in which 'contract' *means* 'legally binding agreement based on written document', we would not attribute to him a mistaken belief that a contract requires written agreement, although the lawyer might have to point out that there are other legally binding agreements that do not require documents. Similarly, the client's other propositional attitudes would no longer involve the notion of contract, but another more restricted notion.

f. People sometimes make mistakes about color ranges. They may correctly apply a color term to a certain color, but also mistakenly apply it to shades of a neighboring

color. When asked to explain the color term, they cite the standard cases (for 'red', the color of blood, fire engines, and so forth). But they apply the term somewhat beyond its conventionally established range—beyond the reach of its vague borders. They think that fire engines, including *that* one, are red. They observe that red roses are covering the trellis. But they also think that *those* things are a shade of red (whereas they are not). Second looks do not change their opinion. But they give in when other speakers confidently correct them in unison.

This case extends the point of the contract example. The error is linguistic or conceptual in something like the way that the shopper's mistake involving the notion of brisket is. It is not an ordinary empirical error. But one may reasonably doubt that the subjects misunderstand the dictionary meaning of the color term. Holding their non-intentional phenomenal experience, physical history, and behavioral dispositions constant, we can imagine that 'red' were applied as they mistakenly apply it. In such cases, we would no longer ascribe content-clauses involving the term 'red' in oblique position. The attribution of the correct beliefs about fire engines and roses would be no less affected than the attribution of the beliefs that, in the actual case, display the misapplication. Cases bearing out the latter point are common in anthropological reports on communities whose color terms do not match ours. Attributions of content typically allow for the differences in conventionally established color ranges.

Here is not the place to refine our rough distinctions among the various kinds of misconceptions that serve the thought experiment. Our philosophical purposes do not depend on how these distinctions are drawn. Still, it is important to see what an array of conceptual errors is common among us. And it is important to note that such errors do not always or automatically prevent attribution of mental content provided by the very terms that are incompletely understood or misapplied. The thought experiment is nourished by this aspect of common practice.

IIc. Expansion and Delineation of the Thought Experiment

As I have tried to suggest in the preceding examples, the relevant attributions in the first step of the thought experiment need not display the subject's error. They may be attributions of a true content. We can begin with a propositional attitude that involved the misconceived notion, but in a true, unproblematic application of it: for example, the patient's belief that he, like his father, developed arthritis in the ankles and wrists at age 58 (where 'arthritis' occurs obliquely).

One need not even rely on an underlying *mis*conception in the thought experiment. One may pick a case in which the subject only partially understands an expression. He may apply it firmly and correctly in a range of cases, but be unclear or agnostic about certain of its applications or implications which, in fact, are fully established in common practice. Most of the examples we gave previously can be reinterpreted in this way. To take a new one, imagine that our protagonist is unsure whether his father has mortgages on the car and house, or just one on the house. He is a little uncertain about exactly how the loan and collateral must be arranged in

order for there to be a mortgage, and he is not clear about whether one may have mortgages on anything other than houses. He is sure, however, that Uncle Harry paid off his mortgage. Imagine our man constant in the ways previously indicated and that 'mortgage' commonly applied only to mortgages on houses. But imagine banking practices themselves to be the same. Then the subject's uncertainty would plausibly not involve the notion of mortgage. Nor would his other propositional attitudes be correctly attributed with the term 'mortgage' in oblique position. Partial understanding is as good as misunderstanding for our purposes.

On the other hand, the thought experiment does appear to depend on the possibility of someone's having a propositional attitude despite an incomplete mastery of some notion in its content. To see why this appears to be so, let us try to run through a thought experiment, attempting to avoid any imputation of incomplete understanding. Suppose the subject thinks falsely that all swans are white. One can certainly hold the features of swans and the subject's non-intentional phenomenal experience, physical history, and non-intentional dispositions constant, and imagine that 'swan' meant 'white swan' (and perhaps some other term, unfamiliar to the subject, meant what 'swan' means). Could one reasonably interpret the subject as having different attitude contents without at some point invoking a misconception? The questions to be asked here are about the subject's dispositions. For example, in the actual case, if he were shown a black swan and told that he was wrong, would he fairly naturally concede his mistake? Or would he respond, "I'm doubtful that that's a swan," until we brought in dictionaries, encyclopedias, and other native speakers to correct his usage? In the latter case, his understanding of 'swan' would be deviant. Suppose then that in the actual situation he would respond normally to the counterexample. Then there is reason to say that he understands the notion of swan correctly; and his error is not conceptual or linguistic, but empirical in an ordinary and narrow sense. (Of course, the line we are drawing here is pretty fuzzy.) When one comes to the counterfactual stage of the thought experiment, the subject has the same dispositions to respond pliably to the presentation of a black specimen. But such a response would suggest a misunderstanding of the term 'swan' as counterfactually used. For in the counterfactual community, what they call "swans" could not fail to be white. The mere presentation of a black swan would be irrelevant to the definitional truth 'All swans are white'. I have not set this case up as an example of the thought experiment's going through. Rather I have used it to support the conjecture that *if* the thought experiment is to work, one must at some stage find the subject believing (or having some attitude characterized by) a content, despite an incomplete understanding or misapplication. An ordinary empirical error appears not to be sufficient.

It would be a mistake, however, to think that incomplete understanding, in the sense that the argument requires, is in general an unusual or even deviant phenomenon. *What I have called "partial understanding" is common or even normal in the case of a large number of expressions in our vocabularies.* 'Arthritis' is a case in point. Even if by the grace of circumstance a person does not fall into views that

run counter to the term's meaning or application, it would not be in the least deviant or "socially unacceptable" to have no clear attitude that would block such views. 'Brisket', 'contract', 'recession', 'sonata', 'deer', 'elm' (to borrow a well-known example), 'pre-amplifier', 'carburetor', 'gothic', 'fermentation', probably provide analogous cases. Continuing the list is largely a matter of patience. The sort of "incomplete understanding" required by the thought experiment includes quite ordinary, nondeviant phenomena.

It is worth remarking that the thought experiment as originally presented might be run in reverse. The idea would be to start with an ordinary belief or thought involving no incomplete understanding. Then we find the incomplete understanding in the second step. For example, properly understanding 'arthritis', a patient may think (correctly) that he has arthritis. He happens to have heard of arthritis only occurring in joints, and he correctly believes that that is where arthritis always occurs. Holding his physical history, dispositions, and pain constant, we imagine that 'arthritis' commonly applies to rheumatoid ailments of all sorts. Arthritis has not been singled out for special mention. If the patient were told by a doctor 'You also have arthritis in the thigh', the patient would be disposed (as he is in the actual case) to respond, 'Really? I didn't know that one could have arthritis except in joints'. The doctor would answer, 'No, arthritis occurs in muscles, tendons, bursas, and elsewhere'. The patient would stand corrected. The notion that the doctor and patient would be operating with in such a case would not be that of arthritis.

My reasons for not having originally set out the thought experiment in this way are largely heuristic. As will be seen, discussion of the thought experiment will tend to center on the step involving incomplete understanding. And I wanted to encourage you, dear reader, to imagine actual cases of incomplete understanding in your own linguistic community. Ordinary intuitions in the domestic case are perhaps less subject to premature warping in the interests of theory. Cases involving not only mental content attribution, but also translation of a foreign tongue are more vulnerable to intrusion of side issues.

A secondary reason for not beginning with this "reversed" version of the thought experiment is that I find it doubtful whether the thought experiment always works in symmetric fashion. There may be special intuitive problems in certain cases—perhaps, for example, cases involving perceptual natural kinds. We may give special interpretations to individuals' misconceptions in imagined foreign communities, when those misconceptions seem to match our conceptions. In other words, there may be some systematic intuitive bias in favor of at least certain of our notions for purposes of interpreting the misconceptions of imagined foreigners. I do not want to explore the point here. I think that any such bias is not always crucial, and that the thought experiment frequently works "symmetrically." We have to take account of a person's community in interpreting his words and describing his attitudes—and this holds in the foreign case as well as in the domestic case.

The reversal of the thought experiment brings home the important point that *even those propositional attitudes not infected by incomplete understanding* depend for their content on social factors that are independent of the individual,

asocially and non-intentionally described. For if the social environment had been appropriately different, the contents of those attitudes would have been different.

Even *apart* from reversals of the thought experiment, it is plausible (in the light of its original versions) that our well-understood propositional attitudes depend partly for their content on social factors independent of the individual, asocially and non-intentionally construed. For each of us can reason as follows. Take a set of attitudes that involve a given notion and whose contents are well-understood by me. It is only contingent that I understand that notion as well as I do. Now holding my community's practices constant, imagine that I understand the given notion incompletely, but that the deficient understanding is such that it does not prevent my having attitude contents involving that notion. In fact, imagine that I am in the situation envisaged in the first step of one of the original thought experiments. In such a case, a proper subset of the original set of my actual attitude contents would, or might, remain the same—intuitively, at least those of my actual attitudes whose justification or point is untouched by my imagined deficient understanding. (In the arthritis case, an example would be a true belief that many old people have arthritis.) These attitude contents remain constant despite the fact that my understanding, inference patterns, behavior, dispositions, and so on would in important ways be different and partly inappropriate to applications of the given notion. What is it that enables these unaffected contents to remain applications of the relevant notion? It is not *just* that my understanding, inference patterns, behavior, and so forth are enough like my actual understanding, inference patterns, behavior, and so forth. For if communal practice had *also* varied so as to apply the relevant notion as I am imagining I misapply it, then my attitude contents would not involve the relevant notion at all. This argument suggests that communal practice is a factor (in addition to my understanding, inference patterns, and perhaps behavior, physical activity, and other features) in fixing the contents of my attitudes, even in cases where I fully understand the content.

IId. Independence from Factive-Verb and Indexical-Reference Paradigms

The thought experiment does not play on psychological "success" verbs or "factive" verbs—verbs like 'know', 'regret', 'realize', 'remember', 'foresee', 'perceive'. This point is important for our purposes because such verbs suggest an easy and clearcut distinction between the contribution of the individual subject and the objective, "veridical" contribution of the environment to making the verbs applicable. (Actually the matter becomes more complicated on reflection, but we shall stay with the simplest cases.) When a person knows that snow is common in Greenland, his knowledge obviously depends on more than the way the person is. It depends on there actually being a lot of snow in Greenland. His mental state (belief that snow is common in Greenland) must be successful in a certain way (true). By changing the environment, one could change the truth value of the content, so that the subject could no longer be said to know the content. It is part of the burden of our argument that even intentional mental states of the individual like beliefs,

which carry no implication of veridicality or success, cannot be understood by focusing purely on the individual's acts, dispositions, and "inner" goings on.

The thought experiment also does not rest on the phenomenon of indexicality, or on *de re* attitudes, in any direct way. When Alfred refers to an apple, saying to himself "That is wholesome," what he refers to depends not just on the content of what he says or thinks, but on what apple is before him. Without altering the meaning of Alfred's utterance, the nature of his perceptual experiences, or his physical acts or dispositions, we could conceive an exchange of the actual apple for another one that is indistinguishable to Alfred. We would thereby conceive him as referring to something different and even as saying something with a different truth value.

This rather obvious point about indexicality has come to be seen as providing a model for understanding a certain range of mental states or events—*de re* attitudes. The precise characterization of this range is no simple philosophical task. But the clearest cases involve non-obliquely occurring terms in content clauses. When we say that Bertrand thinks of some water that it would not slake his thirst (where 'water' occurs in purely non-oblique position), we attribute a *de re* belief to Bertrand. We assume that Bertrand has something like an indexical relation to the water. The fact that Bertrand believes something of some water, rather than of a portion of some other liquid that is indistinguishable to him, depends partly on the fact that it is water to which Bertrand is contextually, "indexically" related. For intuitively we could have exchanged the liquids without changing Bertrand and thereby changed what Bertrand believed his belief content *of*—and even whether his belief was true of it.[3] It is easy to interpret such cases by holding that the subject's mental states and contents (with allowances for brute differences in the contexts in which he applies those contents) remain the same. The differences in the situations do not pertain in any fundamental way to the subject's mind or the nature of his mental content, but to how his mind or content is related to the world.

I think this interpretation of standard indexical and *de re* cases is broadly correct, although it involves oversimplifications and demands refinements. But what I want to emphasize here is that it is inapplicable to the cases our thought experiment fixes upon.

It seems to me clear that the thought experiment need not rely on *de re* attitudes at all. The subject need not have entered into special *en rapport* or quasi-indexical relations with objects that the misunderstood term applies to in order for the argument to work. We can appeal to attitudes that would usually be regarded as paradigmatic cases of *de dicto,* non-indexical, *non-de-re,* mental attitudes or events. The primary mistake in the contract example is one such, but we could choose others to suit the reader's taste. To insist that such attitudes must all be indexically infected or *de re* would, I think, be to trivialize and emasculate these notions, making nearly all attitudes *de re.* All *de dicto* attitudes presuppose *de re* attitudes. But it does not follow that indexical or *de re* elements survive in every attitude. (Cf. notes 2 and 3.)

I shall not, however, argue this point here. The claim that is crucial is not that our argument does not fix on *de re* attitudes. It is, rather, that the social differences

between the actual and counterfactual situations affect the *content* of the subject's attitudes. That is, the difference affects standard cases of obliquely occurring, cognitive-content-conveying expressions in content clauses. For example, still with his misunderstanding, the subject might think that this (referring to his disease in his hands) is arthritis. Or he might think *de re* of the disease in his ankle (or of the disease in his thigh) that his arthritis is painful. It does not really matter whether the relevant attitude is *de re* or purely *de dicto*. What is crucial to our argument is that the occurrence of 'arthritis' is oblique and contributes to a characterization of the subject's mental content. One might even hold, implausibly I think, that all the subject's attitudes involving the notion of arthritis are *de re*, that 'arthritis' in that-clauses *indexically* picks out the property of being arthritis, or something like that. The fact remains that the term occurs obliquely in the relevant cases and serves in characterizing the *dicta* or contents of the subject's attitudes. The thought experiment exploits this fact.

Approaches to the mental that I shall later criticize as excessively individualistic tend to assimilate environmental aspects of mental phenomena to either the factive-verb or indexical-reference paradigm. (Cf. note 2.) This sort of assimilation suggests that one might maintain a relatively clearcut distinction between extramental and mental aspects of mentalistic attributions. And it may encourage the idea that the distinctively mental aspects can be understood fundamentally in terms of the individual's abilities, dispositions, states, and so forth, considered in isolation from his social surroundings. Our argument undermines this latter suggestion. Social context infects even the distinctively mental features of mentalistic attributions. No man's intentional mental phenomena are insular. Every man is a piece of the social continent, a part of the social main.

III. REINTERPRETATIONS

IIIa. Methodology

I find that most people unspoiled by conventional philosophical training regard the three steps of the thought experiment as painfully obvious. Such folk tend to chafe over my filling in details or elaborating on strategy. I think this naivete appropriate. But for sophisticates the three steps require defense.

Before launching a defense, I want to make a few remarks about its methodology. My objective is to better understand our common mentalistic notions. Although such notions are subject to revision and refinement, I take it as evident that there is philosophical interest in theorizing about them as they now are. I assume that a primary way of achieving theoretical understanding is to concentrate on our *discourse* about mentalistic notions. Now it is, of course, never obvious at the outset how much idealization, regimentation, or special interpretation is necessary in order to adequately understand ordinary discourse. Phenomena such as ambiguity, ellipsis, indexicality, idioms, and a host of others certainly demand some regimentation or special interpretation for purposes of linguistic theory. Moreover, more global consid-

erations—such as simplicity in accounting for structural relations—often have effects on the cast of one's theory. For all that, there is a methodological bias in favor of taking natural discourse literally, other things being equal. For example, unless there are clear reasons for construing discourse as ambiguous, elliptical or involving special idioms, we should not so construe it. Literal interpretation is *ceteris paribus* preferred. My defense of the thought experiment, as I have interpreted it, partly rests on this principle.

This relatively non-theoretical interpretation of the thought experiment should be extended to the gloss on it that I provided in Section IIc. The notions of misconception, incomplete understanding, conceptual or linguistic error, and ordinary empirical error are to be taken as carrying little theoretical weight. I assume that these notions mark defensible, common-sense distinctions. But I need not take a position on available philosophical interpretations of these distinctions. In fact, I do not believe that understanding, in our examples, can be explicated as independent of empirical knowledge, or that the conceptual errors of our subjects are best seen as "purely" mistakes about concepts and as involving no "admixture" of error about "the world." With Quine, I find such talk about purity and mixture devoid of illumination or explanatory power. But my views on this matter neither entail nor are entailed by the premises of the arguments I give (cf. e.g., IIId). Those arguments seem to me to remain plausible under any of the relevant philosophical interpretations of the conceptual-ordinary-empirical distinction.

I have presented the experiment as appealing to ordinary intuition. I believe that common practice in the attribution of propositional attitudes is fairly represented by the various steps. This point is not really open to dispute. Usage may be divided in a few of the cases in which I have seen it as united. But broadly speaking, it seems to me undeniable that the individual steps of the thought experiment are acceptable to ordinary speakers in a wide varity of examples. The issue open to possible dispute is whether the steps should be taken in the literal way in which I have taken them, and thus whether the conclusion I have drawn from those steps is justified. In the remainder of Section III, I shall try to vindicate the literal interpretation of our examples. I do this by criticizing, in order of increasing generality or abstractness, a series of attempts to reinterpret the thought experiment's first step. Ultimately, I suggest (IIId and IV) that these attempts derive from characteristically philosophical models that have little or no independent justification. A thoroughgoing review of these models would be out of bounds, but the present paper is intended to show that they are deficient as accounts of our actual practice of mentalistic attribution.

I shall have little further to say in defense of the second and third steps of the thought experiment. Both rest on their intuitive plausibility, not on some particular theory. The third step, for example, certainly does not depend on a view that contents are merely sentences the subject is disposed to utter, interpreted as his community interprets them. It is compatible with several philosophical accounts of mental contents, including those that appeal to more abstract entities such as Fregean thoughts or Russellian propositions, and those that seek to deny that content-clauses indicate any *thing* that might be called a content. I also do not claim that the fact that our subject lacks the relevant beliefs in the third step follows

from the facts I have described. The point is that it is plausible, and certainly possible, that he would lack those beliefs.

The exact interpretation of the second step is relevant to a number of causal or functional theories of mental phenomena that I shall discuss in Section IV. The intuitive idea of the step is that none of the different physical, non-intentionally described causal chains set going by the differences in communal practice need affect our subjects in any way that would be relevant to an account of their mental contents. Differences in the behavior of other members of the community will, to be sure, affect the gravitational forces exerted on the subject. But I assume that these differences are irrelevant to macro-explanations of our subjects' physical movements and inner processes. They do not relevantly affect ordinary non-intentional physical explanations of how the subject acquires or is disposed to use the symbols in his repertoire. Of course, the social origins of a person's symbols do differ between actual and counterfactual cases. I shall return to this point in Sections IV and V. The remainder of Section III will be devoted to the first step of the thought experiment.

IIIb. Incomplete Understanding and Standard Cases of Reinterpretation

The first step, as I have interpreted it, is the most likely to encounter opposition. In fact, there is a line of resistance that is second nature to linguistically oriented philosophers. According to this line, we should deny that, say, the patient really believed or thought that arthritis can occur outside of joints because he misunderstood the word 'arthritis'. More generally, we should deny that a subject could have any attitudes whose contents he incompletely understands.

What a person understands is indeed one of the chief factors that bear on what thoughts he can express in using words. If there were not deep and important connections between propositional attitudes and understanding, one could hardly expect one's attributions of mental content to facilitate reliable predictions of what a person will do, say, or think. But our examples provide reason to believe that these connections are not simple entailments to the effect that having a propositional attitude strictly implies full understanding of its content.

There are, of course, numerous situations in which we normally reinterpret or discount a person's words in deciding what he thinks. Philosophers often invoke such cases to bolster their animus against such attributions as the ones we made to our subjects: "If a foreigner were to mouth the words 'arthritis may occur in the thigh' or 'my father had arthritis', not understanding what he uttered in the slightest, we would not say that he believed that arthritis may occur in the thigh, or that his father had arthritis. So why should we impute the belief to the patient?" Why, indeed? Or rather, why do we?

The question is a good one. We do want a general account of these cases. But the implied argument against our attribution is anemic. We tacitly and routinely distinguish between the cases I described and those in which a foreigner (or anyone) utters something without any comprehension. The best way to understand mental-

istic notions is to recognize such differences in standard practice and try to account for them. One can hardly justify the assumption that full understanding of a content is in general a necessary condition for believing the content by appealing to some cases that tend to support the assumption in order to reject others that conflict with it.

It is a good method of discovery, I think, to note the sorts of cases philosophers tend to gravitate toward when they defend the view that the first step in the thought experiment should receive special interpretation. By reflecting on the differences between these cases and the cases we have cited, one should learn something about principles controlling mentalistic attribution.

I have already mentioned foreigners without command of the language. A child's imitation of our words and early attempts to use them provide similar examples. In these cases, mastery of the language and responsibility to its precepts have not been developed; and mental content attribution based on the meaning of words uttered tends to be precluded.

There are cases involving regional dialects. A person's deviance or ignorance judged by the standards of the larger community may count as normality or full mastery when evaluated from the regional perspective. Clearly, the regional standards tend to be the relevant ones for attributing content when the speaker's training or intentions are regionally oriented. The conditions for such orientation are complex, and I shall touch on them again in Section V. But there is no warrant in actual practice for treating each person's idiolect as always analogous to dialects whose words we automatically reinterpret—for purposes of mental content attribution—when usage is different. People are frequently held, and hold themselves, to the standards of their community when misuse or misunderstanding are at issue. One should distinguish these cases, which seem to depend on a certain *responsibility* to communal practice, from cases of automatic reinterpretation.

Tongue slips and Spoonerisms form another class of example where reinterpretation of a person's words is common and appropriate in arriving at an attribution of mental content. In these cases, we tend to exempt the speaker even from commitment to a homophonically formulated assertion content, as well as to the relevant mental content. The speaker's own behavior usually follows this line, often correcting himself when what he uttered is repeated back to him.

Malapropisms form a more complex class of examples. I shall not try to map it in detail. But in a fairly broad range of cases, we reinterpret a person's words at least in attributing mental content. If Archie says, 'Lead the way and we will precede', we routinely reinterpret the words in describing his expectations. Many of these cases seem to depend on the presumption that there are simple, superficial (for example, phonological) interference or exchange mechanisms that account for the linguistic deviance.

There are also examples of quite radical misunderstandings that sometimes generate reinterpretation. If a generally competent and reasonable speaker thinks that 'orangutan' applies to a fruit drink, we would be reluctant, and it would unquestionably be misleading, to take his words as revealing that he thinks he has

been drinking orangutans for breakfast for the last few weeks. Such total misunderstanding often *seems* to block literalistic mental content attribution, at least in cases where we are not directly characterizing his mistake. (Contrary to philosophical lore, I am not convinced that such a man cannot correctly and literally be attributed a belief that an orangutan is a kind of fruit drink. But I shall not deal with the point here.)

There are also some cases that do not seem generally to prevent mental content attribution on the basis of literal interpretation of the subject's words in quite the same way as the others, but which deserve some mention. For almost any content except for those that directly display the subject's incomplete understanding, there will be many contexts in which it would be misleading to attribute that content to the subject without further comment. Suppose I am advising you about your legal liabilities in a situation where you have entered into what may be an unwritten contract. You ask me what Al would think. It would be misleading for me to reply that Al would think that you do not have a contract (or even do not have any legal problems), if I know that Al thinks a contract must be based on a formal document. Your evaluation of Al's thought would be crucially affected by his inadequate understanding. In such cases, it is incumbent on us to cite the subject's eccentricity: "(He would think that you do not have a contract, but then) he thinks that there is no such thing as a verbally based contract."

Incidentally, the same sort of example can be constructed using attitudes that are abnormal, but that do not hinge on misunderstanding of any one notion. If Al had thought that only traffic laws and laws against violent crimes are ever prosecuted, it would be misleading for me to tell you that Al would think that you have no legal problems.

Both sorts of cases illustrate that in reporting a single attitude content, we typically suggest (implicate, perhaps) that the subject has a range of other attitudes that are normally associated with it. Some of these may provide reasons for it. In both sorts of cases, it is usually important to keep track of, and often to make explicit, the nature and extent of the subject's deviance. Otherwise, predictions and evaluations of his thought and action, based on normal background assumptions, will go awry. When the deviance is huge, attributions demand reinterpretation of the subject's words. Radical misunderstanding and mental instability are cases in point. But frequently, common practice seems to allow us to cancel the misleading suggestions by making explicit the subject's deviance, retaining literal interpretation of his words in our mentalistic attributions all the while.

All of the foregoing phenomena are relevant to accounting for standard practice. But they are no more salient than cases of straightforward belief attribution where the subject incompletely understands some notion in the attributed belief content. I think any impulse to say that common practice is *simply* inconsistent should be resisted (indeed, scorned). We cannot expect such practice to follow general principles rigorously. But even our brief discussion of the matter should have suggested the beginnings of generalizations about differences between cases where reinterpretation is standard and cases where it is not. A person's overall linguistic

competence, his allegiance and responsibility to communal standards, the degree, source, and type of misunderstanding, the purposes of the report—all affect the issue. From a theoretical point of view, it would be a mistake to try to assimilate the cases in one direction or another. We do not want to credit a two-year-old who memorizes '$e = mc^2$' with belief in relativity theory. But the patient's attitudes involving the notion of arthritis should not be assimilated to the foreigner's uncomprehending pronunciations.

For purposes of defending the thought experiment and the arguments I draw from it, I can afford to be flexible about exactly how to generalize about these various phenomena. The thought experiment depends only on there being some cases in which a person's incomplete understanding does not force reinterpretation of his expressions in describing his mental contents. Such cases appear to be legion.

IIIc. Four Methods of Reinterpreting the Thought Experiment

I now want to criticize attempts to argue that even in cases where we ordinarily do ascribe content clauses despite the subject's incomplete understanding of expressions in those clauses, such ascriptions should not be taken literally. In order to overturn our interpretation of the thought experiment's first step, one must argue that none of the cases I have cited is appropriately taken in the literal manner. One must handle (apparent) attributions of unproblematically true contents involving incompletely mastered notions, as well as attributions of contents that display the misconceptions or partial understandings. I do not doubt that one can erect logically coherent and metaphysically traditional reinterpretations of all these cases. What I doubt is that such reinterpretations taken *in toto* can present a plausible view, and that taken individually they have any claim to superiority over the literal interpretations—either as accounts of the language of ordinary mentalistic ascription, or as accounts of the evidence on which mental attributions are commonly based.

Four types of reinterpretation have some currency. I shall be rather short with the first two, the first of which I have already warned against in Section IId. Sometimes relevant mentalistic ascriptions are reinterpreted as attributions of *de re* attitudes *of* entities not denoted by the misconstrued expressions. For example, the subject's belief that he has arthritis in the thigh might be interpreted as a belief *of* the non-arthritic rheumatoid ailment that it is in the thigh. The subject will probably have such a belief in this case. But it hardly accounts for the relevant attributions. In particular, it ignores the oblique occurrence of 'arthritis' in the original ascription. Such occurrences bear on a characterization of the subject's viewpoint. The subject thinks of the disease in his thigh (and of his arthritis) in a certain way. He thinks of each disease that it is arthritis. Other terms for arthritis (or for the actual trouble in his thigh) may not enable us to describe his attitude content nearly as well. The appeal to *de re* attitudes in this way is not adequate to the task of reinterpreting these ascriptions so as to explain away the difference between actual and counterfactual situations. It simply overlooks what needs explication.

A second method of reinterpretation, which Descartes proposed (cf. Section IV) and which crops up occasionally, is to claim that in cases of incomplete understanding, the subject's attitude or content is indefinite. It is surely true that in cases where a person is extremely confused, we are sometimes at a loss in describing his attitudes. Perhaps in such cases, the subject's mental content *is* indefinite. But in the cases I have cited, common practice lends virtually no support to the contention that the subject's mental contents are indefinite. The subject and his fellows typically know and agree on precisely *how to confirm or infirm* his beliefs—both in the cases where they are unproblematically true (or just empirically false) and in the cases where they display the misconception. Ordinary attributions typically specify the mental content without qualifications or hesitations.

In cases of partial understanding—say, in the mortgage example—it may indeed be unclear, short of extensive questioning, just how much mastery the subject has. But even this sort of unclarity does not appear to prevent, under ordinary circumstances, straightforward attributions utilizing 'mortgage' in oblique position. The subject is uncertain whether his father has two mortgages; he knows that his uncle has paid off the mortgage on his house. The contents are unhesitatingly attributed and admit of unproblematic testing for truth value, despite the subject's partial understanding. There is thus little *prima facie* ground for the appeal to indefiniteness. The appeal appears to derive from a prior assumption that attribution of a content entails attribution of full understanding. Lacking an easy means of attributing something other than the misunderstood content, one is tempted to say that there *is* no definite content. But this is unnecessarily mysterious. It reflects on the prior assumption, which so far has no independent support.

The other two methods of reinterpretation are often invoked in tandem. One is to attribute a notion that just captures the misconception, thus replacing contents that are apparently false on account of the misconception, by true contents. For example, the subject's belief (true or false) that that is a sofa would be replaced by, or reinterpreted as, a (true) belief that that is a *chofa*, where 'chofa' is introduced to apply not only to sofas, but also to the armchairs the subject thinks are sofas. The other method is to count the error of the subject as purely metalinguistic. Thus the patient's apparent belief that he had arthritis in the thigh would be reinterpreted as a belief that 'arthritis' applied to something (or some disease) in his thigh. The two methods can be applied simultaneously, attempting to account for an ordinary content attribution in terms of a reinterpreted object-level content together with a metalinguistic error. It is important to remember that in order to overturn the thought experiment, these methods must not only establish that the subject held the particular attitudes that they advocate attributing; they must also justify a *denial* of the ordinary attributions literally interpreted.

The method of invoking object-level notions that precisely capture (and that replace) the subject's apparent misconception has little to be said for it as a natural and generally applicable account of the language of mentalistic ascriptions. We do not ordinarily seek out true object-level attitude contents to attribute to victims of

errors based on incomplete understanding. For example, when we find that a person has been involved in a misconception in examples like ours, we do not regularly reinterpret those ascriptions that involved the misunderstood term, but were untuitively unaffected by the error. An attribution to someone of a true belief that he is eating brisket, or that he has just signed a contract, or that Uncle Harry has paid off his mortgage, is not typically reformulated when it is learned that the subject had not fully understood what brisket (or a contract, or a mortgage) is. A similar point applies when we know about the error at the time of the attribution—at least if we avoid misleading the audience in cases where the error is crucial to the issue at hand. Moreover, we shall frequently see the subject as sharing beliefs with others who understand the relevant notions better. In counting beliefs as shared, we do not require, in every case, that the subjects "fully understand" the notions in those belief contents, or understand them in just the same way. Differences in understanding are frequently located as differences over other belief contents. We agree that you have signed a contract, but disagree over whether someone else could have made a contract by means of a verbal agreement.

There are reasons why ordinary practice does not follow the method of object-level reinterpretation. In many cases, particularly those involving partial understanding, finding a reinterpretation in accord with the method would be entirely nontrivial. It is not even clear that we have agreed upon means of pursuing such inquiries in all cases. Consider the arthritic patient. Suppose we are to reinterpret the attribution of his erroneous belief that he has arthritis in the thigh. We make up a term 'tharthritis' that covers arthritis and whatever it is he has in his thigh. The appropriate restrictions on the application of this term and of the patient's supposed notion are unclear. Is just any problem in the thigh that the patient wants to call 'arthritis' to count as tharthritis? Are other ailments covered? What would decide? The problem is that there are no recognized standards governing the application of the new term. In such cases, the method is patently *ad hoc*.

The method's willingness to invoke new terminology whenever conceptual error or partial understanding occurs is *ad hoc* in another sense. It proliferates terminology without evident theoretical reward. We do not engender better understanding of the patient by inventing a new word and saying that he thought (correctly) that tharthritis can occur outside joints. It is simpler and equally informative to construe him as thinking that arthritis may occur outside joints. When we are making other attributions that do not directly display the error, we must simply bear the deviant belief in mind, so as not to assume that all of the patient's inferences involving the notion would be normal.

The method of object-level reinterpretation often fails to give a plausible account of the evidence on which we base mental attributions. When caught in the sorts of errors we have been discussing, the subject does not normally respond by saying that his views had been misunderstood. The patient does not say (or think) that he had thought he had some-category-of-disease-like-arthritis-and-including-arthritis-but-also-capable-of-occurring-outside-of-joints in the thigh *instead* of the error commonly attributed. This sort of response would be disingenuous. Whatever

other beliefs he had, the subject thought that he had arthritis in the thigh. In such cases, the subject will ordinarily give no evidence of having maintained a true object-level belief. In examples like ours, he typically admits his mistake, changes his views, and leaves it at that. Thus the subject's own behavioral dispositions and inferences often fail to support the method.

The method may be seen to be implausible as an account of the relevant evidence in another way. The patient knows that he has had arthritis in the ankle and wrists for some time. Now with his new pains in the thigh, he fears and believes that he has got arthritis in the thigh, that his arthritis is spreading. Suppose we reinterpret all of these attitude attributions in accord with the method. We use our recently coined term 'tharthritis' to cover (somehow) arthritis and whatever it is he has in the thigh. On this new interpretation, the patient is right in thinking that he has tharthritis in the ankle and wrists. His belief that it has lodged in the thigh is true. His fear is realized. But these attributions are out of keeping with the way we do and should view his actual beliefs and fears. His belief is not true, and his fear is not realized. He will be relieved when he is told that one cannot have arthritis in the thigh. His relief is bound up with a network of assumptions that he makes about his arthritis: that it is a kind of disease, that there are debilitating consequences of its occurring in multiple locations, and so on. When told that arthritis cannot occur in the thigh, the patient does not decide that his fears were realized, but that perhaps he should not have had those fears. He does not think: Well, my tharthritis *has* lodged in the thigh; but judging from the fact that what the doctor called "arthritis" cannot occur in the thigh, tharthritis may not be a single kind of disease; and I suppose I need not worry about the effects of its occurring in various locations, since evidently the tharthritis in my thigh is physiologically unrelated to the tharthritis in my joints. There will rarely if ever be an empirical basis for such a description of the subject's inferences. The patient's behavior (including his reports, or thinkings-out-loud) in this sort of case will normally not indicate any such pattern of inferences at all. But this is the description that the object-level reinterpretation method appears to recommend.

On the standard attributions, the patient retains his assumptions about the relation between arthritis, kinds of disease, spreading, and so on. And he concludes that his arthritis is not appearing in new locations—at any rate, not in his thigh. These attributions will typically be supported by the subject's behavior. The object-level reinterpretation method postulates inferences that are more complicated and different in focus from the inferences that the evidence supports. The method's presentation in such a case would seem to be an *ad hoc* fiction, not a description with objective validity.

None of the foregoing is meant to deny that frequently when a person incompletely understands an attitude content he has some other attitude content that more or less captures his understanding. For example, in the contract example, the client will probably have the belief that if one breaks *a legally binding agreement based on formal documents,* then one may get into trouble. There are also cases in which it is reasonable to say that, at least in a sense, a person has a notion that is

expressed by his dispositions to classify things in a certain way—even if there is no conventional term in the person's repertoire that neatly corresponds to that "way." The sofa case may be one such. Certain animals as well as people may have nonverbal notions of this sort. On the other hand, the fact that such attributions are justifiable *per se* yields no reason to deny that the subject (also) has object-level attitudes whose contents involve the relevant incompletely understood notion.

Whereas the third method purports to account for the subject's thinking at the object level, the fourth aims at accounting for his error. The error is construed as purely a metalinguistic mistake. The relevant false content is seen to involve notions that denote or apply to linguistic expressions. In examples relevant to our thought experiment, we ordinarily attribute a metalinguistic as well as an object-level attitude to the subject, at least in the case of non-occurrent propositional attitudes. For example, the patient probably believes that 'arthritis' applies in English to the ailment in his thigh. He believes that his father had a disease called "arthritis." And so on. Accepting these metalinguistic attributions, of course, does nothing *per se* toward making plausible a denial that the subjects in our examples have the counterpart object-level attitudes.

Like the third method, the metalinguistic reinterpretation method has no *prima facie* support as an account of the language of mentalistic ascriptions. When we encounter the subject's incomplete understanding in examples like ours, we do not decide that all the mental contents which we had been attributing to him with the misunderstood notion must have been purely metalinguistic in form. We also count people who incompletely understand terms in ascribed content clauses as sharing true and unproblematic object-level attitudes with others who understand the relevant terms better. For example, the lawyer and his client may share a wish that the client had not signed the contract to buy the house without reading the small print. A claim that these people share *only* attitudes with metalinguistic contents would have no support in linguistic practice.

The point about shared attitudes goes further. If the metalinguistic reinterpretation account is to be believed, we cannot say that a relevant English speaker shares a view (for example) that many old people have arthritis, with *anyone* who does not use the English word 'arthritis'. For the foreigner does not have the word 'arthritis' to hold beliefs about, though he does have attitudes involving the notion arthritis. And the attribution to the English speaker is to be interpreted metalinguistically, making reference to the word, so as not to involve attribution of the notion arthritis. This result is highly implausible. Ascriptions of such that-clauses as the above, regardless of the subject's language, serve to provide single descriptions and explanations of similar patterns of behavior, inference, and communication. To hold that we cannot accurately ascribe single content-clauses to English speakers and foreigners in such cases would not only accord badly with linguistic practice. It would substantially weaken the descriptive and explanatory power of our common attributions. In countless cases, unifying accounts of linguistically disparate but cognitively and behaviorally similar phenomena would be sacrificed.

The method is implausible in other cases as an account of standard evidence on which mental attributions are based. Take the patient who fears that his arthritis is spreading. According to the metalinguistic reinterpretation method, the patient's reasoning should be described as follows. He thinks that the word 'arthritis' applies to a single disease in him, that the disease in him called "arthritis" is debilitating if it spreads, that 'arthritis' applies to the disease in his wrists and ankles. He fears that the disease called "arthritis" has lodged in his thigh, and so on. Of course, it is often difficult to find evidential grounds for attributing an object-level attitude *as opposed* to its metalinguistic counterpart. As I noted, when a person holds one attitude, he often holds the other. But there are types of evidence, in certain contexts, for making such discriminations, particularly contexts in which *occurrent* mental events are at issue. The subject may maintain that his reasoning did not fix upon words. He may be brought up short by a metalinguistic formulation of his just-completed ruminations, and may insist that he was not interested in labels. In such cases, especially if the reasoning is not concerned with linguistic issues in any informal or antecedently plausible sense, attribution of an object-level thought content is supported by the relevant evidence, and metalinguistic attribution is not. To insist that the occurrent mental event really involved a metalinguistic content would be a piece of *ad hoc* special pleading, undermined by the evidence we actually use for deciding whether a thought was metalinguistic.

In fact, there appears to be a general presumption that a person is reasoning at the object level, other things being equal. The basis for this presumption is that metalinguistic reasoning requires a certain self-consciousness about one's words and social institutions. This sort of sophistication emerged rather late in human history. (Cf. any history of linguistics.) Semantical notions were a product of this sophistication.

Occurrent propositional attitudes prevent the overall reinterpretation strategy from providing a plausible total account which would block our thought experiment. For such occurrent mental events as the patient's thought that his arthritis is especially painful in the knee this morning are, or can be imagined to be, clear cases of object-level attitudes. And such thoughts may enter into or connect up with pieces of reasoning—say the reasoning leading to relief that the arthritis had not lodged in the thigh—which cannot be plausibly accounted for in terms of object-level reinterpretation. The other reinterpretation methods (those that appeal to *de re* contents and to indefiniteness) are non-starters. In such examples, the literally interpreted ascriptions appear to be straightforwardly superior accounts of the evidence that is normally construed to be relevant. Here one need not appeal to the principle that literal interpretation is, other things equal, preferable to reinterpretation. Other things are not equal.

At this point, certain philosophers may be disposed to point out that what a person says and how he behaves do not infallibly determine what his attitude contents are. Despite the apparent evidence, the subject's attitude contents may in all cases I cited be metalinguistic, and may fail to involve the incompletely understood

notion. It is certainly true that how a person acts and what he says, even sincerely, do not determine his mental contents. I myself have mentioned a number of cases that support the point. (Cf. IIIb.) But the point is often used in a sloppy and irresponsible manner. It is incumbent on someone making it (and applying it to cases like ours) to indicate considerations that override the linguistic and behavioral evidence. In Section IIId, I shall consider intuitive or *a priori* philosophical arguments to this end. But first I wish to complete our evaluation of the metalinguistic reinterpretation method as an account of the language of mentalistic ascription in our examples.

In this century philosophers have developed the habit of insisting on metalinguistic reinterpretation for any content attribution that directly *displays* the subject's incomplete understanding. These cases constitute but a small number of the attributions that serve the thought experiment. One could grant these reinterpretations and still maintain our overall viewpoint. But even as applied to these cases, the method seems dubious. I doubt that any evidentially supported account of the language of these attributions will show them in general to be attributions of metalinguistic contents—contents that involve denotative reference to linguistic expressions.

The ascription 'He believes that broad overstuffed armchairs are sofas', as ordinarily used, does not in general *mean* "He believes that broad, overstuffed armchairs are covered by the expression 'sofas' " (or something like that). There are clear grammatical and semantical differences between

(i) broad, overstuffed armchairs are covered by the expression 'sofas'

and

(ii) broad, overstuffed armchairs are sofas.

When the two are embedded in belief contexts, they produce grammatically and semantically distinct sentences.

As noted, ordinary usage approves ascriptions like

(iii) He believes that broad, overstuffed armchairs are sofas.

It would be wildly *ad hoc* and incredible from the point of view of linguistic theory to claim that there is *no* reading of (iii) that embeds (ii). But there is no evidence from speaker behavior that *true* ascriptions of (iii) always (or perhaps even *ever*) derive from embedding (i) rather than (ii). In fact, I know of no clear evidence that (iii) is ambiguous between embedding (i) and (ii), or that (ii) is ambiguous, with one reading identical to that of (i). People do not in general seem to regard ascriptions like (iii) as elliptical. More important, in most cases no amount of nonphilosophical badgering will lead them to withdraw (iii), under some interpretation, *in favor of* an ascription that clearly embeds (i). At least in the cases of *non-occurrent* propositional attitudes, they will tend to agree to a clearly metalinguistic ascription—a belief sentence explicitly embedding something like (i)—in cases where they make an ascription like (iii). But this is evidence that they regard ascriptions that embed (i) and (ii) as both true. It hardly tells against counting belief ascriptions that embed

(ii) as true, or against taking (iii) in the obvious, literal manner. In sum, there appears to be no ordinary empirical pressure on a theory of natural language to represent true ascriptions like (iii) as *not* embedding sentences like (ii). And other things being equal, literal readings are correct readings. Thus it is strongly plausible to assume that ordinary usage routinely accepts as true and justified even ascriptions like (iii), literally interpreted as embedding sentences like (ii).

There are various contexts in which we may be indifferent over whether to attribute a metalinguistic attitude or the corresponding object-level attitude. I have emphasized that frequently, though not always, we may attribute both. Or we might count the different contents as describing what contextually "amount to the same attitude." (Cf. Section I.) Even this latter locution remains compatible with the thought experiment, as long as both contents are *equally attributable* in describing "the attitude." In the counterfactual step of the thought experiment, the metalinguistic content (say, that broad, overstuffed armchairs are called "sofas") will still be attributable. But in these circumstances it contextually "amounts to the same attitude" as an object-level attitude whose content is in no sense equivalent to, or "the same as," the original object-level content. For they have different truth values. Thus, assuming that the object-level and metalinguistic contents are equally attributable, it remains informally plausible that the person's attitudes are different between actual and counterfactual steps in the thought experiment. This contextual conflation of object-level and metalinguistic contents is not, however, generally acceptable even in describing non-occurrent attitudes, much less occurrent ones. There are contexts in which the subject himself may give evidence of making the distinction.

IIId. Philosophical Arguments for Reinterpretation

I have so far argued that the reinterpretation strategies that I have cited do not provide a plausible account of evidence relevant to a theory of the language of mentalistic ascriptions or to descriptions of mental phenomena themselves. I now want to consider characteristically philosophical arguments for revising ordinary discourse or for giving it a nonliteral reading, arguments that rely purely on intuitive or *a priori* considerations. I have encountered three such arguments, or argument sketches.[4]

One holds that the content clauses we ascribed must be reinterpreted so as to make reference to words because they clearly concern linguistic matters—or are about language. Even if this argument were sound, it would not affect the thought experiment decisively. For most of the mental contents that vary between actual and counterfactual situations are not in any intuitive sense "linguistic." The belief that certain armchairs are sofas is intuitively linguistic. But beliefs that some sofas are beige, that Kirkpatrick is playing a clavichord, and that Milton had severe arthritis in his hands are not.

But the argument is unpersuasive even as applied to the contents that, in an intuitive sense, do concern linguistic matters. A belief that broad, overstuffed armchairs are sofas is linguistic (or "about" language) in the same senses as an "analyti-

cally" true belief that no armchairs are sofas. But the linguistic nature of the latter belief does not make its logical form metalinguistic. So citing the linguistic nature of the former belief does not suffice to show it metalinguistic. No semantically relevant component of either content applies to or denotes linguistic expressions.

Both the "analytically" true and the "analytically" false attitudes are linguistic in the sense that they are tested by consulting a dictionary or native linguistic intuitions, rather than by ordinary empirical investigation. We do not scrutinize pieces of furniture to test these beliefs. The pragmatic focus of expressions of these attitudes will be on usage, concepts, or meaning. But it is simply a mistake to think that these facts entail, or even suggest, that the relevant contents are metalinguistic in form. Many contents with object-level logical forms have primarily linguistic or conceptual implications.

A second argument holds that charitable interpretation requires that we not attribute to rational people beliefs like the belief that one may have arthritis in the thigh. Here again, the argument obviously does not touch most of the attitudes that may launch the thought experiment; for many are straightforwardly true, or false on ordinary empirical grounds. Even so, it is not a good argument. There is nothing irrational or stupid about the linguistic or conceptual errors we attribute to our subjects. The errors are perfectly understandable as results of linguistic misinformation.

In fact, the argument makes sense only against the background of the very assumption that I have been questioning. A belief that arthritis may occur in the thigh appears to be inexplicable or uncharitably attributed only if it is assumed that the subject must fully understand the notions in his attitude contents.

A third intuitive or *a priori* argument is perhaps the most interesting. Sometimes it is insisted that we should not attribute contents involving incompletely understood notions because *the individual must mean something different by the misunderstood word than what we non-deviant speakers mean by it*. Note again that it would not be enough to use this argument from deviant speaker meaning to show that the subject has notions that are not properly expressed in the way he thinks they are. In some sense of 'expressed', this is surely often the case. To be relevant, the argument must arrive at a negative conclusion: that the subject cannot have the attitudes that seem commonly to be attributed.

The expression 'the individual meant something different by his words' can be interpreted in more than one way. On one group of interpretations, the expression says little more than that the speaker incompletely understood his words: The patient thought 'arthritis' meant something that included diseases that occur outside of joints. The client would have misexplained the meaning, use, or application of 'contract'. The subject applied 'sofa' to things that, unknown to him, are not sofas. A second group of interpretations emphasizes that not only does the speaker misconstrue or misapply his words, but he had *in mind* something that the words do not denote or express. The subject sometimes had in mind certain armchairs when he used 'sofa.' The client regarded the notion of legal agreement based on written documents as approximately interchangeable with what is expressed by 'contract', and thus had such a notion in mind when he used 'contract'. A person

with a problem about the range of red might sometimes have in mind a mental image of a non-red color when he used 'red'.

The italicized premise of the argument is, of course, always true in our examples under the first group of interpretations, and often true under the second. But interpreted in these ways, the argument is a *non sequitur*. It does not follow from the assumption that the subject thought that a word means something that it does not (or misapplies the word, or is disposed to misexplain its meaning) that the word cannot be used in literally describing his mental contents. It does not follow from the assumption that a person has in mind something that a word does not denote or express that the word cannot occur obliquely (and be interpreted literally) in that-clauses that provide some of his mental contents. As I have pointed out in Section IIIb, there is a range of cases in which we commonly reinterpret a person's incompletely understood words for purposes of mental-content attribution. But the present argument needs to show that deviant speaker-meaning always forces such reinterpretation.

In many of our examples, the idea that the subject has some deviant notion *in mind* has no intuitively clear application. (Consider the arthritis and mortgage examples). But even where this expression does seem to apply, the argument does not support the relevant conclusion. At best it shows that a notion deviantly associated with a word plays a role in the subject's attitudes. For example, someone who has in mind the notion of an agreement based on written documents when he says, "I have just entered into a contract," may be correctly said to believe that he has just entered into an agreement based on written documents. It does not follow from this that he *lacks* a belief or thought that he has just entered into a contract. In fact, in our view, the client's having the deviant notion in mind is a *likely consequence* of the fact that he believes that contracts are impossible without a written document.

Of course, given the first, more liberal set of interpretations of 'means something different', the fact that in our examples the subject means something different by his words (or at least applies them differently) is *implied* by certain of his beliefs. It is implied by a belief that he has arthritis in the thigh. A qualified version of the converse implication also holds. Given appropriate background assumptions, the fact that the subject has certain deviant (object-level) beliefs is implied by his meaning something different by his words. So far, no argument has shown that we cannot accept these implications and retain the literal interpretation of common mentalistic ascriptions.

The argument from deviant speaker-meaning downplays an intuitive feature that can be expected to be present in many of our examples. The subject's willingness to submit his statement and belief to the arbitration of an authority suggests a willingness to have his words taken in the normal way—regardless of mistaken associations with the word. Typically, the subject will regard recourse to a dictionary, and to the rest of us, as at once a check on his usage and his belief. When the verdict goes against him, he will not usually plead that we have simply misunderstood his views. This sort of behavior suggests that (given the sorts of background assump-

tions that common practice uses to distinguish our examples from those of foreigners, radical misunderstandings, and so forth) we can say that in a sense our man meant by 'arthritis' *arthritis*—where *'arthritis'* occurs, of course, obliquely. We can say this despite the fact that his incomplete understanding leads us, in one of the senses explicated earlier, to say that he meant something different by 'arthritis'.

If one tries to turn the argument from deviant speaker-meaning into a valid argument, one arrives at an assumption that seems to guide all three of the philosophical arguments I have discussed. The assumption is that what a person thinks his words mean, how he takes them, fully determines what attitudes he can express in using them: the contents of his mental states and events are strictly limited to notions, however idiosyncratic, that he understands; a person cannot think with notions he incompletely understands. But supplemented with this assumption, the argument begs the question at issue.

The least controversial justification of the assumption would be an appeal to standard practice in mentalistic attributions. But standard practice is what brought the assumption into question in the first place. Of course, usage is not sacred if good reasons for revising it can be given. But none have been.

The assumption is loosely derived, I think, from the old model according to which a person must be directly acquainted with, or must immediately apprehend, the contents of his thoughts. None of the objections explicitly invoke this model—and many of their proponents would reject it. But I think that all the objections derive some of their appeal from philosophical habits that have been molded by it. I shall discuss this model further in Section IV.

One may, of course, quite self-consciously neglect certain aspects of common mentalistic notions in the interests of a revised or idealized version of them. One such idealization could limit itself to just those attitudes involving "full understanding" (for some suitably specified notion of understanding). This limitation is less clearcut than one might suppose, since the notion of understanding itself tends to be used according to misleading stereotypes. Still, oversimplified models, idealizations, of mentalistic notions are defensible, as long as the character and purpose of the oversimplifications are clear. In my opinion, limiting oneself to "fully understood" attitudes provides no significant advantage in finding elegant and illuminating formal semantical theories of natural language. Such a strategy has perhaps a better claim in psychology, though even there its propriety is controversial. (Cf. Section IV.) More to the point, I think that models that neglect the relevant social factors in mentalistic attributions are not likely to provide long-run philosophical illumination of our actual mentalistic notions. But this view hardly admits of detailed support here and now.

Our argument in the preceding pages may, at a minimum, be seen as inveighing against a long-standing philosophical habit of denying that it *is* an oversimplification to make "full understanding" of a content a necessary condition for having a propositional attitude with that content. The oversimplification does not constitute neglect of some quirk of ordinary usage. Misunderstanding and partial understanding are pervasive and inevitable phenomena, and attributions of content despite them are an integral part of common practice.

I shall not here elaborate a philosophical theory of the social aspects of mentalistic phenomena, though in Section V I shall suggest lines such a theory might take. One of the most surprising and exciting aspects of the thought experiment is that its most literal interpretation provides a perspective on the mental that has received little serious development in the philosophical tradition. The perspective surely invites exploration.

IV. APPLICATIONS

I want to turn now to a discussion of how our argument bears on philosophical approaches to the mental that may be termed *individualistic*. I mean this term to be somewhat vague. But roughly, I intend to apply it to philosophical treatments that seek to see a person's intentional mental phenomena ultimately and purely in terms of what happens to the person, what occurs within him, and how he responds to his physical environment, without any essential reference to the social context in which he or the interpreter of his mental phenomena are situated. How I apply the term 'individualistic' will perhaps become clearer by reference to the particular cases that I shall discuss.

a. As I have already intimated, the argument of the preceding sections affects the traditional intro- (or extro-) spectionist treatments of the mind, those of Plato, Descartes, Russell, and numerous others. These treatments are based on a model that likens the relation between a person and the contents of his thought to seeing, where seeing is taken to be a kind of direct, immediate experience. On the most radical and unqualified versions of the model, a person's inspection of the contents of his thought is infallible: the notion of incompletely understanding them has no application at all.

The model tends to encourage individualistic treatments of the mental. For it suggests that what a person thinks depends on what occurs or "appears" within his mind. Demythologized, what a person thinks depends on the power and extent of his comprehension and on his internal dispositions toward the comprehended contents. The model is expressed in perhaps its crudest and least qualified form in a well-known passage by Russell:

> Whenever a relation of supposing or judging occurs, the terms to which the supposing or judging mind is related by the relation of supposing or judging must be terms with which the mind in question is acquainted. . . . It seems to me that the truth of this principle is evident as soon as the principle is understood.[5]

Acquaintance is (for Russell) direct, infallible, non-propositional, non-perspectival knowledge. "Terms" like concepts, ideas, attributes, forms, meanings, or senses are entities that occur in judgments more or less immediately before the mind on a close analogy to the way sensations are supposed to.

The model is more qualified and complicated in the writings of Descartes. In particular, he emphasizes the possibility that one might perceive the contents of

one's mind unclearly or indistinctly. He is even high-handed enough to write, "Some people throughout their lives perceive nothing so correctly as to be capable of judging it properly."[6] This sort of remark appears to be a concession to the points made in Sections I and II about the possibility of a subject's badly understanding his mental contents. But the concession is distorted by the underlying introspection model. On Descartes' view, the person's faculty of understanding, properly so-called, makes no errors. Failure to grasp one's mental contents results from either blind prejudice or interference by "mere" bodily sensations and corporeal imagery. The implication is that with sufficiently careful reflection on the part of the individual subject, these obstacles to perfect understanding can be cleared. That is, one need only be careful or properly guided in one's introspections to achieve full understanding of the content of one's intentional mental phenomena. Much that Descartes says suggests that where the subject fails to achieve such understanding, no definite content can be attributed to him. In such cases, his "thinking" consists of unspecifiable or indeterminate imagery; attribution of definite conceptual content is precluded. These implications are reinforced in Descartes' appeal to self-evident, indubitable truths:

> There are some so evident and at the same time so simple that we cannot think of them without believing them to be true. . . . For we cannot doubt them unless we think of them; and we cannot think of them without at the same time believing them to be true, i.e. we can never doubt them.[7]

The self-evidence derives from the mere understanding of the truths, and fully understanding them is a precondition for thinking them at all. It is this last requirement that we have been questioning.

In the Empiricist tradition Descartes' qualifications on the direct experience model—particularly those involving the interfering effects of sensations and imagery—tend to fall away. What one thinks comes to be taken as a sort of impression (whether more imagistic or more intellectual) on or directly grasped by the individual's mind. The tendency to make full comprehension on the part of the subject a necessary condition for attributing a mental content to him appears both in philosophers who take the content to be a Platonic abstraction and in those who place it, in some sense, inside the individual's mind. This is certainly the direction in which the model pulls, with its picture of immediate accessibility to the individual. Thus Descartes' original concessions to cases of incomplete understanding became lost as his model became entrenched. What Wölfflin said of painters is true of philosophers: they learn more from studying each other than from reflecting on anything else.

The history of the model makes an intricate subject. My remarks are meant merely to provide a suggestive caricature of it. It should be clear, however, that in broad outline the model mixes poorly with the thought experiment of Section II, particularly its first step. The thought experiment indicates that certain "linguistic truths" that have often been held to be indubitable can be thought yet doubted. And it shows that a person's thought *content* is not fixed by what goes on in him, or by what is accessible to him simply by careful reflection. The reason for this last

point about "accessibility" need not be that the content lies too deep in the unconscious recesses of the subject's psyche. Contents are sometimes "inaccessible" to introspection simply because much mentalistic attribution does not presuppose that the subject has fully mastered the content of his thought.

In a certain sense, the metaphysical model has fixed on some features of our use of mentalistic notions to the exclusion of others. For example, the model fastens on the facts that we are pretty good at identifying our own beliefs and thoughts, and we have at least a *prima facie* authority in reporting a wide range of them. It also underlines the point that for certain contents we tend to count understanding as a sufficient condition for acknowledging their truth. (It is debatable, of course, how well it explains or illumines these observations.) The model also highlights the truism that a certain measure of understanding is required of a subject if we are to attribute intentional phenomena on the basis of what he utters. As we have noted, chance or purely rote utterances provide no ground for mental content attributions; certain verbal pathologies are discounted. The model extrapolates from these observations to the claim that a person can never fail to understand the content of his beliefs or thoughts, or that the remedy for such failure lies within his own resources of reflection (whether autonomous and conscious, or unconscious and guided). It is this extrapolation that requires one to pass over the equally patent practice of attributing attitudes where the subject incompletely understands expressions that provide the content of those attitudes. Insistence on metalinguistic reinterpretation and talk about the indefiniteness of attitude contents in cases of incomplete understanding seem to be rearguard defenses of a vastly overextended model.

The Cartesian-Russellian model has few strict adherents among prominent linguistic philosophers. But although it has been widely rejected or politely talked around, claims that it bore and nurtured are commonplace, even among its opponents. As we have seen in the objections to the first step of the argument of Section II, these claims purport to restrict the contents we can attribute to a person on the basis of his use of language. The restrictions simply mimic those of Descartes. Freed of the picturesque but vulnerable model that formed them, the claims have assumed the power of dogma. Their strictures, however, misrepresent ordinary mentalistic notions.

b. This century's most conspicuous attempt to replace the traditional Cartesian model has been the behaviorist movement and its heirs. I take it as obvious that the argument of Section II provides yet another reason to reject the most radical version of behaviorism—"philosophical," "logical" or "analytical" behaviorism. This is the view that mentalistic attributions can be "analytically" defined, or given strict meaning equivalences, purely in non-mental, behavioral terms. No analysis resting purely on the individual's dispositions to behavior can give an "analytic" definition of a mental content attribution because we can conceive of the behavioral definiens applying while the mentalistic definiendum does not. But a new argument for this conclusion is hardly needed since "philosophical" behaviorists are, in effect, extinct.

There is, however, an heir of behaviorism that I want to discuss at somewhat greater length. The approach sometimes goes by the name "functionalism," although

that term is applied to numerous slogans and projects, often vaguely formulated. Even views that seem to me to be affected by our argument are frequently stated so sketchily that one may be in considerable doubt about what is being proposed. So my remarks should be taken less as an attempt to refute the theses of particular authors than as an attack on a way of thinking that seems to inform a cluster of viewpoints. The quotations I give in footnotes are meant to be suggestive, if not always definitive, of the way of thinking the argument tells against.[8]

The views affected by the argument of Section II attempts to give something like a philosophical "account" of the mental. The details and strategy—even the notion of "account"—vary from author to author. But a recurrent theme is that mental notions are to be seen ultimately in terms of the individual subject's input, output, and inner dispositions and states, where these latter are characterized purely in terms of how they lead to or from output, input, or other inner states similarly characterized. Mental notions are to be explicated or identified in functional, nonmentalistic, non-intentional terminology. Proponents of this sort of idea are rarely very specific about what terms may be used in describing input and output, or even what sorts of terms count as "functional" expressions. But the impression usually given is that input and output are to be specified in terms (acceptable to a behaviorist) of irritations of the subject's surfaces and movements of his body. On some versions, neurophysiological terms are allowed. More recently, there have been liberalized appeals to causal input and output relations with particular, specified physical objects, stuffs, or magnitudes. Functional terms include terms like 'causes', 'leads to with probability n', and the like. For our purposes, the details do not matter much, as long as an approach allows no mentalistic or other intentional terms (such as 'means' or that-clauses) into its vocabulary, and as long as it applies to individuals taken one by one.

A difference between this approach and that of philosophical behaviorism is that a whole array of dispositional or functional states—causally or probabilistically interrelated—may enter into the "account" of a single mental attribution. The array must be ultimately secured to input and output, but the internal states need not be so secured one by one. The view is thus not immediately vulnerable to claims against simplistic behaviorisms, that a *given* stimulus-response pattern may have different contents in different social contexts. Such claims, which hardly need a defender, have been tranquilly accepted on this view. The view's hope is that differences in content depend on functional differences in the individual's larger functional structure. From this viewpoint, analytical behaviorism erred primarily in its failure to recognize the interlocking or wholistic character of mental attributions and in its oversimplification of theoretical explanation.

As I said, the notion of an account of the mental varies from author to author. Some authors take over the old-fashioned ideal of an "analysis" from philosophical behaviorism and aim at a definition of the meaning of mentalistic vocabulary, or a definitional elimination of it. Others see their account as indicating a series of scientific hypotheses that identify mental states with causal or functional states, or roles, in the individual. These authors reject behaviorism's goal of providing mean-

ing equivalences, as well as its restrictive methods. The hypotheses are supposed to be type or property identities and are nowadays often thought to hold necessarily, even if they do not give meaning relations. Moreover, these hypotheses are offered not merely as speculation about the future of psychology, but as providing a philosophically illuminating account of our ordinary notion of the mental. Thus if the view systematically failed to make plausible type identities between functional states and mental states, ordinarily construed, then by its own lights it would have failed to give a philosophical "account" of the mental. I have crudely over-schematized the methodological differences among the authors in this tradition. But the differences fall roughly within the polar notions of *account* that I have described. I think our discussion will survive the oversimplifications.[9]

Any attempt to give an account of specific beliefs and thoughts along the lines I have indicated will come up short. For we may fix the input, output, and total array of dispositional or functional states of our subject, as long as these are non-intentionally described and are limited to what is relevant to accounting for his activity taken in isolation from that of his fellows. But we can still conceive of his mental contents as varying. Functionally equivalent people—on any plausible notion of functional equivalence that has been sketched—may have non-equivalent mental-state and event contents, indicated by obliquely non-equivalent content clauses. Our argument indicates a systematic inadequacy in attempts of the sort I described.

Proponents of functionalist accounts have seen them as revealing the true nature of characteristic marks of the mental and as resolving traditional philosophical issues about such marks. In the case of beliefs, desires, and thoughts, the most salient mark is intentionality—the ill-specified information-bearing, representational feature that seems to invest these mental states and events.[10] In our terminology, accounting for intentionality largely amounts to accounting for the content of mental states and events. (There is also, of course, the application of content in *de re* cases. But we put this aside here.) Such content is clearly part of what the functional roles of our subjects' states fail to determine.

It is worth re-emphasizing here that the problem is unaffected by suggestions that we specify input and output in terms of causal relations to particular objects or stuffs in the subject's physical environment. Such specifications may be thought to help with some examples based on indexicality or psychological success verbs, and perhaps in certain arguments concerning natural kind terms (though even in these cases I think that one will be forced to appeal to intentional language). (Cf. note 2.) But this sort of suggestion has no easy application to our argument. For the relevant causal relations between the subject and the physical environment to which his terms apply—where such relations are non-intentionally specified—were among the elements held constant while the subject's beliefs and thoughts varied.

The functionalist approaches I have cited seem to provide yet another case in which mental contents are not plausibly accounted for in non-intentional terms. They are certainly not explicable in terms of causally or functionally specified states and events of the *individual* subject. The intentional or semantical role of mental

states and events is not a function merely of their functionally specified roles in the individual. The failure of these accounts of intentional mental states and events derives from an underestimation of socially dependent features of cognitive phenomena.

Before extending the application of our argument, I want to briefly canvass some ways of being influenced by it, ways that might appeal to someone fixed on the functionalist ideal. One response might be to draw a strict distinction between mental states, ordinarily so-called, and psychological states. One could then claim that the latter are the true subject matter of the science of psychology and may be identified with functional states functionally specified, after all. Thus one might claim that the subject was in the same psychological (functional) states in both the actual and the imagined situations, although he had different beliefs and thoughts ordinarily so-called.

There are two observations that need to be entered about this position. The first is that it frankly jettisons much of the philosophical interest of functionalist accounts. The failure to cope with mental contents is a case in point. The second observation is that it is far from clear that such a distinction between the psychological and the mental is or will be sanctioned by psychology itself. Functionalist accounts arose as philosophical interpretations of developments in psychology influenced by computer theory. The interpretations have been guided by philosophical interests, such as throwing light on the mind-body problem and accounting for mentalistic features in non-mentalistic terms. But the theories of cognitive psychologists, including those who place great weight on the computer analogy, are not ordinarily purified of mentalistic or intentional terminology. Indeed, intentional terminology plays a central role in much contemporary theorizing. (This is also true of theories that appeal to "sub-personal" states or processes. The "sub-personal" states themselves are often characterized intentionally.) Purifying a theory of mentalistic and intentional features in favor of functional or causal features is more clearly demanded by the goals of philosophers than by the needs of psychology. Thus it is at least an open question whether functional approaches of the sort we have discussed give a satisfactory account of *psychological* states and events. It is not evident that psychology will ever be methodologically "pure" (or theoretically purifiable by some definitional device) in the way these approaches demand. *This* goal of functionalists may be simply a meta-psychological mistake.

To put the point another way, it is not clear that functional states, characterized purely in functional, non-intentional terms (and non-intentional descriptions of input and output) are the natural subject matter of psychology. Psychology would, I think, be an unusual theory if it restricted itself (or could be definitionally restricted) to specifying abstract causal or functional structures in purely causal or functional terms, together with vocabulary from other disciplines. Of course, it *may* be that functional states, functionally specified, form a psychological natural kind. And it is certainly not to be assumed that psychology will respect ordinary terminology in its individuation of types of psychological states and events. Psychology must run its own course. But the assumption that psychological terminology will be ultimately non-intentional and purely functional seems without strong support.

More important from our viewpoint, if psychology did take the individualistic route suggested by the approaches we have cited, then its power to illumine the everyday phenomena alluded to in mentalistic discourse would be correspondingly limited.

These remarks suggest a second sort of functionalist response to the argument of Section II, one that attempts to take the community rather than the individual as the object of functional analysis. One might, for example, seek to explain an individual's responsibility to communal standards in terms of his having the right kind of interaction with other individuals who collectively had functional structures appropriate to those standards. Spelling out the relevant notions of interaction and appropriateness is, of course, anything but trivial. (Cf. Section V.) Doing so in purely functional, non-intentional terms would be yet a further step. Until such a treatment is developed and illustrated in some detail, there is little point in discussing it. I shall only conjecture that, if it is to remain non-intentional, such a treatment is likely to be so abstract—at least in our present state of psychological and sociological ignorance—that it will be unilluminating from a philosophical point of view. Some of the approaches we have been discussing already more than flirt with this difficulty.

c. Individualistic assumptions about the mental have infected theorizing about the relation between mind and meaning. An example is the Gricean project of accounting for conventional or linguistic meaning in terms of certain complex intentions and beliefs of individuals.[11] The Gricean program analyzes conventional meaning in terms of subtle "mutual knowledge," or beliefs and intentions about each others' beliefs and intentions, on the part of most or all members of a community. Seen as a quasi-definitional enterprise, the program presupposes that the notion of an individual's believing or intending something is always "conceptually" independent of the conventional meaning of symbols used to express that something. Insofar as 'conceptually' has any intuitive content, this seems not to be the case. Our subject's belief or intention contents can be conceived to vary simply by varying conventions in the community around him. The content of individuals' beliefs seems sometimes to depend partly on social conventions in their environment. It is true that our subjects are actually rather abnormal members of their community, at least with respect to their use and understanding of a given word. But normality here is judged against the standards set by communal conventions. So stipulating that the individuals whose mental states are used in defining conventional meaning be relevantly normal will not avoid the circularity that I have indicated. I see no way to do so. This charge of circularity has frequently been raised on intuitive grounds. Our argument gives the intuitions substance. Explicating convention in terms of belief and intention may provide various sorts of insight. But it is not defining a communal notion in terms of individualistic notions. Nor is it reducing, in any deep sense, the semantical, or the intentional generally, to the psychological.

d. Individualistic assumptions have also set the tone for much discussion of the ontology of the mental. This subject is too large to receive detailed consideration here. It is complicated by a variety of crosscurrents among different projects, methodologies, and theses. I shall only explore how our argument affects a certain line

of thinking closely allied to the functionalist approaches already discussed. These approaches have frequently been seen as resuscitating an old argument for the materialist identity theory. The argument is three-staged. First, one gives a philosophical "account" of each mentalistic locution, an account that is *prima facie* neutral as regards ontology. For example, a belief or a thought that sofas are comfortable is supposed to be accounted for as one functionally specified state or event within an array of others—all of which are secured to input and output. Second, the relevant functionally specified states or events are expected to be empirically correlated or correlatable with physiological states or events in a person (states or events that have those functions). The empirical basis for believing in these correlations is claimed to be provided by present or future physical science. The nature of the supposed correlations is differently described in different theories. But the most prevalent views expect only that the correlations will hold for each organism and person (perhaps at a given time) taken one by one. For example, the functionally specified event type that is identified with a thought that sofas are comfortable may be realized in one person by an instance (or "token") of one physiological event type, and in another person by an instance of another physiological event type. Third, the ("token") mental state or event in the person is held to be identical with the relevant ("token") physiological state or event, on general grounds of explanatory simplicity and scientific method. Sometimes, this third stage is submerged by building uniqueness of occupancy of functional role into the first stage.[12]

I am skeptical about this sort of argument at every stage. But I shall doubt only the first stage here. The argument we gave in Section II directly undermines the attempt to carry out the first stage by recourse to the sort of functionalist approaches that we discussed earlier. Sameness of functional role, individualistically specified, is compatible with difference of content. I know of no better non-intentional account of mentalistic locutions. If a materialist argument of this genre is to arrive, it will require a longer first step.

I shall not try to say whether there is a philosophically interesting sense in which intentional mental phenomena are physical or material. But I do want to note some considerations against materialist *identity* theories.

State-like phenomena (say, beliefs) raise different problems from event-like phenomena (say, occurrent thoughts). Even among identity theorists, it is sometimes questioned whether an identity theory is the appropriate goal for materialism in the case of states. Since I shall confine myself to identity theories, I shall concentrate on event-like phenomena. But our considerations will also bear on views that hope to establish some sort of token identity theory for mental states like beliefs.

One other preliminary. I want to remain neutral about how best to describe the relation between the apparent event-like feature of occurrent thoughts and the apparent relational feature (their relation to a content). One might think of there being an event, the token thought event, that is in a certain relation to a content (indicated by the that-clause). One might think of the event as consisting—as not being anything "over and above"—the relevant relation's holding at a certain time

between a person and a content. Or one might prefer some other account. From the viewpoint of an identity theory, the first way of seeing the matter is most advantageous. So I shall fit my expositon to that point of view.

Our ordinary method of identifying occurrent thought events and differentiating between them is to make reference to the person or organism to whom the thought occurs, the time of its occurrence, and the content of the thought. If person, time, and content are the same, we would normally count the thought event the same. If any one of these parameters differs in descriptions of thought events (subject to qualifications about duration), then the events or occurrences described are different. Of course, we can differentiate between events using descriptions that do not home in on these particular parameters. But these parameters are dominant. (It is worth noting that differentiations in terms of causes and effects usually tend to rely on the content of mental events or states at some point, since mental states or events are often among the causes or effects of a given mental event, and these causes or effects will usually be identified partly in terms of their content.) The important point for our purposes is that in ordinary practice, sameness of thought content (or at least some sort of strong equivalence of content) is taken as a necessary condition for sameness of thought occurrence.

Now one might codify and generalize this point by holding that no occurrence of a thought (that is, no token thought event) could have a different (or extensionally non-equivalent) content and be the very same token event. If this premise is accepted, then our argument of Section II can be deployed to show that a person's thought event is not *identical* with any event in him that is described by physiology, biology, chemistry, or physics. For let *b* be any given event described in terms of one of the physical sciences that occurs in the subject while he thinks the relevant thought. Let *'b'* be such that it denotes the same physical event occurring in the subject in our counterfactual situation. (If you want, let *'b'* be rigid in Kripke's sense, though so strong a stipulation is not needed.) The second step of our argument in Section II makes it plausible that *b* need not be affected by counterfactual differences in the communal use of the word 'arthritis'. Actually, the subject thinks that his ankles are stiff from arthritis, while *b* occurs. But we can conceive of the subject's *lacking* a thought event that his ankles are stiff from arthritis, while *b* occurs. Thus in view of our initial premise, *b* is not identical with the subject's occurrent thought.[13]

Identity theorists will want to reject the first premise—the premise that no event with a different content could be identical with a given thought event. On such a view, the given thought event that his ankles are stiff from arthritis might well have been a thought that his ankles are stiff from tharthritis, yet be precisely the same token thought event. Such a view is intuitively very implausible. I know of only one reasonably spelled-out basis of support for this view. Such a basis would be provided by showing that mentalistic phenomena are causal or functional states, in one of the strong senses discussed earlier, and that mental events are physical tokens or realizations of those states. If 'that thought that his ankles are stiff from arthritis' could be accounted for in terms like 'that event with such and such a

causal or functional role' (where 'such and such' does not itself involve intentional terminology), and if independently identified physical events systematically filled these roles (or realized these states), we could perhaps see a given thought event as having a different role—and hence content—in different possible situations. Given such a view, the functional specification could perhaps be seen as revealing the contingency of the intentional specification as applied to mental event tokens. Just as we can imagine a given physiological event that actually plays the role of causing the little finger to move two inches, as playing the role of causing the little finger to move three inches (assuming compensatory differences in its physiological environment), so we could perhaps imagine a given thought as having a different functional role from its actual one—and hence, assuming the functionalist account, as having a different content. But the relevant sort of functionalist account of intentional phenomena has not been made good.[14]

The recent prosperity of materialist-functionalist ways of thinking has been so great that it is often taken for granted that a given thought event might have been a thought with a different, obliquely non-equivalent content. Any old event, on this view, could have a different content, a different significance, if its surrounding context were changed. But in the case of occurrent thoughts—and intentional mental events generally—it is hardly obvious, or even initially plausible, that anything is more essential to the identity of the event than the content itself. Materialist identity theories have schooled the imagination to picture the content of a mental event as varying while the event remains fixed. But whether such imaginings are possible fact or just philosophical fancy is a separate question.[15]

At any rate, functionalist accounts have not provided adequate specification of what it is to be a thought that ____, for particular fillings of the blank. So a specification of a given thought event in functionalist terms does not reveal the contingency of the usual, undisputed intentional specifications.

Well, *is* it possible for a thought event to have had a different content from the one it has and be the very same event? It seems to me natural and certainly traditional to assume that this is not possible. Rarely, however, have materialists seen the identity theory as natural or intuitive. Materialists are generally revisionist about intuitions. What is clear is that we currently do identify and distinguish thought events primarily in terms of the person who has them, the rough times of their occurrence, and their contents. And we do assume that a thought event with a different content is a different thought event (insofar as we distinguish at all between the thinking event and the person's being related to a thought content at a time). I think these facts give the premise *prima facie* support and the argument against the identity theory some interest. I do not claim that we have *"a priori"* certainty that no account of intentional phenomena will reveal intentional language to be only contingently applicable to belief states or thought events. I am only dubious.

One might nurture faith or hope that some more socially oriented functionalist specification could be found. But no such specification is ready to hand. And I see no good reason to think that one must be found. Even if such a specification were found, it is far from clear that it would deflect the argument against the iden-

tity theory just considered. The "functional" states envisaged would depend not merely on what the individual does and what inner causal states lead to his activity—non-intentionally specified—but also on what his fellows do. The analogy between functional states and physiological states in causing the individual's internal and external activity was the chief support for the view that a given token mental event might have been a token of a different content. But the envisaged socially defined "functional states" bear no intuitive analogy to physiological states or other physical causal states within the individual's body. Their function is not simply that of responding to environmental influences and causing the individual's activity. It is therefore not clear (short of *assuming* an identity theory) that any event that is a token of one of the envisaged socially defined "functional states" could have been a token of a different one. The event might be essentially identified in terms of its social role. There is as yet no reason to identify it in terms of physically described events in the individual's body. Thus it is not clear that such a socially oriented functional account of thought contents would yield grounds to believe that the usual intentional specifications of mental events are merely contingent. It is, I think, even less clear that an appropriate socially oriented functional account is viable.

Identity theories, of course, do not exhaust the resources of materialism. To take one example, our argument does not speak directly to a materialism based on composition rather than identity. On such a view, the same physical material might compose different thoughts in different circumstances. I shall say nothing evaluative about this sort of view. I have also been silent about other arguments for a token identity theory—such as those based on philosophical accounts of the notions of causality or explanation. Indeed, my primary interest has not been ontology at all. It has been to identify and question individualistic assumptions in materialist as well as Cartesian approaches to the mental.

V. MODELS OF THE MENTAL

Traditional philosophical accounts of mind have offered metaphors that produce doctrine and carry conviction where argument and unaided intuition flag. Of course, any such broad reconstructions can be accused of missing the pied beauties of the natural article. But the problem with traditional philosophy of mind is more serious. The two overwhelmingly dominant metaphors of the mental—the infallible eye and the automatic mechanism—have encouraged systematic neglect of prominent features of a wide range of mental phenomena, broadly speaking, social features. Each metaphor has its attractions. Either can be elaborated or doctored to fit the facts that I have emphasized. But neither illumines those facts. And both have played some part in inducing philosophers to ignore them.

I think it optimistic indeed to hope that any one picture, comparable to the traditional ones, will provide insight into all major aspects of mental phenomena. Even so, a function of philosophy is to sketch such pictures. The question arises whether one can make good the social debts of earlier accounts while retaining at

least some of their conceptual integrity and pictorial charm. This is no place to start sketching. But some summary remarks may convey a sense of the direction in which our discussion has been tending.

The key feature of the examples of Section II was the fact that we attribute beliefs and thoughts to people even where they incompletely understand contents of those very beliefs and thoughts. This point about intentional mental phenomena is not everywhere applicable: non-linguistic animals do not seem to be candidates for misunderstanding the contents of their beliefs. But the point is certainly salient and must be encompassed in any picture of intentional mental phenomena. Crudely put, wherever the subject has attained a certain competence in large relevant parts of his language and has (implicitly) assumed a certain general commitment or responsibility to the communal conventions governing the language's symbols, the expressions the subject uses take on a certain inertia in determining attributions of mental content to him. In particular, the expressions the subject uses sometimes provide the content of his mental states or events even though he only partially understands, or even misunderstands, some of them. Global coherence and responsibility seem sometimes to override localized incompetence.

The detailed conditions under which this "inertial force" is exerted are complicated and doubtless more than a little vague. Clearly, the subject must maintain a minimal internal linguistic and rational coherence and a broad similarity to others' use of the language. But meeting this condition is hardly sufficient to establish the relevant responsibility. For the condition is met in the case of a person who speaks a regional dialect (where the same words are sometimes given different applications). The person's aberrations relative to the larger community may be normalities relative to the regional one. In such cases, of course, the regional conventions are dominant in determining what contents should be attributed. At this point, it is natural to appeal to etiological considerations. The speaker of the dialect developed his linguistic habits from interaction with others who were a party to distinctively regional conventions. The person is committed to using the words according to the conventions maintained by those from whom he learned the words. But the situation is more complicated than this observation suggests. A person born and bred in the parent community might simply decide (unilaterally) to follow the usage of the regional dialect or even to fashion his own usage with regard to particular words, self-consciously opting out of the parent community's conventions in these particulars. In such a case, members of the parent community would not, and should not, attribute mental contents to him on the basis of homophonic construal of his words. Here the individual's intentions or attitudes toward communal conventions and communal conceptions seem more important than the causal antecedents of his transactions with a word—unless those intentions are simply included in the etiological story.

I shall not pursue these issues here. The problem of specifying the conditions under which a person has the relevant general competence in a language and a responsibility to its conventions is obviously complicated. The mixture of "causal" and intentional considerations relevant to dealing with it has obvious near analogs

in other philosophical domains (etiological accounts of perception, knowledge, reference). I have no confidence that all of the details of the story would be philosophically interesting. What I want to stress is that to a fair degree, mentalistic attribution rests not on the subject's having mastered the contents of the attribution, and not on his having behavioral dispositions peculiarly relevant to those contents, but on his having a certain responsibility to communal conventions governing, and conceptions associated with, symbols that he is disposed to use. It is this feature that must be incorporated into an improved model of the mental.

I think it profitable to see the language of content attribution as constituting a complex *standard* by reference to which the subject's mental states and events are estimated, or an abstract grid on which they are plotted. Different people may vary widely in the degree to which they master the elements and relations within the standard, even as it applies to them all. This metaphor may be developed in several directions and with different models: applied geometry, measurement of magnitudes, evaluation by a monetary standard, and so forth. A model I shall illustrate briefly here borrows from musical analysis.

Given that a composer has fulfilled certain general conditions for establishing a musical key, his chordal structures are plotted by reference to the harmonic system of relations appropriate to the tonic key. There is vast scope for variation and novelty within the harmonic framework. The chords may depart widely from traditional "rules" or practices governing what count as interesting or "reasonable" chordal structures and progressions. And the composer may or may not grasp the harmonic implications and departures present in his composition. The composer may sometimes exhibit harmonic incompetence (and occasionally harmonic genius) by radically departing from those traditional rules. But the harmonic system of relations applies to the composition in any case. Once established, the tonic key and its associated harmonic framework are applied unless the composer takes pains to set up another tonic key or some atonal arrangement (thereby intentionally opting out of the original tonal framework), or writes down notes by something like a slip of the pen (suffering mechanical interference in his compositional intentions), or unintentionally breaks the harmonic rules in a massive and unprincipled manner (thereby indicating chaos or complete incompetence). The tonic key provides a standard for describing the composition. The application of the standard depends on the composer's maintaining a certain overall coherence and minimal competence in conforming to the standard's conventions. And there are conditions under which the standard would be replaced by another. But once applied, the harmonic framework—its formal interrelations, its applicability even to deviant, pointless progressions—is partly independent of the composer's degree of harmonic mastery.

One attractive aspect of the metaphor is that it has some application to the case of animals. In making sounds, animals do sometimes behave in such a way that a harmonic standard can be roughly applied to them, even though the standard, at least in any detail, is no part of what they have mastered. Since they do not master the standard (though they may master some of its elements), they are not candidates for partial understanding or misunderstanding. (Of course, this may be said of

many people as regards the musical standard.) The standard applies to both animals and people. But the conditions for its application are sensitive in various ways to whether the subject himself has mastered it. Where the subject does use the standard (whether the language, or a system of key relationships), his uses take on special weight in applications of the standard to him.

One of the metaphor's chief virtues is that it encourages one to seek social explications for this special weight. The key to our attribution of mental contents in the face of incomplete mastery or misunderstanding lies largely in social functions associated with maintaining and applying the standard. In broad outline, the social advantages of the "special weight" are apparent. Symbolic expressions are the overwhelmingly dominant source of detailed information about what people think, intend, and so forth. Such detail is essential not only to much explanation and prediction, but also to fulfilling many of our cooperative enterprises and to relying on one another for second-hand information. Words interpreted in conventionally established ways are familiar, palpable, and public. They are common coin, a relatively stable currency. These features are crucial to achieving the ends of mentalistic attribution just cited. They are also critical in maximizing interpersonal comparability. And they yield a bias toward taking others at their word and avoiding *ad hoc* reinterpretation, once overall agreement in usage and commitment to communal standards can be assumed.

This bias issues in the practice of expressing even many differences in understanding without reinterpreting the subject's words. Rather than reinterpret the subject's word 'arthritis' and give him a trivially true object-level belief and merely a false metalinguistic belief about how 'arthritis' is used by others, it is common practice, and correct, simply to take him at his word.

I hardly need re-emphasize that the situation is vastly more complicated than I have suggested in the foregoing paragraphs. Insincerity, tongue slips, certain malapropisms, subconscious blocks, mental instability all make the picture more complex. There are differences in our handling of different sorts of expressions, depending, for example, on how clear and fixed social conventions regarding the expressions are. There are differences in our practices with different subject matters. There are differences in our handling of different degrees of linguistic error. There are differences in the way meaning-, assertion-, and mental-contents are attributed. (Cf. note 4.) I do not propose ignoring these points. They are all parameters affecting the inertial force of "face value" construal. But I want to keep steadily in mind the philosophically neglected fact about social practice: Our attributions do not require that the subject always correctly or fully understand the content of his attitudes.

The point suggests fundamental misorientations in the two traditional pictures of the mental. The authority of a person's reports about his thoughts and beliefs (*modulo* sincerity, lack of subconscious interference, and so forth) does not issue from a special intellectual vision of the contents of those thoughts and beliefs. It extends even to some cases in which the subject incompletely understands those contents. And it depends partly on the social advantages of maintaining communally established standards of communication and mentalistic attribution. Likewise,

the descriptive and explanatory role of mental discourse is not adequately modeled by complex non-intentional mechanisms or programs for the production of an individual's physical movement and behavior. Attributing intentional mentalistic phenomena to individuals serves not only to explain their behavior viewed in isolation but also to chart their activity (intentional, verbal, behavioral, physical) by complex comparison to others—and against socially established standards.[16] Both traditional metaphors make the mistake, among others, of treating intentional mental phenomena individualistically. New approaches must do better. The sense in which man is a social animal runs deeper than much mainstream philosophy of mind has acknowledged.[17]

Notes

1. Our examples suggest points about learning that need exploration. It would seem naive to think that we first attain a mastery of expressions or notions we use and then tackle the subject matters we speak and think about in using those expressions or notions. In most cases, the processes overlap. But while the subject's understanding is still partial, we sometimes attribute mental contents in the very terms the subject has yet to master. Traditional views take mastering a word to consist in matching it with an already mastered (or innate) concept. But it would seem, rather, that many concepts (or mental content components) are like words in that they may be employed before they are mastered. In both cases, employment appears to be an integral part of the process of mastery.

2. A development of a similar theme may be found in Hilary Putnam's notion of a division of linguistic labour. Cf. "The Meaning of 'Meaning'," *Philosophical Papers* 2 (London, 1975) pp. 227 ff. Putnam's imaginative work is in other ways congenial with points I have developed. Some of his examples can be adapted in fairly obvious ways so as to give an argument with different premises, but a conclusion complementary to the one I arrive at in Section IIa:

Consider Alfred's belief contents involving the notion of water. Without changing Alfred's (or his fellows') non-intentional phenomenal experiences, internal physical occurrences, or dispositions to respond to stimuli on sensory surfaces, we can imagine that not water (H_2O), but a different liquid with different structure but similar macro-properties (and identical phenomenal properties) played the role in his environment that water does in ours. In such a case, we could ascribe no content clauses to Alfred with 'water' in oblique position. His belief contents would differ. The conclusion (with which I am in sympathy) is that mental contents are affected not only by the physical and qualitatively mental way the person is, but by the nature of his *physical environment*.

Putnam himself does not give quite this argument. He nowhere states the first and third steps, though he gives analogs of them for the meaning of 'water'. This is partly just a result of his concentration on meaning instead of propositional attitudes. But some of what he says even seems to oppose the argument's conclusion. He remarks in effect that the subject's *thoughts* remain constant between his actual and counterfactual cases (p. 224). In his own argument he explicates the difference between actual and counterfactual cases in terms of a difference in the extension of terms, not a difference in those aspects of their meaning that play a role in the cognitive life of the subject. And he tries to explicate his examples in terms of indexicality—a mistake, I think, and one that tends to divert attention from major implications of the examples he gives. (Cf. Section IId.) In my view, the examples do illustrate the fact that all attitudes involving natural kind notions, including *de dicto* attitudes, presuppose *de re* attitudes. But the examples do not show that natural kind linguistic expressions are in any ordinary sense indexical. Nor do they show that beliefs involving natural kind notions are always *de re*. Even if they did, the change from actual to counterfactual cases would affect oblique occurrences of natural kind terms in that-clauses—occurrences that are the key to attributions of cognitive content.

(Cf. above and note 3.) In the cited paper and earlier ones, much of what Putnam says about psychological states (and implies about mental states) has a distinctly individualistic ring. Below in Section IV, I criticize viewpoints about mental phenomena influenced by and at least strongly suggested in his earlier work on functionalism. (Cf. note 9.)

On the other hand, Putnam's articulation of social and environmental aspects of the meaning of natural kind terms complements and supplements our viewpoint. For me, it has been a rich rewarder of reflection. More recent work of his seems to involve shifts in his viewpoint on psychological states. It may have somewhat more in common with our approach than the earlier work, but there is much that I do not understand about it.

The argument regarding the notion of water that I extracted from Putnam's paper is narrower in scope than our argument. The Putnam-derived argument seems to work only for natural kind terms and close relatives. And it may seem not to provide as direct a threat to certain versions of functionalism that I discuss in Section IV: At least a few philosophers would claim that one could accommodate the Putnamian argument in terms of *non*-intentional formulations of input-output relations (formulations that make reference to the specific nature of the physical environment). Our argument does not submit to this maneuver. In our thought experiment, the physical environment (sofas, arthritis, and so forth in our examples) and the subject's causal relations with it (at least as these are usually conceived) were held constant. The Putnamian argument, however, has fascinatingly different implications from our argument. I have not developed these comparisons and contrasts here because doing justice to Putnam's viewpoint would demand a distracting amount of space, as the ample girth of this footnote may suggest.

3. I have discussed *de re* mental phenomena in "Belief *De Re*," *The Journal of Philosophy* 74 (1977):338-62. There I argue that all attitudes with content presuppose *de re* attitudes. Our discussion here may be seen as bearing on the details of this presupposition. But for reasons I merely sketch in the next paragraph, I think it would be a superficial viewpoint that tried to utilize our present argument to support the view that nearly all intentional mental phenomena are covertly indexical or *de re*.

4. Cf. my "Belief and Synonymy," *The Journal of Philosophy* 75 (1978):119-38, Section III, where I concentrate on attribution of belief contents containing "one criterion" terms like 'vixen' or 'fortnight' which the subject misunderstands. The next several pages interweave some of the points in that paper. I think that a parallel thought experiment involving even these words is constructible, at least for a narrowly restricted set of beliefs. We can imagine that the subject believes that some female foxes—say, those that are virgins—are not vixens. Or he could believe that a fortnight is a period of ten days. (I believed this for many years.) Holding his physical history, qualitative experience, and dispositions constant, we can conceive of his linguistic community defining these terms as he actually misunderstands them. In such a case, his belief contents would differ from his actual ones.

5. Bertrand Russell, *Mysticism and Logic* (London, 1959), p. 221. Although Russell's statement is unusually unqualified, its kinship to Descartes' and Plato's model is unmistakable. Cf. Plato, *Phaedrus*, 249b-c, *Phaedo*, 47b6-c4; Descartes, *Philosophical Works*, eds. Haldane and Ross 2 vols. (New York, 1955), *Rules for the Direction of the Mind*, section XII, Vol. I, pp. 41-42, 45; *Principles of Philosophy*, Part I, XXXII-XXXV. Vol. I, pp. 232-33; *Replies*, Vol. II, 52; Hume, *A Treatise of Human Nature*, I, 3,5; II, 2,6; Kant, *A Critique of Pure Reason*, A7-B11; Frege, *The Foundations of Arithmetic*, section 105; G. E. Moore, *Principia Ethica*, 86.

6. Descartes, *Principles of Philosophy*, XLV-XLI.

7. Descartes, *Philosophical Works*, Vol. II., *Replies*, p. 42.

8. Certain movements sometimes called "functionalist" are definitely not my present concern. Nothing I say is meant to oppose the claim that hypotheses in psychology do and should make reference to "sub-personal" states and processes in explaining human action and ordinary mental states and processes. My remarks may bear on precisely how such hypotheses are construed philosophically. But the hypotheses themselves must be judged primarily by their fruits. Similarly, I am not concerned with the claim that computers provide an illuminating perspec-

tive for viewing the mind. Again, our view may bear on the interpretation of the computer analogy, but I have no intention of questioning its general fruitfulness. On the other hand, insofar as functionalism is merely a slogan to the effect that "once you see how computers might be made to work, you realize such and such about the mind," I am inclined to let the cloud condense a little before weighing its contents.

9. A representative of the more nearly "analytical" form of functionalism is David Lewis, "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy* 50 (1972):249-58: "Applied to common-sense psychology—folk science rather than professional science, but a theory nonetheless—we get the hypothesis . . . that a mental state M . . . is definable as the occupant of a certain causal role R—that is, as the state, of whatever sort, that is causally connected in specified ways to sensory stimuli, motor responses, and other mental states" (249-50). Actually, it should be noted that the argument of Section I applies to Lewis's position less directly than one might suppose. For reasons unconnected with matters at hand, Lewis intends his *definition* to apply to relational mentalistic predicates like 'thinks' but not to complex predicates that identify actual mental states or events, like 'thinks that snow is white'. Cf. *Ibid.*, p. 256, n13. This seems to me a puzzling halfway house for some of Lewis's philosophical purposes. But our argument appears to apply anyway, since Lewis is explicit in holding that physical facts about a person taken in isolation from his fellows "determine" all his specific intentional events and states. Cf. 'Radical Interpretation', *Synthese* 27 (1974):331ff. I cite Lewis's definitional approach because it has been the most influential recent piece of its genre, and many of those influenced by it have not excluded its application to specific intentional mental states and events. Other representatives of the definitional approach are J. J. C. Smart, "Further Thoughts on the Identity Theory," *Monist* 56 (1972):149-62; D. W. Armstrong, *A Materialist Theory of Mind* (London, 1968), pp. 90-91 and *passim*; Sidney Shoemaker, "Functionalism and Qualia," *Philosophical Studies* 27 (1975):306-7. A representative of the more frequently held "hypothesis" version of functionalism is Hilary Putnam, "The Mental Life of Some Machines," *Philosophical Papers* 2 (Cambridge, 1975), and "The Nature of Mental States," *Ibid.*, cf. p. 437: ". . . if the program of finding psychological laws that are not species specific . . . ever succeeds, then it will bring in its wake a delineation of the kind of functional organization that is necessary and sufficient for a given psychological state, as well as a precise definition of the notion 'psychological state'." In more recent work, Putnam's views on the relation between functional organization and psychological (and also mental) states and events have become more complicated. I make no claims about how the argument of Section II bears on them. Other representatives of the "hypothesis" approach are Gilbert Harman, "Three Levels of Meaning," *The Journal of Philosophy* 65 (1968); "An Introduction to 'Translation and Meaning'," *Words and Objections*, eds. D. Davidson and J. Hintikka (Reidel, 1969), p. 21; and *Thought* (Princeton, 1973), pp. 43-46, 56-65, for example, p. 45: ". . . mental states and processes are to be functionally defined (by a psychological theory). They are constituted by their function or role in the relevant programme"; Jerry Fodor, *The Language of Thought* (New York, 1975), Chapter I; Armstrong, *A Materialist Theory of Mind*, p. 84. An attempt to articulate the common core of the different types of functionalist "account" occurs in Ned Block and Jerry Fodor's "What Psychological States are Not," *Philosophical Review* 81 (1972), p. 173: ". . . functionalism in the broad sense of that doctrine which holds that type identity conditions for psychological states refer only to their relations to inputs, outputs and one another."

10. Often functionalists give mental contents only cursory discussion, if any at all. But claims that a functional account explains intentionality by accounting for all specific intentional states and events in non-intentional, functional language occur in the following: Daniel Dennett, *Content and Consciousness* (London, 1969), Chapter II and *passim*; Harman, *Thought*, for example, p. 60: "To specify the meaning of a sentence used in communication is partly to specify the belief or other mental state expressed; and the representative character of that state is determined by its functional role"; Fodor, *The Language of Thought*, Chapters I and II, for

example, p. 75: "The way that information is stored, computed . . . or otherwise processed by the organism explains its cognitive states and in particular, its propositional attitudes"; Smart, "Further Thoughts on the Identity Theory"; Hartry Field, "Mental Representation," *Erkenntnis* 13 (1978): 9-61. I shall confine discussion to the issue of intentionality. But it seems to me that the individualistic cast of functionalist accounts renders them inadequate in their handling of another major traditional issue about intentional mental states and events—first-person authority.

11. H. P. Grice, "Meaning," *Philosophical Review* 66 (1957):377-88; "Utterer's Meaning, Sentence-Meaning, and Word-Meaning," *Foundations of Language* 4 (1968):225-42; Stephen Schiffer, *Meaning* (Oxford, 1972), cf. especially pp. 13, 50, 63ff; Jonathan Bennett, "The Meaning-Nominalist Strategy," *Foundations of Language* 10 (1974):141-68. Another example of an individualistic theory of meaning is the claim to explicate all kinds of meaning ultimately in psychological terms, and these latter in functionalist terms. See, for example Harman, "Three Levels of Meaning," note 9. This project seems to rest on the functionalist approaches just criticized.

12. Perhaps the first reasonably clear modern statement of the strategy occurs in J. J. C. Smart, "Sensations and Brain Processes," *Philosophical Review* 68 (1959):141-56. This article treats qualitative experiences; but Smart is explicit in applying it to specific intentional states and events in "Further Thoughts on the Identity Theory." Cf. also David Lewis, "An Argument for the Identity Theory," *The Journal of Philosophy* 63 (1966):17-25; "Psychophysical and Theoretical Identifications"; Armstrong, *A Materialist Theory of Mind, passim;* Harman, *Thought*, pp. 42-43; Fodor, *The Language of Thought*, Introduction.

13. The argument is basically Cartesian in style, (cf. *Meditations* II), though the criticism of functionalism, which is essential to its success, is not in any obvious sense Cartesian. (Cf. note 14.) Also the conclusion gives no special support to Cartesian ontology. The terminology of rigidity is derived from Saul Kripke, "Naming and Necessity," *Semantics of Natural Language*, eds., Davidson and Harman (Dordrecht, 1972), though as mentioned above, a notion of rigidity is not essential for the argument. Kripke has done much to clarify the force of the Cartesian sort of argument. He gives such an argument aimed at showing the non-identity of sensations with brain processes. The argument as presented seems to suffer from a failure to criticize materialistic accounts of sensation language and from not indicating clearly how token physical events and token sensation events that are *prima facie* candidates for identification could have occurred independently. For criticism of Kripke's argument, see Fred Feldman, "Kripke on the Identity Theory," *The Journal of Philosophy* 71 (1974):665-76; William G. Lycan, "Kripke and the Materialists," *Ibid.*, pp. 677-89; Richard Boyd, "What Physicalism Does Not Entail," *Readings in the Philosophy of Psychology*, ed. N. Block (forthcoming); Colin McGinn, "Anomalous Monism and Kripke's Cartesian Intuitions," *Analysis* 37 (1977):78-80. It seems to me, however, that these issues are not closed.

14. It is important to note that our argument against functionalist specifications of mentalistic phenomena did not depend on the assumption that no occurrent thought could have a different content from the one it has and be the very same occurrence or event. If it did, the subsequent argument against the identity theory would, in effect, beg the question. The strategy of the latter argument is rather to presuppose an independent argument that undermines non-intentional functionalist specifications of what it is to be *a* thought that (say) sofas are comfortable; then to take as plausible and undefeated the assumption that no occurrent thought could have a different (obliquely non-equivalent) content and be the same occurrence or event; and, finally, to use this assumption with the modal considerations appealed to earlier, to arrive at the non-identity of an occurrent thought event with any event specified by physical theory (the natural sciences) that occurs within the individual.

Perhaps it is worth saying that the metaphorical claim that mental events are identified by their *role* in some "inference-action language game" (to use a phrase of Sellars's) does not provide a plausible ground for rejecting the initial premise of the argument against the identity

theory. For even if one did not reject the "role-game" idea as unsupported metaphor, one could agree with the claim on the understanding that the roles are largely the intentional contents themselves and the same event in *this* sort of "game" could not have a different role. A possible view in the philosophy of mathematics is that numbers are identified by their role in a progression and such roles are essential to their identity. The point of this comparison is just that appeal to the role metaphor, even if accepted, does not settle the question of whether an intentional mental event or state could have had a different content.

15. There are *prima facie* viable philosophical accounts that take sentences (whether tokens or types) as truth bearers. One might hope to extend such accounts to mental contents. On such treatments, contents are not things over and above sentences. They simply *are* sentences interpreted in a certain context, treated in a certain way. Given a different context of linguistic interpretation, the content of the same sentence might be different. One could imagine mental events to be analogous to the sentences on this account. Indeed, some philosophers have thought of intentional mental events as being inner, physical sentence (or symbol) tokens—a sort of brain writing. Here again, there is a picture according to which the same thought event might have had a different content. But here again the question is whether there is any reason to think it is a true picture. There is the prior question of whether sentences can reasonably be treated as contents. (I think sentence types probably can be; but the view has hardly been established, and defending it against sophisticated objections is treacherous.) Even if this question is answered affirmatively, it is far from obvious that the analogy between sentences and contents, on the one hand, and thought events and contents, on the other, is a good one. Sentences (types or tokens) are commonly identified independently of their associated contents (as evidenced by inter- and intra-linguistic ambiguity). It is *relatively* uncontroversial that sentences can be identified by syntactical, morphemic, or perceptual criteria that are in principle specifiable independently of what particular content the sentence has. The philosophical question about sentences and contents is whether discourse about contents can be reasonably interpreted as having an ontology of nothing more than sentences (and intentional agents). The philosophical question about mental events and contents is "What is the nature of the events?" "Regardless of what contents are, could the very same thought event have a different content?" The analogous question for sentences—instead of thought events—has an uncontroversial affirmative answer. Of course, we know that when and where non-intentionally identifiable physical events have contents, the same physical event could have had a different content. But it can hardly be *assumed* for purposes of arguing a position on the mind-body problem that mental events are non-intentionally identifiable physical events.

16. In emphasizing social and pragmatic features in mentalistic attributions, I do not intend to suggest that mental attributions are any the less objective, descriptive, or on the ontological up and up. There are substantial arguments in the literature that might lead one to make such inferences. But my present remarks are free of such implications. Someone might want to insist that from a "purely objective viewpoint" one can describe "the phenomena" equally well in accord with common practice, literally interpreted, or in accord with various reinterpretation strategies. Then our arguments would, perhaps, show only that it is "objectively indeterminate" whether functionalism and the identity theory are true. I would be inclined to question the application of the expressions that are scare-quoted.

17. I am grateful to participants at a pair of talks given at the University of London in the spring of 1978, and to Richard Rorty for discussions earlier. I am also indebted to Robert Adams and Rogers Albritton whose criticisms forced numerous improvements. I appreciatively acknowledge support of the John Simon Guggenheim Foundation.