

## *Is There Mental Representation?*

We tend to take it for granted that there is such a thing as mental representation. People obviously have beliefs, desires, hopes, and fears, and it seems obvious that such mental states involve mental representation. Consider the following mental states: the belief that Paul is in Pittsburgh, the desire that Paul should be in Pittsburgh, the hope that he is in Pittsburgh, and the fear that he is. These are different attitudes *toward* Paul's being in Pittsburgh: belief, hope, desire, fear. They are attitudes *about* Paul—a belief about him, a desire about him, a hope about him, a fear about him. Presumably these attitudes therefore must involve mental representation. For, it seems, to speak of mental representation is simply to make the obvious point that mental states are often attitudes *toward* one or another state of affairs, that they are attitudes *about* one thing or another.

In the same way we assume there are mental representations. For the belief that Paul is in Pittsburgh is itself a mental representation of Paul's being in Pittsburgh; the desire that Paul should be in Pittsburgh is a different representation of the same thing. Similarly for other mental states that are attitudes toward some state of affairs or other, attitudes about one thing or another. All such mental states are, or seem to be, themselves mental representations.

It seems almost as obvious that there is some sort of system to mental representation. We cannot accept as a brute unanalyzable fact about a mental state that it represents Paul's being in Pittsburgh. Such a mental state must surely have something nontrivial in common with other mental states that represent Paul's being in various other places and must also have something nontrivial in com-

mon with other mental states that represent other persons' being in Pittsburgh. There must be some identifiable aspect of the mental state, an aspect it can share with other mental states, an aspect by virtue of which it is a representation of *Paul*. It must also have a different aspect by virtue of which it and other states with that aspect are representations of someone's being in *Pittsburgh*. What a mental state represents must surely be determined by the way in which certain elements are combined in that state, just as what a sentence represents is determined by the way in which certain words are combined in that sentence. Just as a finite stock of words can be combined in an infinite number of possible ways to form an infinite number of sentences, so too, it seems, a finite stock of mental elements must be combinable in a indefinite number of ways to form an indefinite number of mental representations. Mental states must have elements and structure in a way that is analogous to the way in which sentences have elements and structure. There must be, as it were, mental words, mental structure, mental names, mental predicates, mental connectives, and mental quantifiers.

This should not be surprising. We presuppose as much whenever we offer reasons to explain why someone has done something or has a given desire or belief. Suppose someone explains why Ned wants to hold a conference on mechnonetics in Pittsburgh, by citing Ned's desire that Paul should be in Pittsburgh and Ned's belief that Paul will be in Pittsburgh if he, Ned, holds a conference on mechnonetics there. This explanation invokes the commonsense principle that a desire that *P* and a belief that *P, if Q* can lead to a desire that *Q*. As Israel Scheffler has observed, such commonsense principles clearly presuppose that mental states can have logical structures, in this case that a belief can have a conditional structure. The idea that there is some sort of system of mental representation seems therefore required by common sense.

Now, given that there is such a system of mental representation, it can be argued that a natural language like English or German will be part of the system of mental representation possessed by someone who speaks that language. For one thing, there is the experience familiar to second language learners of no longer having to translate between their first language and the language they are learning, when they come, as we sometimes say, to be able to "think in" the new

language. Second, when we learn a new theory of some sort, a new branch of physics or mathematics, learning the language of the theory and learning to think in the new way required by the theory seem impossible to separate. Third, giving arbitrary labels to aspects of the natural environment seems to have an immediate effect on the way we perceive that environment. Fourth, in most conversations we do not plan our remarks ahead of time but, as it were, simply “think out loud.”

I would suggest that these four points are best explained by supposing that language learning involves modifying our system of mental representation so as to incorporate the language being learned. This is not of course to say that all mental representation is in language. Much, perhaps most, is not. This other nonlinguistic representation may be very much like linguistic representation, or alternatively may involve something like the sort of representation that is involved in pictures, maps, or diagrams. I am not sure what to say about this. The main claim I wish to make is that some mental representation is in language.

Now, however, I must confess that I have some doubts about what I have just said. What worries me is not, as you might suppose, the last suggestion that some mental representation is in language. I am instead worried by the more basic claim that there is such a thing as mental representation in the first place. This strikes me as much less obvious than I have been pretending. On the other hand, I am strongly inclined to think that, if the basic claim is correct, then almost certainly some mental representation is in language. But, as I say, I am worried about the basic claim.

Notice, first of all, that not all beliefs, desires, hopes, or fears consist in *explicit* mental representations of what is believed, desired, hoped, or feared. Almost everyone believes that  $104 + 3 = 107$ , but few people believe this in any explicit way. Few people have in their minds, either consciously or unconsciously, an explicit mental representation of this particular sum. Rather, we might be inclined to say, they believe it implicitly. It is, we suppose, clearly implied by things they believe more explicitly.

The example reveals something very important. The fact that someone believes, desires, hopes, or fears that something is the case does not by itself entail that he or she has an explicit mental repre-

sentation of that being the case. So the claim that there is mental representation does not follow trivially from the existence of beliefs, desires, hopes, and fears. How then is that claim to be supported? That is my worry.

What does it mean to say that a given person only implicitly believes that  $104 + 3 = 107$ ? Presumably, to say that he does not explicitly believe this is to deny that his belief that  $104 + 3 = 107$  is a real state of him, a state that is part of the causal order, a state that might play a role in getting him to believe other things or act in various ways. To say that he believes implicitly that  $104 + 3 = 107$  is to say rather that this belief is merely implicit, perhaps implicit in things he believes explicitly so that it is obviously implied by his explicit beliefs and easily “reachable” by him from his explicit beliefs without any real thought.

But then we must at least consider the possibility that all alleged mental representation might be somehow merely implicit. Do mental representations ever play a real causal role? Are beliefs, desires, hopes, and fears ever part of the real causal order or are they always to some degree implicit in something else in approximately the way in which an average person’s belief that  $104 + 3 = 107$  is thought to be a merely implicit belief?

One reason we take an average person to believe that  $104 + 3 = 107$  is that, if we ask him what the sum of  $104 + 3$  he will say that it is 107, and he will say this immediately without laborious calculation. He can give this answer immediately even though he has no prior explicit mental representation of this sum. But then maybe all beliefs are like this—involving no explicit mental representation but only the ability to respond appropriately if one is asked certain questions or if one is put in certain other situations.

The question whether a belief is explicit, in this sense, is of course not the same as the question whether it is an “occurrent belief,” i.e., a belief of which one is now consciously aware. A defender of mental representation might well maintain that one has a large number of explicit beliefs that are part of the causal order but that for the most part never become conscious or “occurrent.” On the other hand, I am worried whether any beliefs, including “occurrent” conscious beliefs, are ever explicit in this sense, being themselves part

of the causal order and not merely implicit in other things that are part of the causal order.

The issue here is that raised by behaviorism. Do mental states ever play an explanatory role? Are they part of the causal order? Or are they merely part of an interpretation we make, something that is only implicit in the causal order in the way that the average person's belief that  $104 + 3 = 107$  is merely implicit? What is at issue here is not the existence of mental states. Of course mental states exist; the issue is whether they ever exist explicitly rather than merely implicitly. Nor is there a serious issue here of verbal analysis. The issue is not whether mental terminology can be translated into some combination of behavioral and neurophysiological terminology. Whether or not such translation is possible, the challenge of behaviorism remains: What reason is there to suppose that mental states are real parts of the causal order? In other words, what reason is there to suppose that mental representation is ever anything more than implicit representation?

Now it might be argued that implicit representation somehow depends on explicit mental representation. I am inclined to think it does, but I am not sure how the argument is to be put. One possibility would be to argue like this:

There is implicit mental representation only if there is a disposition to form a corresponding explicit mental representation under certain conditions. We can ascribe an implicit belief that  $104 + 3 = 107$  to the average person because the average person is disposed to believe this explicitly if asked, "What is  $104 + 3$ ?"

This argument is hardly compelling, however. Why must we suppose that the relevant disposition is a disposition to form an explicit *mental* representation? Consider the following analogy. Someone might suppose that linguistic representation is sometimes implicit in mental representation, without having to suppose that mental representation is ever explicit representation in language. This person might agree that there is implicit linguistic representation only if there is some sort of disposition to produce an explicit linguistic representation. But this does not have to be a disposition to produce an explicit *mental* representation in language. It could be (this person might suppose) a disposition to produce an explicit linguistic representation out loud or in writing.

Objection: sometimes one simply says something silently to oneself. Won't that have to count as an explicit mental representation in language? Not necessarily. It might be argued that this is a case in which one produces an explicit mental but nonlinguistic representation *of* a nonmental linguistic representation.

I am not endorsing this position, since I am inclined to believe there is mental representation in language. But I think this other position is a possible position to take, one that is in no way incoherent. If so, the claim that there is implicit mental representation of a certain sort does not logically entail the claim that there is a disposition to form an explicit mental representation of that sort. But then, the argument as I have stated it must be rejected since it assumed that there is such an entailment.

Perhaps we could make do with a weaker assumption. Consider this argument.

There is implicit mental representation only if there is a disposition to form a corresponding explicit representation. Sometimes this will be a disposition to form an outer nonmental representation, such as an utterance in a language, but not always. There are thoughts that cannot be expressed in language. Furthermore, animals and human infants have beliefs, desires, hopes, and fears without possessing any language. They must therefore be disposed to form corresponding explicit representations that are not in language. Such representations can only be explicit mental representations. So there are after all some explicit mental representations.

I am not sure this is conclusive. Precisely because animals and human infants do not have language, some philosophers and psychologists deny that they have beliefs, desires, fears, and so forth, in the full sense of these terms and hold that implicit mental representations can be ascribed to such creatures only to the extent that they are disposed to act in ways that can be interpreted as behavioral representations of one or another state of affairs.

A somewhat different and possibly more compelling argument might go like this:

There is implicit representation only where there is something that is implicit in—implied by—the content something has explicitly. So there is implicit mental representation only if there is explicit mental representation.

As an a priori argument, this will not do, since behaviorists are not

committing a logical error in rejecting the argument. But the argument seems plausible to me if interpreted as saying that:

The best account we have at present of implicit mental representation assumes that there is explicit mental representation and takes something to be implicitly mentally represented only if it is fairly obviously implied by what is explicitly mentally represented.

This seems plausible to me even though I am not altogether sure what the alleged “best account” is supposed to be. I do not know of any very specific plausible account of mental representation. There is, of course, the ordinary commonsensical idea that people act as they do because of their beliefs, desires, hopes, fears, and so forth. According to common sense, such mental states do sometimes play a part in the causal order, which is sufficient if true for them to count as explicit mental representations. And common sense is sometimes willing to speak of belief in a case in which the belief is merely implicit in what someone explicitly believes. So I suppose there is some sort of commonsensical account of implicit and explicit mental representation. That is, I guess, our “best account.” I do not know of any clearly better account.

True, there are a number of interesting proposals by psychologists that postulate mental representations of various sorts. But where these proposals clearly go beyond common sense, they seem to me so speculative that I hesitate to include any of them as part of our “best current account.” Maybe I am being overly cautious here. If so, I hope someone will tell me why.

Of course, if I am right about common sense being the best account, the situation is unstable. Common sense does not seem to yield much in the way of an explicit theory and, if psychologists do develop a fairly specific theory to account for various aspects of mental life, it is not obvious to me that their theory will have to be in any way an extension of common sense, nor is it obvious that such a theory will have to include an important role for mental representations. But, at least for the time being, our best account does assume that explicit mental representations are part of the causal order, which gives us some reason to *believe* that explicit mental representations are part of the causal order.

There is the same weak sort of reason to think that mental representations have logical structure. For, as I have already mentioned,

commonsensical explanations that appeal to a person's reasons presuppose that beliefs and other mental representations sometimes have, for example, the structure of a conditional.

And there is, I think, almost the same weak sort of reason to think that "one's system of mental representation," (if I may call it that), incorporates one's natural language. True, common sense does not seem to be committed to this assumption. But the four points mentioned earlier are best explained, I think, by supposing that language learning modifies one's system of mental representation by incorporating the language being learned. To repeat those points, they are (1) that language learners come to be able to "think in" the new language, (2) that learning a new theoretical terminology seems to involve learning new methods of thinking, (3) that giving arbitrary labels to things helps to structure one's perceptions, and (4) that conversation is often unplanned and involves a certain amount of something it is natural to call "thinking out loud." The explanation here is quite unspecific, so these points provide at best very weak evidence for the claim that some thought is in language. But our evidence that there is any such thing as mental representation at all seems to be also of this extremely weak sort.