Knowing One's Own Mind

Donald Davidson

There is no secret about the nature of the evidence we use to decide what other people think: we observe their acts, read their letters, study their expressions, listen to their words, learn their histories, and note their relations to society. How we are able to assemble such material into a convincing picture of a mind is another matter; we know how to do it without necessarily knowing how we do it. Sometimes I learn what I believe in much the same way someone else does, by noticing what I say and do. There may be times when this is my only access to my own thoughts. According to Graham Wallas:

The little girl had the making of a poet in her who, being told to be sure of her meaning before she spoke, said "How can I know what I think till I see what I say?"

A similar thought was expressed by Robert M. Young: "I would say that most good painters don't know what they think until they paint it."

Gilbert Ryle was with the poet and the painter all the way in this matter; he stoutly maintained that we know our own minds in exactly the same way we know the minds of others, by observing what we say, do, and paint. Ryle was wrong. It is seldom the case that I need or appeal to evidence or observation in order to find out what I believe; normally I know what I think before I speak or act. Even when I have evidence, I seldom make use of it. I can be wrong about my own thoughts, and so the appeal to what can be publicly determined is not irrelevant. But the possibility that one may be mistaken about one's own thoughts cannot defeat the overriding presumption that a person knows what he or she believes; in general, the belief that one has a thought is enough to justify that belief. But though this is true, and even obvious to most of us, the fact has, so far as I can see, no easy explanation. While it is clear enough, at least in outline, what we have to go on in trying to fathom the thoughts of others, it is obscure why, in our own case, we can so often know what we think without appeal to evidence or recourse to observation.

Because we usually know what we believe (and desire and doubt and intend) without needing or using evidence (even when it is available), our sincere avowals concerning our present states of mind are not subject to the failings of conclusions based on evidence. Thus sincere first person present-tense claims about thoughts, while neither infallible nor incorrigible, have an authority no second or third person claim, or first person other-tense claim, can have. To recognize this fact is not, however, to explain it.

Since Wittgenstein it has become routine to try to relieve worries about "our knowledge of
other minds” by remarking that it is an essential aspect of our use of certain mental predicates that we apply them to others on the basis of behavioral evidence but to ourselves without benefit of such aid. The remark is true, and when properly elaborated, it ought to answer someone who wonders how we can know the minds of others. But as a response to the skeptic, Wittgenstein’s insight (if it is Wittgenstein’s) should give little satisfaction. For, first, it is a strange idea that claims made without evidential or observational support should be favored over claims with such support. Of course, if evidence is not cited in support of a claim, the claim cannot be impugned by questioning the truth or relevance of the evidence. But these points hardly suffice to suggest that in general claims without evidential support are more trustworthy than those with. The second, and chief, difficulty is this. One would normally say that what counts as evidence for the application of a concept helps define the concept, or at least places constraints on its identification. If two concepts regularly depend for their application on different criteria or ranges of evidential support, they must be different concepts. So if what is apparently the same expression is sometimes correctly employed on the basis of a certain range of evidential support and sometimes on the basis of another range of evidential support (or none), the obvious conclusion would seem to be that the expression is ambiguous. Why then should we suppose that a predicate like “believe” that Ras Dahan is the highest mountain in Ethiopia, which is applied sometimes on the basis of behavioral evidence and sometimes not, is unambiguous? If it is ambiguous, then there is no reason to suppose it has the same meaning when applied to oneself that it has when applied to another. If we grant (as we should) that the necessarily public and interpersonal character of language guarantees that we often correctly apply these predicates to others, and that therefore we often do know what others think, then the question must be raised what grounds each of us has for thinking he knows what (in the same sense) he thinks. The Wittgensteinian style of answer may solve the problem of other minds, but it creates a corresponding problem about knowledge of one’s own mind. The correspondence is not quite complete, however. The original problem of other minds invoked the question how one knows others have minds at all. The problem we now face must be put this way: I know what to look for in attributing thoughts to others. Using quite different criteria (or none), I apply the same predicates to myself; so the skeptical question arises why I should think it is thoughts I am attributing to myself. But since the evidence I use in the case of others is open to the public, there is no reason why I shouldn’t attribute thoughts to myself in the same way I do to others, in the mode of Graham Wallas, Robert Motherwell, and Gilbert Ryle. In other words, I don’t, but I could, treat my own mental states in the same way I do those of others. No such strategy is available to someone who seeks the same sort of authority with respect to the thoughts of others as he apparently has in dealing with his own thoughts. So the asymmetry between the cases remains a problem, and it is first person authority that creates the problem.

I have suggested an answer to this problem in another paper. In that paper I argued that attention to how we attribute thoughts and meanings to others would explain first person authority without invoking skeptical doubts. In recent years, however, some of the very facts about the attribution of attitudes on which it is relied to defend first person authority have been employed to attack that authority: it has been argued, on what are thought to be new grounds, that while the methods of the third person interpreter determine what we usually deem to be the contents of an agent’s mind, the contents so determined may be unknown to the agent. In the present essay I consider some of these arguments, and urge that they do not constitute a genuine threat to first person authority. The explanation I offered in my earlier paper of the symmetry between first and other-person attributions of attitudes seems to me if anything to be strengthened by the new considerations, or those of them that seem valid.

It should be stressed again that the problem I am concerned with does not require that our beliefs about our own contemporary states of mind be infallible or incorrigible. We can and do make mistakes about what we believe, desire, approve, and intend; there is also the possibility of self-deceit. But such cases, though not infrequent, are not and could not be standard; I do not argue for this now, but take it as one of the facts to be explained.

Setting aside, then, self-deception and other anomalous or borderline phenomena, the question is whether we can, without irrationality, inconsistency, or confusion, simply and straightforwardly think we have a belief we do not have, or think we do not have a belief we do have. A number of philosophers and philosophically minded psychologists have recently entertained views that entail or suggest that this could easily happen — indeed, that it must happen all the time.

The threat was there in Russell’s idea of propositions that could be known to be true even though they contained “ingredients” with which the mind of the knower was not acquainted; and as the study of the de re attitudes evolved the peril grew more acute.

But it was Hilary Putnam who pulled the plug. Consider Putnam’s 1975 argument to show that meanings, as he put it, “just ain’t in the head.” Putnam argues persuasively that what words mean depends on more than “what is in the head.” He tells a number of stories the moral of which is that aspects of the natural history of how someone learned the use of a word necessarily make a difference to what the word means. It seems to follow that two people might be in physically identical states, and yet mean different things by the same words.

The consequences are far-reaching. For if people can (usually) express their thoughts correctly in words, then their thoughts — their beliefs, desires, intentions, hopes, expectations — also must in part be identified by events and objects outside the person. If meanings aren’t in the head, then neither, it would seem, are beliefs and desires and the rest.

Since some of you may be a little weary of Putnam’s doppelganger on Twin Earth, let me tell you my own science fiction story — if that is what it is. My story avoids some irrelevant difficulties in Putnam’s story, though it introduces some new problems of its own. I’ll come back to Earth, and Twin Earth, a little later.) Suppose lightning strikes a dead tree in a swamp; I am standing nearby. My body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, The Swampman, moves exactly as I did; according to its nature it departs the swamp, encounters and seems to recognize my friends, and appears to return their greetings in English. It moves into my house and seems to write articles on radical interpretation. No one can tell the difference.

But there is a difference. My replica can’t recognize my friends; it can’t recognize anything, since it never cognized anything in the first place. It can’t know my friends’ names (though of course it seems to), it can’t remember my house. It can’t mean what I do by the word “house,” for example, since the sound “house” it makes was not learned in a context that would give it the right meaning — or any meaning at all. Indeed, I don’t see how my replica can be said to mean anything by the sounds it makes, nor to have any thoughts.

Putnam might not go along with this last claim, for he says that if two people (or objects) are in relevantly similar physical states, it is “absurd” to think their psychological states are “one bit different.” It would be a mistake to be sure that Putnam and I disagree on this point, however, since it is not yet clear how the phrase “psychological state” is being used.

Putnam holds that many philosophers have wrongly assumed that psychological states like belief and knowing the meaning of a word are both (I) “inner” in the sense that they do not presuppose the existence of any individual other than the subject to whom the state is ascribed, and (II) that these are the very states which we normally identify and individuate as we do beliefs and the other propositional attitudes. Since we normally identify and individuate mental states and meanings in terms partly of relations to objects and events other than the subject, Putnam believes (I) and (II) come apart in his opinion, no states can satisfy both conditions.

Putnam calls psychological states satisfying condition (I) “narrow.” He thinks of such states as solipsistic, and associates them with Descartes’ view of the mental. Putnam may consider these states to be the only “true” psychological states; in much of his paper he omits the qualifier “narrow,” despite the fact that narrow psychological states (so called) do not correspond to the propositional attitudes as normally identified. Not everyone has been persuaded that there is an intelligible distinction to be drawn between narrow (or inner, or Cartesian, or individualistic — all these terms are current) psychological states
and psychological states identified (if any are) in terms of external facts (social or otherwise). Thus John Searle has claimed that our ordinary propositional attitudes satisfy condition (I), and so there is no need of states satisfying condition (II), while Tyler Burge has denied that there are, in any interesting sense, propositional attitudes that satisfy condition (I).1 But there seems to be universal agreement that no states satisfy both conditions.

The thesis of this essay is that there is no reason to suppose that ordinary mental states do not satisfy both conditions (I) and (II): I think such states are "inner," in the sense of being identical with states of the body, and so identifiable without reference to objects or events outside the body; they are at the same time "nonindividualistic" in the sense that they can be, and usually are, identified in part by their causal relations to events and objects outside the subject whose states they are. A corollary of this thesis will turn out to be that contrary to what is often assumed, first person authority can without contradiction apply to states that are regularly identified by their relations to events and objects outside the person.

I begin with the corollary. Why is it natural to assume that states that satisfy condition (II) may not be known to the person who is in those states?

Now I must talk about Putnam’s Twin Earth. He asks us to imagine two people exactly alike physically and (therefore) alike with respect to all "narrow" psychological states. One of the two people, an inhabitant of Earth, learned to use the word water by being shown water, and hearing about it, etc. The other, an inhabitant of Twin Earth, has learned to use the word "water" under conditions not observably different, but the substance to which she has been exposed is not water but a lookalike substance we may call "twater." Under the circumstances, Putnam claims, the first speaker refers to water when she uses the word "water"; her twin refers to twater when she uses the word "water." So we seem to have a case where "narrow" psychological states are identical, and yet the speakers mean different things by the same word.

How about the thoughts of these two speakers? The first says herself to facing a glass of water, "Here’s a glass of water"; the second matters exactly the same sounds to herself when facing a glass of water. Each speaks the truth, since their words mean different things. And since each is sincere, it is natural to suppose they believe different things, the first believing there is a glass of water in front of her, the second believing there is a glass of twater in front of her. But do they know what they believe? If the meanings of their words, and thus the beliefs expressed by using those words, are partly determined by external factors about which the agents are ignorant, their beliefs and meanings are not narrow in Putnam’s sense. There is therefore nothing on the basis of which either speaker can tell which state she is in, for there is no internal or external clue to the difference available. We ought, it seems, to conclude that neither speaker knows what she means or thinks. The conclusion has been drawn explicitly by a number of philosophers, among them Putnam. Putnam declares that he "totally abandons the idea that if there is a difference in meaning ... then there must be some difference in our concepts (or in our psychological state). What determines meaning and extension "is not, in general, fully known to the speaker." Here "psychological state" means narrow psychological state, and it is assumed that only such states are "fully known." Jerry Fodor believes that ordinary propositional attitudes are (pretty nearly) "in the head," but he agrees with Putnam that if propositional attitudes were partly identified by factors outside the agent, they would not be in the head, and would not necessarily be known to the agent. John Searle also, though his reasons are not Fodor’s, holds that meanings are in the head ("there is nowhere else for them to be"), but seems to accept the inference that if this were not the case, first person authority would be lost. Perhaps the plainest statement of the position appears in Andrew Woodfield’s introduction to a book of essays on the objects of thought. Referring to the claim that the contents of the mind are often determined by facts external to and perhaps unknown to the person whose mind it is, he says:

Because the external relation is not determined subjectively, the subject is not authoritative about that. A third person might well be in a better position than the subject to know which object the subject is thinking about, hence be better placed to know which thought it was.12

Those who accept the thesis that the contents of propositional attitudes are partly determined in terms of external factors seem to have a problem similar to the problem of the skeptic who finds we may be altogether mistaken about the "outside" world. In the present case, ordinary skepticism of the senses is avoided by supposing the world itself more or less correctly determines the contents of thoughts about the world. (The speaker who thinks it is water is probably right, for he learned the use of the word "water" in a watery environment; the speaker who thinks twater is probably right, for he learned the word "water" in a twater environment.) But skepticism is not defeated; it is only displaced onto knowledge of our own minds. Our ordinary beliefs about the external world are (on this view) directed onto the world, but we don’t know what we believe.

There is, of course, a difference between water and twater, and it can be discovered by normal means, whether it is discovered or not. So a person might find out what he believes by discovering the difference between water and twater, and finding out enough about his own relations both to determine which one his talk and beliefs are about. The skeptical conclusion we seem to have reached concerns the extent of first person authority: it is far more limited than we supposed. Our beliefs about the world are mostly true, but we may easily be wrong about what we think. It is a transposed image of Cartesian skepticism.

Those who hold that the contents of our thoughts and the meanings of our words are often fixed by factors of which we are ignorant have not been much concerned with the apparent consequence of their views which I have been emphasizing. They have, of course, realized that if they were right, the Cartesian idea that the one thing we can be certain of is the contents of our own minds, and the Fregean notion of meanings fully "grasped," must be wrong. But they have not made much of an attempt, so far as I know, to resolve the seeming conflict between their views and the strong intuition that first person authority exists.

One reason for the lack of concern may be that some seem to see the problem as confined to a fairly limited range of cases, cases where concepts or words latch on to objects that are picked out or referred to using proper names, indexicals, and words for natural kinds. Others, though, argue that the ties between language and thought on the one hand and external affairs on the other are so pervasive that no aspect of thought as usually conceived is untouched. In this vein Daniel Dennett remarks that "one must be firmly informed about, intimately connected with, the world at large, its occupants and properties, in order to be said with any propriety to have beliefs." He goes on to claim that the identification of all beliefs is infected by the outside, nonsubjective factors that are recognized to operate in the sort of case we have been discussing. Burge also emphasizes the extent to which our beliefs are affected by external factors, though for reasons he does not explain, he apparently does not view this as a threat to first person authority.

The subject has taken a disquieting turn. At one time behaviorism was invoked to show how it was possible for one person to know what was in another’s mind; behaviorism was then rejected in part because it could not explain one of the most obvious aspects of mental states: the fact that they are in general known to the person who has them without appeal to behavioristic evidence. The recent fashion, though not strictly behavioristic, once more identifies mental states partly in terms of social and other external factors, thus making them to that extent publicly discoverable. But at the same time it reinstates the problem of accounting for first person authority.

Those who are convinced of the external dimension of the contents of thoughts as ordinarily identified and individuated have reacted in different ways. One response has been to make a distinction between the contents of the mind as subjectively and internally determined, on the one hand, and ordinary beliefs, desires, and intentions, as we normally attribute them on the basis of social and other outward connections, on the other. This is clearly the trend of Putnam’s argument (although the word "water" has different meanings, and is used to express different beliefs when it is used to refer to water and to
water, people using the word for these different purposes may be in "the same psychological state". Jerry Fodor accepts the distinction for certain purposes, but argues that psychology should adopt the stance of "methodological solipsism" (Putnam's phrase) — that is, it should deal exclusively with inner states, the truly subjective psychological states which owe nothing to their relations to the outside world. 19 Stephen Stich makes essentially the same distinction, but draws a sterner moral: where Fodor thinks we merely need to tinker a bit with propositional attitudes as usually conceived to separate out the purely subjective element, Stich holds that psychological states as we now think of them belong to a crude and confused "folk psychology" which must be replaced by a yet to be invented "cognitive science." The subtitle of his recent book is "The Con Against Belief." 20

Clearly those who draw such a distinction have ensured that the problem of first person authority, at least as I have posed it, cannot be solved. For the problem I have set is how to explain the asymmetry between the way in which a person knows about his contemporary mental states and the way in which others know about them. The mental states in question are beliefs, desires, intentions, and so on, as ordinarily conceived. Those who accept something like Putnam's distinction do not even try to explain first person authority with respect to these states; if there is first person authority at all, it attaches to quite different states. In Stich's case, it is not obvious that it can attach to anything.

I think Putnam, Burge, Dennett, Fodor, Stich, and others are right in calling attention to the fact that ordinary mental states, at least the propositional attitudes, are partly identified by relations to society and the rest of the environment, relations which may in some respects not be known to the person in those states. They are also right, in my opinion, that for this reason (if for no others), the concepts of "folk psychology" cannot be incorporated into a coherent and comprehensive system of laws of the sort for which physics strives. These concepts are part of a commonsense theory for describing, interpreting, and explaining human behavior which is a bit freestyle, but (so I think) indispensable. I can imagine a science concerned with people and pured of "folk psychology," but I cannot think in what its interest would consist. This is not, however, the topic of this essay.

I am here concerned with the puzzling discovery that we apparently do not know what we think — at least in the way we think we do. This is a real puzzle if, like me, you believe it is true that external factors partly determine the contents of thoughts, and also believe that in general we do know, and in a way others don't do, what we think. The problem arises because admitting the identifying and individuating role of external factors seems to lead to the conclusion that our thoughts may not be known to us.

But does this conclusion follow? The answer depends, I believe, on the way in which one thinks the identification of mental contents depends on external factors.

The conclusion does follow, for example, for any theory which holds that propositional attitudes are identified by objects (such as propositions, tokens of propositions, or representations) which are in or "before" the mind, and which contain or incorporate (as "ingredients") objects or events outside the agent; for it is obvious that everyone is ignorant of endless features of every external object. That the conclusion follows from these assumptions is generally conceded. 21 However, for reasons I shall mention below, I reject the assumptions on which the conclusion is in this case based.

Tyler Burge has suggested that there is another way in which external factors enter into the determination of the contents of speech and thought. One of his "thought experiments" happens pretty well to fit me. Until recently I believed arthritis was an inflammation of the joints caused by calcium deposits. I did not know that any inflammation of the joints, for example gout, also counted as arthritis. So when a doctor told me (falsely as it turned out) that I had gout, I believed I had gout but I did not believe I had arthritis. At this point Burge asks us to imagine a world in which I was physically the same but in which the word "arthritis" happened actually to apply only to inflammation of the joints caused by calcium deposits. Then the sentence "Gout is not a form of arthritis" would have been true, not false, and the belief that I expressed by this sentence would not have been the false belief that gout is not a form of arthritis but a true belief about some disease other than arthritis. Yet in the imagined world all my physical states, my "internal qualitative experiences", my behavior and dispositions to behave, are the same as they are in this world. My belief would have changed, but I would have no reason to suppose that it had, and so could not be said to know what I believed.

Burge stresses the fact that his argument depends on the possibility of someone's having a propositional attitude despite an incomplete mastery of some notion in its content ... 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 misapparition. 22

It seems to follow that if Burge is right, whenever a person is wrong confused, or partially misinformed about the meaning of a word, he is wrong, confused, or partially misinformed about any of his beliefs that is (or would be) expressed by using that word. Since such "partial understanding" is "common or even normal in the case of a large number of expressions in our vocabularies" according to Burge, it must be equally common or normal for us to be wrong about what we believe (and, of course, fear, hope for, wish were the case, doubt, and so on).

Burge apparently accepts this conclusion; at least so I interpret his denial that "full understanding of a content is in general a necessary condition for believing the content." He explicitly rejects "the old model according to which a person must be directly acquainted with, or must immediately apprehend, the contents of his thoughts ... 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." 23

I am uncertain how to understand these claims, since I am uncertain how seriously to take the talk of "direct acquaintance" with, and of "immediately apprehending," a content. But in any case I am convinced that if what we mean and think is determined by the linguistic habits of those around us in the way Burge believes they are, then first person authority is very seriously compromised. Since the degree and character of the compromise seem to me incompatible with what we know about the kind of knowledge we have of our own minds, I must reject some premise of Burge's. I agree that what I mean and think is not "fixed" (exclusively) by what goes on in me, so what I must reject is Burge's account of how social and other external factors control the contents of a person's mind.

For a number of reasons, I am inclined to discount the importance of the features of our attributions of attitudes to which Burge points. Suppose that I, who think the word "arthritis" applies to inflammation of the joints only if caused by calcium deposits, and my friend Arthur, who knows better, both sincerely utter to Smith the words "Carl has arthritis." According to Burge, if other things are more or less equal (Arthur and I are both generally competent speakers of English, both have often applied the word "arthritis" to genuine cases of arthritis, etc.) then our words on this occasion mean the same thing. Arthur and I mean the same thing by our words, and we express the same belief. My error about the dictionary meaning of the word (or about what arthritis is) makes no difference to what I meant or thought on this occasion. Burge's evidence for this claim seems to rest on his conviction that this is what anyone (unspoiled by philosophy) would report about Arthur and me. I doubt that Burge is right about this, but even if he is, I don't think it proves his claim. Ordinary attributions of meanings and attitudes rest on vast and vague assumptions about what is and is not shared (linguistically and otherwise) by the addressee, the person to whom the attribution is made, and the addressee's intended audience. When some of these assumptions prove false, we may alter the words we use to make the report, often in substantial ways. When nothing much hinges on it, we do not change the way we take someone at his word, even if this does not quite reflect some aspect of the speaker's thought or meaning. But this is not because we are bound (outside of a law court, anyway) to be legalistic about it. And often we aren't. If Smith (unspoiled by philosophy) reports to still another party (perhaps a distant doctor attempting a diagnosis on the basis of a telephone report) that Arthur and I both have said, and believe, that Carl has arthritis, he may actively mislead his hearer. If this danger were to arise, Smith, alert to the facts, would not simply say "Arthur and Davidson both believe Carl has arthritis"; he would add something like, "But Davidson thinks arthritis must be
caused by calcium deposits." The need to make this addition I take to show that the simple attribution was not quite right; there was a relevant difference in the thoughts Arthur and I expressed when we said "Carl has arthritis." Burge does not have to be budged by this argument, of course, since he can insist that the report is literally correct, but could, like any reporter, be misleading. I think, on the other hand, that this reply would overlook the extent to which the contents of one belief necessarily depend on the contents of others. Thoughts are not independent atoms, and so there can be no simple, rigid, rule for the correct attribution of a single thought.18

Though I reject Burge’s insistence that we are bound to give a person’s words the meaning they have in his linguistic community, and to interpret his propositional attitudes on the same basis, I think there is a somewhat different, but very important, sense in which social factors do control what a speaker can mean by his words. If a speaker wishes to be understood, he must intend his words to be interpreted in a certain way, and so must intend to provide his interlocution with the clues they need to arrive at the intended interpretation. This holds whether the hearer is sophisticated in the use of a language the speaker knows or is the learner of a first language. It is the requirement of learnability, interpretability, that provides the irreducible social factor, and that shows why someone can’t mean something by his words that can’t be correctly deciphered by another. (Burge seems to make this point himself in a later paper.)19

Now I would like to return to Putnam’s Twin Earth example, which does not depend on the idea that social linguistic usage dictates (under more or less standard conditions) what speakers mean by their words, nor, of course, what their (narrow) psychological states are. I am, as I said, persuaded that Putnam is right; what our words mean is fixed in part by the circumstances in which we learned, and used, the words. Putnam’s single example (water) is not enough, perhaps, to nail down this point, since it is possible to insist that “water” doesn’t apply just to stuff with the same molecular structure as water but also to stuff enough like water in structure to be odorless, potable, to support swimming and sailing, etc. (I realize that this remark, like many others in this piece, may show that I don’t know

and which I hold whether or not others do, since I think this much “externalism” is required to explain how language can be learned, and how words and attitudes can be identified by an interpreter.

Why does Putnam think that if the reference of a word is (sometimes) fixed by the natural history of how the word was acquired, a user of the word may lose first person authority? Putnam claims (correctly, in my view) that two people can be in all relevant physical (chemical, physiological, etc.) respects the same and yet mean different things by their words and have different propositional attitudes (as these are normally identified). The differences are due to environmental differences about which the two agents may, in some respects, be ignorant. Why, under these circumstances, should we suppose these agents may not know what they mean and think? Talking with them will not easily show this. As we have noted, each, when faced with a glass of water or twat water says honestly, “Here’s a glass of water.” If they are in their home environments, each is right; if they have switched earths, each is wrong. If we ask each one what he means by the word “water,” he gives the right answer, using the same words, of course. If we ask each one what he believes, he gives the right answer. These answers are right because though verbally identical, they must be interpreted differently. And what is it that they do not know (in the usual authoritative way) about their states? As we have seen, Putnam distinguishes the states we have just been discussing from “narrow” psychological states which do not presuppose the existence of any individual other than the subject in that state. We may now start to wonder why Putnam is interested in narrow psychological states. Part of the answer is, of course, that it is these states that he thinks have the “Cartesian” property of being known in a special way by the person who is in them. (The other part of the answer has to do with constructing a “scientific psychology”; this does not concern us here.)

The reasoning depends, I think, on two largely unquestioned assumptions. These are:

1. If a thought is identified by a relation to something outside the head, it isn’t wholly in the head. (It ain’t in the head.)

2. If a thought isn’t wholly in the head, it can’t be “grasped” by the mind in the way required by first person authority.

That this is Putnam’s reasoning is suggested by his claim that if two heads are the same, narrow psychological states must be the same. Thus if we suppose two people are “molecule for molecule” the same (“in the sense in which two neckties can be ‘identical’”; you may add, if you wish, that each of the two people “thinks the same verbalized thoughts . . . , has the same sense data, the same dispositions, etc.”), then “it is absurd to think [one] psychological state is one bit different from” the other. These are, of course, narrow psychological states, not the ones we normally attribute, which ain’t in the head.21

It is not easy to say in exactly what way the verbalized thoughts, sense data, and dispositions can be identical without reverting to the neckties, so let us revert. Then the idea is this: the narrow psychological states of two people are identical when their physical states cannot be distinguished. There would be no point in disputing this, since narrow psychological states are Putnam’s to define; what I wish to question is assumption (1) above which led to the conclusion that ordinary propositional attitudes aren’t in the head, and that therefore first person authority doesn’t apply to them.

It should be clear that it doesn’t follow, simply from the fact that meanings are identified in part by relations to objects outside the head, that meanings aren’t in the head. To suppose this would be as bad as to argue that because my being sunburned presupposes the existence of the sun, my sunburn ain’t a condition of my skin. My sunburned skin may be indistinguishable from someone else’s skin that achieved its burn by other means (our skins may be identical in “the necktie sense”); yet one of us is really sunburned and the other not. This is enough to show that an appreciation of the external factors that enter into our common ways of identifying mental states does not discredit an identity theory of the mental and the physical. Andrew Woodfield seems to think it does. He writes:

No de re state about an object that is external to the person’s brain can possibly be identical with
a state of that brain, since no brain state presupposes the existence of an external object.\textsuperscript{25} Individual states and events don’t conceptually presuppose anything in themselves; some of their descriptions may, however. My paternal grandfather didn’t presuppose me, but if someone can be described as my paternal grandfather, several people besides my grandfather, including me, must exist.

Burge may make a similar mistake in the following passage:

\[\text{no occurrence of a thought … could have a different content and be the very same token event …} \text{[Then … 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 … b need not be affected by counterfactual differences [that do not change the contents of the thought event]. Thus … b [the physical event] is not identical with the subject’s occurrent thought.}\text{\textsuperscript{26}}

Burge does not claim to have established the premise of this argument, and so not its conclusion. But he holds that the denial of the premise is “intuitively very implausible”. He goes on, “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 imaginations are possible fact or just philosophical fancy is a separate question.” It is because he thinks the denial of the premise to be very improbable that he holds that “materialist identity theories” are themselves “rendered implausible by the non-individualistic thought experiments.”

I accept Burge’s premise; I think its denial not merely implausible but absurd. If two mental events have different contents they are surely different events. What I take Burge’s and Putnam’s imagined cases to show (and what I think The Swampman example shows more directly) is that people who are in all relevant physical respects similar (or “identical” in the necktie sense) can differ in what they mean or think, just as they can differ in being grandfathers or being sunburned. But of course there is something different about them, even in the physical world; their causal histories are different.

I conclude that the mere fact that ordinary mental states and events are individuated in terms of relations to the outside world has no tendency to discredit mental–physical identity theories as such. In conjunction with a number of further (plausible) assumptions, the “externalism” of certain mental states and events can be used, I think, to discredit type–type identity theories; but if anything it supports token–token identity theories. (I see no good reason for calling all identity theories “materialist”; if some mental events are physical events, this makes them no more physical than mental. Identity is a symmetrical relation.)

Putnam and Woodfield are wrong, then, in claiming that it is “absurd” to think two people could be physically identical (in the “necktie” sense) and yet differ in their ordinary psychological states. Burge, unless he is willing to make far stronger play than he has with essentialist assumptions, is wrong in thinking he has shown all identity theories implausible. We are therefore free to hold that people can be in all relevant physical respects identical while differing psychologically: this is in fact the position of “anomalous monism” for which I have argued elsewhere.

One obstacle to non-evidential knowledge of our own ordinary propositional attitudes has now been removed. For if ordinary beliefs and the other attitudes can be “in the head” even though they are identified as the attitudes they are partly in terms of what is not in the head, then the threat to first person authority cannot come simply from the fact that external factors are relevant to the identification of the attitudes.

But an apparent difficulty remains. True, my sunburn, though describable as such only in relation to the sun, is identical with a condition of my skin which can (I assume) be described without reference to such “external” factors. Still, if, as a scientist skilled in all the physical sciences, I have access only to my skin, and am denied knowledge of the history of its condition, then by hypothesis there is no way for me to tell that I am sunburned. Perhaps, then, someone has first person authority with respect to the contents of his mind only as those contents can be described or discovered without reference to external factors. Insofar as the contents are identified in terms of external factors, first person authority necessarily lapses. I can tell by examining my skin what my private or “narrow” condition is, but nothing I can learn in this restricted realm will tell me that I am sunburned. The difference between referring to and thinking of water and referring to and thinking of twater is like the difference between being sunburned and one’s skin being in exactly the same condition through another cause. The semantic difference lies in the outside world, beyond the reach of subjective or sublunar knowledge. So the argument might run.

This analogy, between the limited view of the skin doctor and the tunnel vision of the mind’s eye, is fundamentally flawed. It depends for its appeal on a faulty picture of the mind, a picture which those who have been attacking the subjective character of ordinary psychological states share with those they attack. If we can bring ourselves to give up this picture, first person authority will no longer be seen as a problem; instead, it will turn out that first person authority is dependent on, and explained by, the social and public factors that were supposed to undermine that authority.

There is a picture of the mind which has become so ingrained in our philosophical tradition that it is almost impossible to escape its influence even when its worst faults are recognized and repudiated. In one crude, but familiar, version, it goes like this: the mind is a theater in which the conscious self watches a passing show (the shadows on the wall). The show consists of “appearances,” sense data, qualia, what is given in experience. What appears on the stage are not the ordinary objects of the world that the outer eye registers and the heart loves, but their purported representatives. Whatever we know about the world outside depends on what we can glean from the inner clues.

The difficulty that has been apparent from the start with this description of the mental is to see how it is possible to beat a track from the inside to the outside. Another conspicuous, though perhaps less appreciated, difficulty is to locate the self in the picture. For the self seems on the one hand to include theater, stage, actors, and audience; on the other hand, what is known and registered pertains to the audience alone. This second problem could be as well stated as the problem of the location of the objects of the mind: are they in the mind, or simply viewed by it?

I am not now concerned with such (now largely disavowed) objects of the mind as sense-data, but with their judgmental cousins, the supposed objects of the propositional attitudes, whether thought of as propositions, tokens of propositions, representations, or fragments of mentalese. The central idea I wish to attack is that these are entities that the mind can “entertain,” “grasp,” “have before it,” or be “acquainted” with. (These metaphors are probably instructive: voyeurs merely want to have representations before the mind’s eye, while the more aggressive grasp them; the English may be merely acquainted with the contents of the mind, while more friendly types will actually entertain them.)

It is easy to see how the discovery that external facts enter into the individuation of states of mind disturbs the picture of the mind I have been describing. For it to be in a state of mind is for the mind to be in some relation like grasping to an object, then whatever helps determine what object it is must equally be grasped if the mind is to know what state it is in. This is particularly evident if an external object is an “ingredient” in the object before the mind. But in either case, the person who is in the state of mind may not know what state of mind he is in.

It is at this point that the concept of the subjective—of a state of mind—seems to come apart. On the one hand, there are the true inner states, with respect to which the mind retains its authority; on the other hand there are the ordinary states of belief, desire, intention, and meaning, which are polluted by their necessary connections with the social and public world.

In analogy, there is the problem of the sunburn expert who cannot tell by inspecting the skin whether it is a case of sunburn or merely an identical condition with another cause. We can solve the sunburn problem by distinguishing between sunburn and sunniburn; sunniburn is just like sunburn except that the sun need not be involved. The expert can spot a case of sunniburn just by looking, but not a case of sunburn. This solution works because skin conditions,
Unlike objects of the mind, are not required to be such that there be a special someone who can tell, just by looking, whether or not the condition obtains.

The solution in the case of mental states is different, and simpler: it is to get rid of the metaphor of objects before the mind. Most of us long ago gave up the idea of perceptions, sense data, the flow of experience, as things "given" to the mind; we should treat propositional objects in the same way. Of course people have beliefs, wishes, doubts, and so forth; but to allow this is not to suggest that beliefs, wishes and doubts are entities in or before the mind, or that being in such states requires there to be corresponding mental objects.

This has been said before, in various tones of voice, but for different reasons. Ontological scruples, for example, are not part of my interest. We will always need an infinite supply of objects to help describe and identify attitudes like belief; I am not suggesting for a moment that belief sentences, and sentences that attribute the other attitudes, are not relational in nature. What I am suggesting is that the objects to which we relate people in order to describe their attitudes need not be any sense be psychological objects, objects to be grasped, known, or entertained by the person whose attitudes are described.

This point, too, is familiar; Quine makes it when he suggests that we may use our own sentences to keep track of the thoughts of people who do not know our language. Quine's point is semantic, and he says nothing in this context about the epistemological and psychological aspects of the attitudes. We need to bring these various concerns together. Sentences about the attitudes are relational; for semantic reasons there must therefore be objects to which to relate those who have attitudes. But having an attitude is not having an entity before the mind; for compelling psychological and epistemological reasons we should deny that there are objects of the mind.

The source of the trouble is the dogma that to have a thought is to have an object before the mind. Putnam and Fodor (and many others) have distinguished two sorts of objects, those that are truly inner and thus "before the mind" or "grasped" by it, and those that identify the thought in the usual way. I agree that no objects can serve these two purposes. Putnam (and some of the other philosophers I have mentioned) think the difficulty springs from the fact that an object partly identified in terms of external relations cannot be counted on to coincide with an object before the mind because the mind may be ignorant of the external relation. Perhaps this is so. But it does not follow that we can find other objects which will ensure the desired coincidence. For if the object isn't connected with the world, we can never learn about the world by having that object before the mind; and for reciprocal reasons, it would be impossible to detect such a thought in another. So it seems that what is before the mind cannot include its outside connections - its semantics. On the other hand, if the object is connected with the world, then it cannot be fully "before the mind" in the relevant sense. Yet unless a semantic object can be before the mind in its semantic aspect, thought, conceived in terms of such objects, cannot escape the fate of sense data.

The basic difficulty is simple: if to have a thought is to have an object "before the mind," and the identity of the object determines what the thought is, then it must always be possible to be mistaken about what one is thinking. For unless one knows everything about the object, there will always be senses in which one does not know what object it is. Many attempts have been made to find a relation between a person and an object which will in all contexts hold if and only if the person can intuitively be said to know what object it is. But none of these attempts has succeeded, and I think the reason is clear. The only object that would satisfy the twin requirements of being "before the mind" and also such that it determines what the content of a thought must, like Humle's ideas and impressions, "be what it seems and seem what it is." There are no such objects, public or private, abstract or concrete.

The arguments of Burge, Putnam, Dennett, Fodor, Stich, Kaplan, Evans, and many others to show that propositions can't both determine the contents of our thoughts and be subjectively assured are, in my opinion, so many variants on the simple and general argument I have just sketched. It is not just propositions that can't do the job; no objects could.

When we have freed ourselves from the assumption that thoughts must have mysterious objects, we can see how the fact that mental states as we commonly conceive them are identified in part by their natural history not only fails to touch the internal character of such states or to threaten first person authority; it also opens the way to an explanation of first person authority. The explanation comes with the realization that what a person's words mean depends in the most basic cases on the kinds of objects and events that have caused the person to hold the words to be applicable; similarly for what the person's thoughts are about. An interpreter of another's words and thoughts must depend on scattered information, fortunate training, and imaginative surmise in coming to understand the other. The agent herself, however, is not in a position to wonder whether she is generally using her own words to apply to the right objects and events, since whatever she regularly does apply them to gives her words the meaning they have and her thoughts the contents they have. Of course, in any particular case, she may be wrong in what she believes about the world; what is impossible is that she should be wrong most of the time. The reason is apparent: unless there is a presumption that the speaker knows what she means, i.e., is getting her own language right, there would be nothing for an interpreter to interpret. To put the matter another way, nothing could count as someone regularly misapplying her own words. First person authority, the social character of language, and the external determinants of thought and meaning go naturally together, once we give up the myth of the subjective, the idea that thoughts require mental objects.

Notes

1 Graham Wallas, The Art of Thought, Harcourt Brace, 1926.
4 I make no claim for originality here; Stephen Stich has used a very similar example in "Autonomous psychology and the belief-desire thesis," The Monist, 61 (1978), pp. 573ff. I should emphasize that I am not suggesting that an object accidentally or artificially created could not think; The Swampman simply needs time in which to acquire a causal history that would make sense of the claim that he is speaking of, remembering, identifying, or thinking of items in the world. (I return to this point later.)
5 Hilary Putnam, "The meaning of 'meaning'," p. 144.
7 Hilary Putnam, "The meaning of 'meaning'," pp. 164-5.
9 John Searle, Intentionality, chapter 8.
11 Daniel Dennett, "Beyond belief," in Thought and Object, p. 76.
13 Jerry Fodor, "Methodological solipsism considered as a research strategy in cognitive psychology."
See, for example, Gareth Evans, *The Varieties of Reference*, Oxford University Press, 1982, pp. 45, 199, 201.

Tyler Burge, "Individualism and the mental," p. 83.

Ibid., pp. 90, 102, 104.

Burge suggests that the reason we normally take a person to mean by his words what others in his linguistic community mean, whether or not the speaker knows what others mean, is that "People are frequently held, and hold themselves, to the standards of the community when misuse or misunderstanding are at issue." He also says such cases "depend on a certain responsibility to communal practice" ("Individualism and the mental," p. 90). I don't doubt the phenomenon, but its bearing on what it is supposed to show, (a) It is often reasonable to hold people responsible for knowing what their words mean; in such cases we may treat them as committed to positions they did not know or believe they were committed to. This has nothing (directly) to do with what they mean by their words, nor what they believed. (b) As good citizens and parents we want to encourage practices that enhance the chances for communication; using words as we think others do may enhance communication. This thought (whether or not justified) may help explain why some people tend to attribute meanings and beliefs in a legalistic way; they hope to encourage conformity. (c) A speaker who wishes to be understood must intend his words to be interpreted (and hence interpretable) along certain lines; this intention may be served by using words as others do (though often this is not the case). Similarly, a hearer who wishes to understand a speaker must intend to interpret the speaker's words as the speaker intended (whether or not the interpretation is "standard"). These reciprocal intentions become morally important in endless situations which have no necessary connection with the determination of what someone had in mind.

See, for example, "Two thought experiments reviewed," p. 289.

Burge has described "thought experiments" which do not involve language at all; one of these experiments prompts him to claim that someone brought up in an environment without aluminum could not have "aluminum thoughts" ("Individualism and psychology," p. 5). Burge does not say why he thinks this, but it is no means obvious that counterfactual assumptions are needed to make the point. In any case, the new thought experiments seem to rest on intuitions quite different from the intuitions invoked in "Individualism and the mental"; it is not clear how social norms feature in the new experiments, and the linguistic habits of the community are apparently irrelevant. At this point it may be that Burge's position is close to mine.

"Two thought experiments reviewed," p. 288.

"The meaning of 'meaning'," p. 227.

Andrew Woodfield, in *Thought and Object*, p. viii.

"Individualism and the mental," p. 111.


I am greatly indebted to Akeel Bilgrami and Ernie LePore for criticism and advice. Tyler Burge generously tried to correct my understanding of his work.