1987

What is a Belief State?

Curtis Brown

Trinity University, cbrown@trinity.edu

Repository Citation

What Is a Belief State?

Curtis Brown

What we believe depends on more than the purely intrinsic facts about us: facts about our environment or context also help determine the contents of our beliefs.[1] This observation has led several writers to hope that beliefs can be divided, as it were, into two components: a "core" that depends only on the individual's intrinsic properties; and a periphery that depends on the individual's context, including his or her history, environment, and linguistic community. Thus Jaegwon Kim suggests that "within each noninternal psychological state that enters into the explanation of some action or behavior we can locate an 'internal core state' which can assume the causal-explanatory role of the noninternal state."[2] In the same vein, Stephen Stich writes that "nonautonomous" states, like belief, are best viewed as "conceptually complex hybrids" made up of an autonomous component together with historical and contextual features.[3] John Perry, whose term I have adopted, distinguishes between belief states, which are determined by an individual's intrinsic properties, and objects of belief, which are not.[4] And Daniel Dennett makes use of the same notion when he asks:[5]

What, then, is the *organismic contribution* to the fixation of propositional attitudes? How shall we characterize what we get when we subtract facts about context or embedding from the determining whole? This remainder, however we ought to characterize it, is the proper domain of psychology, 'pure' psychology, or in Putnam's phrase, 'psychology in the narrow sense'.

I propose to explore the notion of a belief state. In section I, I will propose an account of the idea of a belief state that is neutral with respect to such questions as how belief states are best characterized and whether they play a central role in psychology or epistemology. In section II, I use this account to describe some recent positions in the philosophy of psychology as sharing the *notion* of a belief state but disagreeing about the *nature* of belief states. The remaining sections develop my own preferred account of how best to characterize belief states. The account I favor takes belief states to be relations between people and semantic or intensional objects. Section III explains why propositions will not do for this purpose: depending on how we specify the relevant propositions, there will either be too many to characterize belief states completely or too few

to characterize them essentially. Following David Lewis and Roderick Chisholm, I suggest that belief states are best characterized by properties. Finally, section IV sketches an account of how the properties that characterize one's belief state are determined by one's intrinsic properties. The account of belief states defended in sections III and IV is an important part of a larger account of belief proper – an account that, however, can only be hinted at here.

I

I shall assume without defense that the objects of belief are propositions, and in particular that if a sentence of the form 'Joe believes that *S*' is true, then the proposition expressed by *S* is an object of Joe's belief. I believe that this view is correct,[6] but it is certainly a minority view. Those who reject it may want to regard it here as merely a simplifying assumption; it should be possible to adapt my account of the notion of a belief state to any view according to which beliefs have objects (whether propositions or something else) that are determined (in some not necessarily straightforward way) by the 'that'-clauses of true ascriptions of belief.

Dennett's helpful term 'organismic contribution to belief' suggests two conditions that must be met for something to be a belief state. First, a belief state is an *organismic* contribution to belief: it is the organism's contribution to belief, not the environment's. We might put this by saying that one's belief state depends only on one's *intrinsic* properties. An "intrinsic property" of an organism is one that involves only facts solely about the organism. Facts about the organism's environment are wholly irrelevant to its intrinsic properties. Seeing an apple does not depend only on intrinsic properties of an organism, since whether one sees an apple depends not only on facts solely about one's body, but also on whether one's environment includes an apple. For precisely the same reason, believing a singular proposition about something in one's environment does not depend only on one's intrinsic properties. Further, the belief state of an organism at a time should involve only properties of the organism *at that time*. Thus we have condition I: the belief state of an organism at a time should supervene on the intrinsic properties of the organism at that time.

Second, a belief state is an organismic contribution to belief. One believes the things one does because of the total state one is in. But not every facet of one's total state contributes to determining what one believes. One's belief state does not depend on facts about one, such as whether one's body has an odd or even number of water molecules, which make no contribution to what one believes. The notion of a belief state must be such as to allow that if all the intrinsic facts about me were exactly as they are except that my left big toe contained one less

water molecule, I would be in the same belief state I am in fact in. Thus we have *condition 2*: a belief state must be a state of the organism that depends only on those facts about the organism that are relevant to what it believes.

Both of these requirements rule that certain facts are irrelevant to what belief state one is in. They give information about the conditions under which one is in a particular belief state.

Suppose that I am at the moment in a particular belief state. Under what circumstances would I be in this belief state? We know already of one set of circumstances in which I am in it: the actual one. But there are more. Some conditions met by my present circumstances do not need to be met for me to be in the same belief state. If we rule that certain facts about my present situation are irrelevant to what belief state I am in, we thereby increase the number of circumstances in which I would be in int.

Some such facts are ruled out by the first condition. Only facts solely about me are relevant to what belief state I am in. No matter how the facts about my environment alter, as long as the intrinsic facts about me remain the same, my belief state remains the same. According to the first condition, then, I would be in the same belief state in all those circumstances in which, however different my environment might be, the intrinsic facts about me remain the same.

The second condition relaxes the conditions on my being in the same belief state still further. Not all the intrinsic facts about me are relevant: just those facts a difference in which would involve a difference in what I believe. Thus the circumstances in which I would be in the same belief state include not only those in which my environment differs but also those in which the intrinsic facts about me differ in ways that do not affect what I believe.

Let us now try to make this intuitive picture a little more precise. I will employ some quasi-technical abbreviations that are admittedly no more precise than the English expressions they abbreviate. Nevertheless, they are helpful because the complexity of the account, despite the simplicity of its basic idea, makes it difficult to state unambiguously in ordinary English.

I begin by dividing the world up in a mildly unusual way. The facts that determine what an individual believes divide into facts solely about the individual and facts about the individual's environment or situation. So we need to theoretical notions. First, we need the notion of the *total intrinsic state*, or simply "total state" for short, of a person, at a time, in a world. Let 'TS' be a variable over total intrinsic states, and let 'TS(x,t,w)' denote the TS of person x, at time t, in world w. For vividness only, I will think of a total state as a

concrete, momentary, physical entity. (We could almost think of a total state as a time-world slice of an individual. But not quite: the facts about a total state do not include facts about whose state it is. TS(me, now, the actual world) is the very same total state as TS(Twin Curtis, now, the actual world), although our current time-world slices are different.)

Our second notion is that of a situation. The situation of x, at t, in w – S(x,t,w) – is the sum of the intrinsic facts about x's environment up to t, including the environment's and x's own past history. S(x,t,w) includes all the intrinsic facts about w, up to and including time t, except those that constitute TS(x,t,w). Again for vividness, I will think of a situation as a concrete physical entity. If possible worlds are thought of as concrete, we can get the situation of x at t in w by taking w, cutting off everything after t, and then cutting out the stage of x at t.

A TS and an S together determine a set of propositions P. Let P(TS,S) be the set of those propositions P such that, for all x, necessarily, if x is in TS and x's situation is S, then x believes P. If we ignore complications introduced because situations contain facts about a world only up to a certain time, and we think of situations and total states as concrete building blocks, then we can visualize P(TS,S) as follows: TS and S fit together to form a world, and P is the set of propositions believed in that world by the person of whom TS is a slice (at the time of the gap into which TS was inserted).

Recall Putnam's Twin Earth example.[7] I believe that water is wet; but were I on Putnam's Twin Earth in the very same TS I am in fact in, I would believe instead that XYZ is wet. The very same TS, embedded in different contexts, warrants different belief ascriptions. We might diagram this situation as shown in the accompanying figure (letting 'ts$_1$' name a particular total state, the one I am in now). Here, s$_1$ is a particular situation – my actual situation, let us say. Then p$_1$ is the set of propositions that my actual total intrinsic state, ts$_1$, determines in s$_1$; that is, p$_1$ = P(tis$_1$, s$_1$). Assuming, as seems plausible, that my total state together with all the facts about my environment, including my own past history, determines all the things I believe,[8] p$_1$ will be the set of all the things I actually believe. Thus p$_1$ includes the proposition that water is wet, but not the proposition that XYZ is wet.

$$s_1{:}p_1 \qquad\qquad s_2{:}p_2 \qquad \cdots \qquad s_3{:}p_3$$

$$ts_1$$
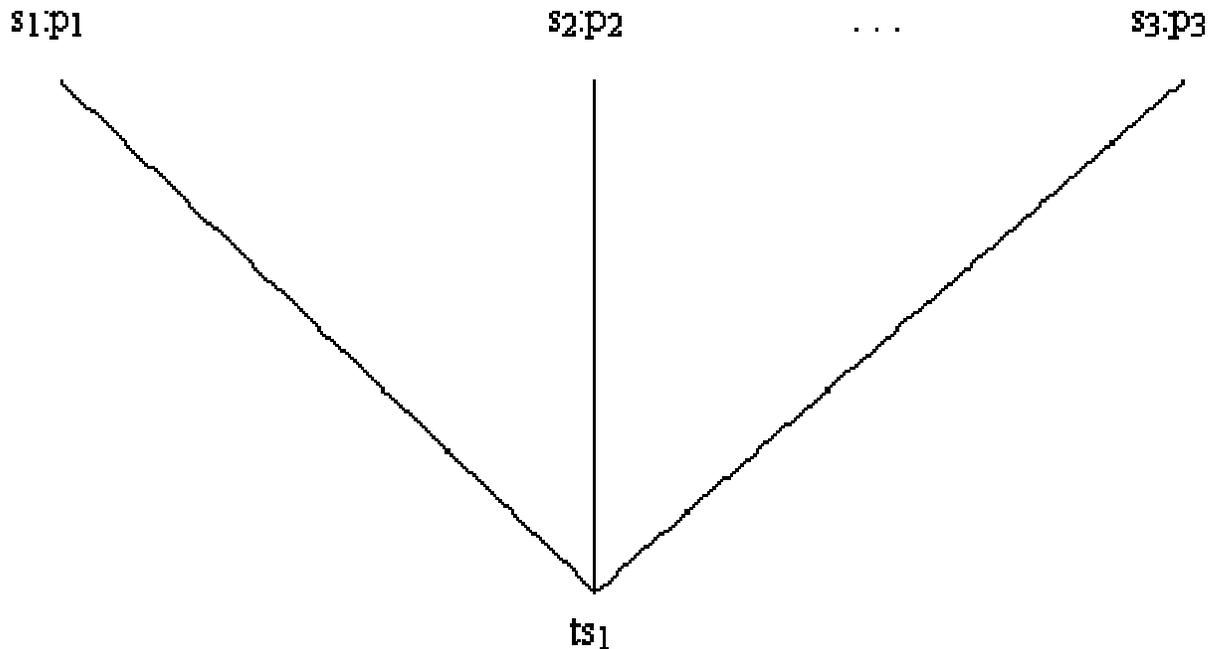
Now s_2 is another situation – perhaps the situation of my Twin Earth counterpart. p_2 is then the set of propositions determined by my actual total state, ts1, together with that situation. p_2 will be the set of propositions believed by my Twin Earth counterpart – including the proposition that XYZ is wet, but not the proposition that water is wet.

The list of situations is then extended to include all the situations, actual and possible, into which ts_1 could be fitted. (The list will include, for example, a counterfactual situation in which I myself grow up on a planet where XYZ fills the lakes. The set of propositions determined by ts_1 and that situation will be like p_2 and unlike p_1 in containing the proposition that water is wet; but it will be like p_1 and unlike p_2 in containing the proposition that I, CB, am named 'Curtis'. Instead, p_2 contains the proposition that Twin Curtis is named 'Curtis'.)

Now, condition 1 says that I am in the same belief state, no matter what situation I am in, provided that I am in ts_1. But according to condition 2, there are also other total states that I (or others) could be in and be in the same belief state: namely, all those total states that make no difference to my beliefs. A total state would make a difference to my beliefs if there were some situation in which it led to different propositions believed. So those total states that make no difference to what I believe are those for which there is no situation in which they produce different propositions believed that ts_1: namely, all those TS such that $P(TS,s_1) = p_1$, $P(TS,s_2) = p_2$, . . ., $P(TS,s_n) = p_n$.

The belief state I am actually in, then, is the state anyone would be in if and only if they were in one of the total intrinsic states just described. For at least some, and perhaps all, purposes we could simply identify this state with the corresponding set of total intrinsic states, just as for at least some and perhaps all purposes we can identify the property of being wooden with the set of all actual and possible wooden things. We can then generalize our account as follows: the belief state of individual x, at time t, and in world w – BS(x,t,w) – is the set of all those TS such that: for all S, if P(TS(x,t,w),S) = Q, then P(TS,S) = Q. In something closer to English, x's belief state (at time t and in world w) is the set of total states that determine, with respect to any situation, the same propositions x's total state (at t and in w) determines with respect to that situation.

My belief state, then, is determined as follows. First, consider the total state I am actually in. Next, consider what beliefs that TS would give me (or others) in all situations of interest. Finally, consider the set of all the TS's that would produce exactly the same beliefs as my actual TS, no matter what they situation. This set of TS's is my belief state.

So my current belief state is the set of all those total states that would produce, in my actual situation, the belief that water is wet (but not the belief that XYZ is wet), together with all the other beliefs I have in the actual situation, and would produce, in the Twin-Earth situation, the belief that SYZ is wet (but not that water is wet), together with all the other beliefs of my Twin-Earth counterpart.

It may be helpful to compare my belief state with my weight state. Consider first the TS I am actually in. Next, consider what weights that TS would produce in all situations of interest (on the Earth, on the Moon, if the Earth were twice as big, etc.). Finally, consider the set of TS's that would produce the same weights as my actual TS, no matter what the situation. Some of these will specify shortness and fatness, some tallness and slenderness; some will specify that their possessor be made of flesh, some that it be made of stone. The set of all of them is my weight state. Or, perhaps more clearly, what they all share and nothing else does is my weight state.

As it happens, weight states are very interesting and important for physics, for a particular weight state is just the property of having a particular mass. Belief states may turn out to be as important for psychology as weight states are for physics. But nothing said so far guarantees that they will be. There may be nothing of interest shared by all and only the total states that make up one's belief state. Belief states as here defined would still exist, but would be of no

systematic importance – like the scattered object consisting of my left thumb and a certain beetle in Africa, or like the property of being a Virgo.

II

We have reached an account of what it could mean to talk about "belief states" or about the "organismic contribution to belief." A good deal of recent discussion can be understood as addressed, more or less explicitly, to the question how best to characterize belief states so understood, or what the nature of belief states is. We have thought of one's belief state as a set of total intrinsic states: those states, roughly, in which one would believe the same things as in one's actual state, no matter what the circumstances. The question of what the nature of a belief state is might then be understood as the question of what, if anything, these total intrinsic states have in common besides the fact that one believes the same things in each of them. If they do share something interesting, then it will seem plausible that it is in virtue of that something that a person in a given belief state believes the things she or he does.

An analogy may help make this clear. We can pick out a certain collection of physical states of material objects as those states that produce in us the sensation of red; and we might then call the whole collection the "state of being red." If all the states in the collection (and no others) share some interesting feature, then it will seem reasonable to say that it is because of that feature that objects appear red to us, and redness will seem to be an intrinsic property of objects. If the states do not share such a feature, then it will seem that the "state of being red" is a gerrymandered state hardly deserving the name at all; redness will seem to be irreducibly a relational property, a property an object has just in case it affects people in a certain way.

Similarly, if the physical states comprised by a given belief state (and no others) share some interesting feature, then it will seem natural to say that it is in virtue of possessing that feature that one believes the things one does, and belief will seem to be an intrinsic property of individuals. If they do not, then "belief states" will seem hardly to bed states at all, and belief will seem to be irreducible relational.

If there is an intrinsic property of redness, then it will be reasonable to say that if things had been different – if people had been constructed differently – redness might have looked blue; if not, there will seem to be a necessary connection between being red and looking red. Similarly, it will seem natural to say that one's belief that P might not have been a belief that P just in case there is an intrinsic property of individuals in virtue of which they believe the things they do.

Recent views about the nature of belief states divide into three groups. First, it may be that they simply have no nature: the various total states that constitute a belief state may share nothing but some long, disjunctive physical property. The "organismic contribution" to belief may simply not be interesting by itself, just as the object's contribution to redness may not be interesting by itself. This seems to be the view of Gareth Evans[9] and of some Wittgensteinian writers;[10] it also seems to be the view of those who think that cognitive psychology needs to devise its own theoretical terms and that these will have no interesting connection with the folk psychological notion of belief. This view is held by Paul and Patricia Churchland[11] and is entertained seriously by Stich.[12] Second, it may be that what the states that constitute the belief state share is somehow specifiable formally or syntactically, as a elation between the believer and a sentence in a "language of thought." This seems to be the view of Jerry Fodor and Hartry Field,[13] and is entertained by Stich, who calls it the "modified Panglossian prospect."[14] Finally, it may be that belief states are best characterized as relations to semantically defined objects such as propositions or properties. This sort of view is at least suggested by work of Robert Stalnaker and David Lewis,[15] and it is a view of this sort that I propose to explore.

The most obvious candidates for semantic objects that characterize our belief states are the propositions we believe. But these propositions clearly do not characterize our belief states essentially. That is the moral of examples like Putnam's: since I could be in the same belief state but believe different propositions, the propositions I in fact believe are not essential to my belief state. This is one reason why many have found the view that belief states are best characterized by semantic objects implausible.

But it would be fainthearted to abandon the idea of a semantic characterization for this reason. Granted, the same belief state yields different propositions believed in different situations. It may nevertheless e that there are *some* propositions that a given belief state determines in *every* situation, and it may be that these provide the characterization we seek.

The leading idea of the account of belief I favor is that of an *immediate object of belief*. Among the things I believe, some are believed indirectly or mediately. The belief that water is wet, the belief that I am married to Karen, the belief that Aristotle wrote the *Nicomachean Ethics* – all these I believe indirectly, since I could be in the same belief state and yet not believe them. Had I grown up on Twin Earth calling XYZ 'water' I would now believe that XYZ is wet, not that water is. Some of my beliefs, then, differ between the actual situation and the counterfactual one. But surely others of my beliefs hold constant across them, and surely it is in virtue of some of these more constant beliefs – such as

the belief that the odorless substance that fills the lakes and is called 'water' is wet – that I believe in the one situation that water is wet, and in the other that XYZ is wet. I suggest that there are some propositions that I believe in every situation in which I am in the same belief state; and that it is in virtue of believing these, together with the varying facts of the situations I could find myself in, that I believe in those situations the rest of the propositions I believe in them. These constant propositions are the immediate objects of belief.

We can define the notion of an immediate object of belief as follows. A proposition P is a (partial) immediate object of (say) Joe's belief just in case, for any possible situation S, if Joe is in the same belief state in S he is actually in, then Joe believes P in S. One's *total* immediate object of belief is then simply the collection of all one's partial immediate objects of belief.

Suppose we want to find an illuminating and helpful way to characterize belief states by means of semantic objects. What is it exactly that we want? I suggest that it is a means of determining, given only intrinsic facts about an individual, a set of semantic objects such that: (1) anyone in the same belief state will determine the same objects – that is, the objects characterize the belief state *essentially*; (2) no one in a different belief state determines exactly the same objects – that is, the objects characterize the belief state *completely*; and (3) the facts about these special objects, together with the facts about one's environment, determine the facts about what one believes in that environment.

My initial hope was that the semantic objects that would meet these conditions would be propositions – in particular, the immediate objects of belief. But it appears that no set of propositions can satisfy both (1) and (2). So the account I will offer is slightly more complex.

This account involves two related notions. The first is the notion of an immediate object of belief. The second, adapted from David Lewis and Roderick Chisholm, [16] is the notion of the property (immediately) self-ascribed by an individual. The total property self-ascribed by an individual is what essentially and completely characterizes the individual's belief state. One's total immediate object of belief *characterizes one's being* in a particular belief state rather than the belief state itself; and it determines, in conjunction with contextual features, the rest of one's beliefs. But the two are intimately connected: from one's total immediate object of belief it is possible t recover one's total self-ascribed property; and one's total self-ascribed property, together with who (and when) one is, determines one's total immediate object of belief.

In the remainder of this paper I will do two things. In section III I will explain why it seems impossible to find a set of propositions that characterize one's belief state both essentially and completely. Reflection on this difficulty will lead to the notion of a self-ascribed property as what best characterizes one's belief state. In section IV, I will sketch an account of how one's intrinsic properties determine one's total self-ascribed property.

III

One's total immediate object of belief, as I have defined it, does not characterize one's belief state essentially. I believe that I am named 'Curtis'. Very likely I would believe this in any situation in which I were in the same belief state I am actually in, so the proposition that I am named 'Curtis' is an immediate object of my belief. Nevertheless, although *I* cannot be in this belief state without believing that I am named 'Curtis', someone else can be. My Twin Earth counterpart is in exactly the same belief state I am in, but does not believe that I am named 'Curtis' – indeed, he has no beliefs about me at all. So although *my being in this* beliefs state is essentially characterized by the proposition that I am named 'Curtis', the belief state itself is not. Of course, if any partial immediate object does not characterize my belief state essentially, neither does my total immediate object of belief.

The definition of 'immediate object of belief' could be modified so as to avoid this result. We could say that an immediate object of my belief is a proposition which anyone in my belief state would believe in any situation. But given this definition, my total immediate object of belief does not characterize my belief state completely.

For consider: Twin Curtis and I are in the same belief state. I believe that I am named 'Curtis', but not that Twin Curtis is; Twin Curtis believes that he is named 'Curtis' but not that I am. There must be something about the belief state we share in virtue of which we believe these different propositions. If our belief state can be completely characterized by means of immediate objects of belief, then there must be some immediate object of belief that we share and in virtue of which we indirectly believe these different propositions. But it seems that there is no such proposition.

The needed proposition cannot be the proposition that the sentence "I am named 'Curtis'" is true: I know that the sentence is true or false only relative to context. Do I believe that the proposition the sentence expresses *in the present context* is true? Let us pas over for the moment problems about direct belief of singular propositions about the present time. There are many different present contexts – for instance, one for me and a different one for Twin Curtis. So

perhaps what I really believe is that the sentence "I am named 'Curtis'" is true in the context I am now in. but then our semantic ascent has accomplished nothing. This cannot be the proposition both Twin Curtis and I believe in virtue of which each of us believes that he is named 'Curtis': Twin Curtis does not believe that he is named 'Curtis' in virtue of believing that the sentence "I am named 'Curtis' is true in the context *I* am now in.

We were evidently wrong to think that the immediate belief we seek is metalinguistic. It might be thought instead that Twin Curtis and I each believe that the unique person who satisfies a certain description is named 'Curtis'. But surely we don't, for we acknowledge that any description we immediately believe we satisfy may be satisfied by someone else as well. Nor does each of us believe that he is named 'Curtis' in virtue of believing that all people who satisfy some description are named 'Curtis', for any description I immediately believe I satisfy could be satisfied by someone who is not really named 'Curtis', and I am aware of this.

If we now take the subjects of attitudes to be persona stages rather than people proper, we may construct similar examples involving different stages of the same person rather than different people. I will present an example showing that if we do not allow singular propositions about oneself (where 'oneself' may refer to a temporal stage of a person), we cannot find propositions in virtue of immediately believing which we believe indirectly some singular propositions about the present time.

Consider first an example that might be thought not to work. In situation 1, Andrea believes, at t_1, that it is then raining. In (counterfactual) situation 2, Andrea is in exactly the same belief state as in situation 1, but she believes at t_2, a different time than t_1, that it is then raining. In situation 1, Andrea does not immediately believe the proposition that it is raining at t_1, since she is in the same belief state in situation 2 but does not believe that proposition. If we are to characterize her belief state, we need something that stays constant across situations 1 and 2.

Let S be the sentence 'It is now raining', P_1 be the proposition that it is raining at t_1, and P_2 be the proposition that it is raining at t_2. If Andrea's belief state can be characterized in terms of propositions, then there must be a proposition Q that she believes in both situations and in virtue of which she believes P_1 in S_1 and P_2 in S_2.

In most such examples, it is plausible that there is an immediate object of belief Q in virtue of which Andrea believes the things she does. We typically identify the present time, I suggest, as the time at which we have a total experience of a

certain character. Let F characterize Andrea's total experience. Then she believes in both S_1 and S_2 that it is raining at the time at which she has F. but the time at which she has F is different in S_1 than in S_2, so in virtue of this immediate belief she had different mediate beliefs in the two situations.

But even if this strategy works for the present example, it will not always work. For consider the following case. Sarah is the subject of a psychological experiment. The experimenters have discovered that by means of a particularly effective hypnotic technique they can get people to relive experiences they have had in the past so exactly that there is no difference at all between the original experience and its later counterpart. They explain this carefully to Sarah. They tell her that on a cold day in January they will take her into their laboratory, hook her up to a variety of devices, and then do nothing but monitor her. On a hot day the following July they will take her into the same laboratory, hook her up to the same devices, and then recreate exactly the same experience she had the previous January. In January, brought into the room and hooked up, she thinks: this is the first time that I have had just this experience, but I will have it again next July. The following summer, hooked up again, she thinks: this is the first time that I have had just this experience, but I will have it again next July. Thus if the character of her experience is F, then at each time she knows that there are two times at which she has F, but also thinks each time that she has never had F before.

Suppose that at both times Sarah accepts the sentence "It is now cold outside." There seems to be no immediate propositional belief on the basis of which she accepts this sentence. She does not believe that it is cold outside at *the* time at which she has F, since she believes that there are two such times. Nor does she believe that it is cold outside at *all* times at which she has F, since she believes that she has F once when it is hot. She does believe that it is cold outside at the *first* time at which she has F. Does this give her reason to accept ""t is now cold outside"" Not unless she also has reason to accept "It is now the first time at which I have F." But if she identifies the current moment as the first time at which she has F, then it seems that the only propositional basis for accepting the latter sentence would be the proposition that the first time at which she has F is the first time at which she has F, and this is no help at all. It seems that there simply is no propositional belief on the basis of which she accepts the sentence "It is now cold outside"; thus, on the revised definition of 'immediate object of belief' we have been considering, there are not enough immediate objects of Sarah's belief to characterize her belief states completely.

Let us revert to the official definition of an immediate object of belief, and let us take the relevant subjects of belief to be temporal stages of Sarah rather than Sarah as a whole. Then we can suggest that in January, the January-stage of

Sarah immediately believes that she (it?) exists at a time at which it is cold. In July, the July-stage of Sarah believes that *it* exists at a time at which it is cold. We now have enough immediate beliefs to fully characterize Sarah's belief state. But too much information, rather than too little, is now captured in the immediate objects of one's belief, so that Sarah's immediate objects of belief to not characterize her belief state essentially. The two Sarah-stages are in exactly the same belief state but do not have the same immediate objects of belief.[17]

Thus it seems that no matter what set of propositions we term one's "immediate object of belief," it will either fail to characterize one's belief state essentially or it will fail to characterize it completely. Is there anything that does both? The account offered by David Lewis and Roderick Chisholm offers hope. I will discuss Lewis's version of this account.

On Lewis's view, belief states are characterized not by the propositions we immediately believe but by the properties we or our temporal stages "self-ascribe." In both winter and summer, temporal stages of Sarah self-ascribe the property of existing at a time at which it is cold outside. The property may be thought of as a set of "centered worlds,"[18] in this case worlds centered on a person-stage that exists at a time at which it is cold outside.

When one self-ascribes a property, one thereby believes that one is the center of one of the centered worlds in the set that constitutes the property. Thus one believes the proposition that one has the property in question. This proposition is an immediate object of one's belief, in my sense, since one believes it in any situation in which one is in the same belief state. Again, it is a proposition that one believes but does not believe in virtue of believing anything else. We have seen that such propositions do no characterize belief states essentially, since someone else could be in the same belief state without believing the same propositions. If Lewis is correct, however, anyone in the same belief state will (immediately) self-ascribe the same properties.

The transition from the account I have given of the immediate objects of belief to the properties that characterize belief states is quite smooth and natural, given the account of belief states in section I. On my account, the immediate objects of one's belief are those propositions that one believes in any situation in which one is in the same belief state. Call this the "official definition." It is instructive to consider the relation between the official definition and an alternative that bears some resemblance to what Lewis and Robert Stalnaker mean by the term 'object of belief'. On this second formulation, the total immediate object of someone's belief is the proposition true at all and only those worlds in which one is in the same belief state one is actually in and all one's beliefs are true. Call this formulation the "alternative definition."

The alternative definition sorts most propositions into the immediate/mediate categories in the same way the official definition does. Consider first a proposition that Ralph believes but that is not an immediate object of his belief: the proposition that water fills the lakes. It is not an immediate object of Ralph's belief on the official definition, since Ralph does not believe it in every situation in which he is in the same belief state: there are Twin-Earth-type situations in which Ralph is in the same belief state but believes rather that XYZ fills the lakes. Also it is not an immediate object of Ralph's belief on the alternative definition, for there are worlds in which Ralph is in the same belief state he is actually in, and all his beliefs are true, but XYZ rather than water fills the lakes.

Consider next a proposition that is an immediate object of Ralph's belief. Let us suppose that the proposition that 'Ralph has arthritis in his thigh' is true is such a proposition. If so, it satisfies the official definition – it is a proposition that Ralph believes in any situation in which he is in the same belief state. But then it also satisfies the alternative definition: if Ralph believes it in every world in which he is in the same belief state, then it must be true in every world in which he is in the same belief state and all his beliefs are true.

The strange case of Sarah shows that the official definition does not provide enough immediate objects of belief to say what it is about Sarah's belief state that remains the same at the two times at which she has the same experience. But Sarah's case also poses an instructive difficulty for the alternative definition: it gives us no clear advice on how to classify certain worlds. Consider the world at which Sarah has the same experience at two different times. And suppose that in the actual world the experiment failed, so that in fact Sarah had the experience in question only once. Is the world in which she had it twice a world in which she is in the same belief state and all her beliefs are true? It depends on which of the occasions when she was in the same belief state we consider. In January she was in the relevant belief state, and all her beliefs (let us suppose) were true; whereas in July she was in exactly the same belief state, but at least one of her beliefs – the belief that it was cold outside – was false. Yet both of these situations are located in the same possible world.[19]

Earlier I compared situations with possible worlds having holes in them. Take the possible world we have just been considering: cut out the relevant stage of Sarah in January and you get the first situation; cut out the relevant stage of Sarah in July and you get the second situation. Plug a Sarah-stage in the relevant belief state into situation 1, and you get a Sarah-stage all of whose beliefs are true; plug the same Sarah-stage into situation 2, and you get a Sarah-stage one of whose beliefs is false.

Situations are similar to the "centered worlds" introduced by Quine and discussed by Lewis. Just as there are many centered worlds for each possible world, so there are many situations for each possible world.

The natural move here seems to be to take as basic not possible worlds but situations. What characterizes one's belief state is not the set of possible worlds that constitutes one's total immediate object of belief, but rather the set of situations in which all the beliefs of anyone in the same beliefs state one is actually in are true. When we first began considering belief states, we found that the notions we needed were that of a state of a person and that of a situation, which might be regarded as a state of the person's environment. These now seem again to be the notions we need to characterize belief states: possible worlds have just gotten in the way.

The picture that emerges from this discussion is one of belief as a three-tiered rather than a two-tiered phenomenon. Instead of having two levels only, those of mediate and immediate belief, we have three: mediate belief, immediate belief, and self-ascription. We believe the things we believe mediately in virtue of believing other things immediately; and some of the things we believe immediately, we believe in virtue of self-ascribing properties.

This can be illustrated as follows. I believe that I am in San Antonio. This is a belief in a singular proposition about San Antonio, and so not an immediate belief. I believe it in virtue of certain further beliefs: for instance, the belief that I am in a town called 'San Antonio'. This latter proposition may be an immediate object of my belief, one I believe but not in virtue of believing anything else. Still this proposition does not characterize my belief state essentially, since others in the same belief state need not believe the proposition that I am in a town called 'San Antonio'. What characterizes my belief state essentially is not the proposition I believe immediately but rather my self-ascription of the property of living in a town called 'San Antonio'.

There is an easy way to derive one's total self-ascribed property from one's total immediate object of belief. For any person-stage x, we can express x's total immediate object of belief by means of a sentence of the form 'x is F'. We can see this as follows. Some of x's immediate objects of belief will be propositions that x has a certain property; some will be propositions that something is the case at a time t; and all will be propositions of some sort or other. So all ascriptions of immediate belief will have one or another of the following forms:

x immediately believes that *x* is F

x immediately believes that Q at *t*

x immediately believes that P.

The first is already in the desired form. For the second, let F be 'exists at a time at which Q'. All remaining cases have the third form; for them let F be 'is such that P' or 'exists in a world in which P'.

Now, if *x*'s total immediate object of belief is the proposition that *x* is F, then, I suggest, *x*'s belief state is essentially and completely characterized by the property F, the total property self-ascribed by *x*. Conversely, if *x*'s belief state is essentially and completely characterized by F, then *x*'s total immediate object of belief is the proposition that *x* is F.

IV

We now have a sketch of an account of belief that takes the notion of an immediately self-ascribed property as basic. But this fundamental notion has been left somewhat obscure. In particular, we need an account of the relation between the intrinsic properties that ground one's belief state and the property that characterizes it. How do the intrinsic facts about me determine the relevant property?

Let us begin by pretending that we are trying to find how one's intrinsic properties determine a set of propositions (the immediate objects of one's belief) rather than of properties. We will be led back to properties soon enough; but it is simpler to begin by considering propositions, and the eventual modification will be minor.

The account we now seek would tell us how to get from information about someone's belief state to information about what propositions are immediate objects of that person's belief. Belief states are in this respect analogous to sentences. Sentences express (in context) propositions, and a semantics of sentences tells us how to get from information about sentences to information about the propositions they express. (A compositional semantics will tell us, for instance, that 'P & Q' is true in any possible world just in case P is true in that world and Q is true in that world.) So it may be appropriate to call the account we seek a "semantics of belief states."

This analogy suggests that consideration of semantics proper may help us to find the sort of account we need of the immediate objects of belief. It is helpful to begin by considering an oversimple picture of the semantics of sentences. This picture is sometimes associated with the term 'procedural semantics'.[20]

The picture begins from the idea that the meaning of a sentence is the proposition it expresses and that propositions are functions from possible worlds to truth values. Sentences mean what people mean by them. So it may seem natural to suppose that what a sentence means (for a person) is a function from worlds to truth values, a function that the person somehow determines. One way to determine such a function would be to be disposed, upon inspection of any world, to pronounce the verdict 'true' or 'false'. Consider for instance the sentence 'The earth is round'. On the simple picture, one understands 'The earth is round' to express the proposition that the earth is round just in case, if asked "Is 'The earth is round' true?" and shown enough of any possible world, one would answer "yes" if the earth were round in that world and "no" otherwise.

Many problems with such a view are immediately apparent. For instance, one can't actually investigate any world but one's own, and the facts about what investigations one pursues are among the facts about that world. But perhaps the picture being considered could be captured along these lines: in any possible world in which all the facts about one relevant to one's understand of 'The earth is round' are the same as they actually are, and in which one pursues the relevant investigations far enough, one assents to 'The earth is round' if the earth is round in that world and dissents from it otherwise. If this approach will not work, then let us simply indulge in the helpful fiction of the Verne-o-scope, a sort of telescope by means of which we can examine as carefully as we like possible worlds other than our own.[21]

More serious difficulties with the simple picture remain. On such difficulty is that we can intend to mean by our expressions propositions that we don't, and know we don't, internalize. Hilary Putnam intends to use the sentence 'There are elms in North America' in such a way that he could not, and knows he could not, reliably say in all situations whether it was true or not. Most of us have similar intentions with regard to sentences containing natural kind terms, indexicals, demonstratives, and proper names. We understand any such sentence to mean some proposition in particular, but *which* proposition we understand it to mean is not determined by facts solely about us.

There is a second difficulty, related to but distinct from the first: the simple picture assumes that a verificationist theory of meaning is correct – that truth conditions for even theoretical sentences depend only on observable facts. This seems plainly false. Suppose that I am an expert on elms. I know not only the way they appear but also the genetic structure that distinguishes them from other trees. I suppose there might be trees that look just like elms but were not. But: show me through the Verne-o-scope a world in which there are in North America trees that look just like elms, but in which people never evolve and so

in which there are no microscopes or other technological aids that enhance our observational skills. I may well not know whether 'There are elms in North America' is true or not. Surely this does not show I do not know what proposition the sentence expresses. If I were given enough information about the world in question, I would know whether the sentence was true. It's just that "enough" information is more than I can glean by unaided observation. But if we try to avoid presupposing verificationism by allowing more information about the world than observation can provide, how are we to specify how much more? If it seems too restrictive to limit the information to the observational, it also seems too generous to admit the information that there elms in North America,[22] and it is not clear how to steer a middle course.

Now let us return to the immediate objects of belief. A simple proposal similar to the simple semantic picture just considered is this: one's total object of belief is a function from worlds to truth values, determined in the following way. The worlds that are assigned the value T are those that, upon examination, the subject would agree could for all he knew be actual. (Notice that not all the worlds the subject would agree could for all he knew be actual will be worlds that really are compatible with all his beliefs. I believe that the lakes on earth are full of water. But show me a world where they are full of XYZ, and I will wrongly claim that it could for all I know be actual. I know more things than I realize.)

Look again at the two criticisms I made of the simple semantic picture. First, we can use sentences to mean propositions we haven't internalized, even ones we couldn't internalize, since we intend social and historical facts to play a role in determining what we mean. The same is true of belief: we believe propositions we haven't, or couldn't, internalize, and for something like the same reasons. But nothing of the sort *could* be true of immediate or direct belief. It is no shortcoming of a procedure for determining the *immediate* objects of belief that it finds no place for social or historical factors.

Second, the simple semantic picture is verificationist, whereas any adequate semantic theory must not be. But perhaps verificationism is not a fault in a semantics of belief states.

The trouble with verificationism is that many sentences have truth conditions that we do not, even could not, have immediate access to. But to be related to a proposition in such a way that one does not have immediate access to its truth conditions is to be related to it in virtue of features of one that might, had matters been different, have related tone to a different proposition. Being related to P as things are, and being related to Q as things might have been, are

indistinguishable from the point of view of the subject. But a proposition that is an immediate object of one'' belief is a proposition one is related to in a way that could not relate one to any other proposition.

Verificationism as a theory of meaning for a public language assumes that there has to be too strong a connection between what a sentence or term means and what is understood by someone who is able to use it correctly. Putnam and others have provided a strong corrective to this view of language. (I do not mean to imply that theirs has been the only important criticism of verificationism.) But the price is to concede that in a sense we don't know the public meaning of most of the terms we use capably enough. We can't immediately believe anything we can't completely understand, so the maneuver that saves us from verificationism with respect to the public language will not save us from verificationism with respect to belief states.

The account we have been considering, according to which the immediate objects of one's belief are determined by which worlds one would agree could be actual, seems able to meet the objections we have so far considered. But a more serious one remains: the account requires us not only to have beliefs but also to be aware of what they are. Dogs have beliefs, but could hardly assent or dissent when presented with alternative possibilities and asked whether they might be actual. Perhaps some of our beliefs are like all of a dog's are in this respect. We need a notion of various possible worlds being compatible with a subject's belief state that does not require the subject to have views about whether they are or are not so compatible. We need a way to determine what a subject's immediate objects of belief are without supposing that the subject can tell us.

Such a procedure for determining what the immediate objects of someone's (say, Art's) belief are, given only nonintentional facts about Art, would be a method of what has been called "radical interpretation." Much of the literature on radical interpretation urges us to adopt one or another "principle of charity": assume that Art's beliefs are mostly true, or that his terms mostly refer, or that words mean the same when he uses them as when we do.[23] If we want to find the immediate objects of Art's belief, however, we will be careful not to apply any such principle. For each of these principles will yield different results in different possible situations in which Art himself remains exactly the same. Take, for instance, the principle that we should interpret Art in such a way that as many of his beliefs as possible come out true. Surely the facts could be different than they in fact are, even radically different, in a situation in which Art was nevertheless exactly the same; and in such a situation the principle would lead us to attribute very different beliefs to Art than we would in the actual situation. Since Art's immediate beliefs remain the same in any situation

in which Art's intrinsic properties are the same, this principle of charity will not lead us to make the correct attributions of immediate belief to Art.

We need instead specific instances of the following two general principles: that we should attribute those direct beliefs that Art would most likely have been led by his experience to acquire; and that we should attribute those direct beliefs that would best explain Art's actions.[24] Even these principles will lead to the desired results only if they employ descriptions of experience and action that are not themselves situation dependent: we shall need to have at hand individualistic accounts of perception and action.

Suppose that there are such principles: principles that will take us from complete descriptions of Art's perceptual input and behavioral output to a correct account of his total set of direct beliefs. Then these principles together with descriptions of Art's perceptions and actions provide something like a functional account of Art's total set of immediate objects of belief. Once we have this general characterization, we can provide a derivative object-by-object account of Art's immediate beliefs: P is a partial immediate object of Art's belief if and only if there is a Q such that Q is Art's total immediate object of belief, and Q implies P.

At its simplest, the account I am suggesting may be put, following Dennett,[25] like this. The set of worlds at which one's total immediate object of belief is true contains those worlds in which one would be most at home or to which one is best adapted: those in which one's behavior would make the most sense and one's expectations be least thwarted. Which worlds these are is determined solely by the intrinsic facts about one, so the appropriate set will not differ with differences in one's environment alone. And this account should make it seem plausible that there are immediate objects of belief, since it should be clear that there will be worlds to which one is especially well adapted.

But now we are in a position to see the need for a modification we have been expecting from the start. Is there really a set of worlds to which Art is best adapted? I am reasonably well adapted to the world I am currently in – provided that I am located toward the end of the twentieth century. Were I in precisely the same intrinsic state, but located in the twelfth century, however, I would be disastrously ill adapted. Again, if I were in my present intrinsic state but located on Venus, I would be ill adapted. What I am well or ill adapted to, then, are situations rather than worlds. To characterize my belief state, we need to consider the set of situations to which I am best adapted.

As noted earlier, a set of situations, like a set of centered worlds, corresponds not to a proposition but to a property: the property one has if and only if one is *in* one of those situations. One's total self-ascribed property is, then, the property of being in one of the situations to which one is in fact best adapted. For any person *x*, if *x* is in the belief state characterized by property F, then *x*'s total immediate object of belief is the proposition that *x* is F.

We have now seen how one's intrinsic properties determine one's total self-ascribed property and how one's total self-ascribed property *k*, together with who (and when) one is, determines one's total immediate object of belief. Thus, to complement the neutral answer to my title question given in section I, we have a sketch of a more substantive answer. To do more would be to go beyond the bounds of the present paper.

But I see this account of belief states as forming part of a larger account of belief proper. Our ordinary or mediate beliefs are propositions we believe in virtue of our immediate beliefs, together with facts about our environment. If the dependence of mediate on immediate belief is systematic, it should be possible to find the principles that govern it; it should be possible, that is, to give an account of belief in general in terms of immediate belief and contextual factors. Such an account would resemble Alvin Goldman's account of action in general in terms of basic actions and "level generation."[26] I believe that such an account is possible, but this is not the place to provide it.[27]

Notes

[1.] Many examples have been used by recent writers to illustrate this point. Such examples all involve cases in which, although all of the relevant intrinsic properties of individuals x and y are the same, x nevertheless has different beliefs than y. One of the most familiar of these examples, provided by Hilary Putnam (see "The Meaning of 'Meaning'," in *Mind, Language and Reality: Philosophical Papers*, vol. 2 (Cambridge, 1974), esp. 223-27), invites us to imagine that I have an exact duplicate on Twin Earth, a distant planet exactly like earth except that the colorless, odorless liquid that falls from the sky, fills the lakes, and is called 'water' by the inhabitants is not H2O but XYZ. Despite sharing all our relevant intrinsic characteristics, my counterpart and I do not share all our beliefs: whereas I believe that water is wet, he does not, believing instead that XYZ is wet.

In Putnam's example, the two individuals considered – myself and my counterpart – are both actual and are exactly alike. Other examples involve individuals who are both actual and who are *not* exactly alike, but who are presumed to be similar in all relevant respects. (See, e.g., John Perry, "The

Essential Indexical," *Nous* 13 (1980): 3-21; Robert Stalnaker, "Indexical Belief," *Synthese* 49 (1981): 129-51; and Stephen P. Stich, *From Folk Psychology to Cognitive Science: The Case against Belief* (Cambridge, Mass., 1983), 63-64). These examples are very useful and illuminating. One could object, though, against Putnam that he requires us to imaging something desperately unlikely (that two actual people could be exactly alike), and against Perry and Stich that differences supposed to be irrelevant might turn out not to be. Both these difficulties are avoided in examples employed by Tyler Burge, which involve the same person in exactly the same state in different situations, at least one of them counterfactual. Having noted that such examples are available, I will continue to use Putnam's for illustration.

That I believe that *water* is wet, and not that XYZ is wet, results in part, Putnam's example suggests, from the actual chemical structure of the salient liquid portion of my environment and not simply from such intrinsic facts about me as how my rods and cones are being stimulated, what words are running through my mind, and how I am disposed to move my body. Other examples in the literature make a similar point about other beliefs – for example, the belief that I have arthritis in my thigh, and the belief that chicory is bitter.

2. Jaegwon Kim, "Psychophysical Supervenience," *Philosophical Studies* 41 (1982): 51-70, at 65.

3. Stich, *From Folk Psychology*, 168-70.

4. See Perry, "The Essential Indexical," as well as "Belief and Acceptance," *Midwest Studies in Philosophy* 5 (1980): 533-42; "A Problem about Continued Belief," *Pacific Philosophical Quarterly* 61 (1980): 317-32.

5. Daniel Dennett, "Beyond Belief," in *Thought and Object*, edited by Andrew Woodfield (Oxford, 1982), 19. A similar notion is employed in Jerry Fodor, "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," in *Representations*(Cambridge, Mass., 1981).

6. See, for example, Tom McKay, "On Proper Names in Belief Ascriptions," *Philosophical Studies* 39 (1981): 287-303; Takashi Yagisawa, "The Pseudo-Mates Argument," *Philosophical Review* 93 (1984): 407-18; and chapter 3 of *my Beliefs and Their Objects* (Ph.D. dissertation, Princeton University, 1982).

7. The example is summarized in note 1.

8. Since situations include facts about a world only up to a certain time, this involves the assumption that future things and events do not affect the content of one's beliefs, though they may affect their truth or falsity. David Kaplan once introduced the name 'Newman 1' as a rigid designator of the first child actually to be born in the twenty-first century. ("Quantifying In," *in Words and Objections: Essays on the Work of W. V. Quine*, edited by D. Davidson and J. Hintikka (Dordrecht, 1969), 206-42, at 228-29.) Did he then believe singular propositions about Newman 1? Not if the account sketched here is correct. The facts about Kaplan's situation and his total state were compatible with Newman 1's being either Sally or Joe, so if anything determined that Kaplan believed singular propositions about one rather than the other, it was something other than the facts about his situation and total state. But on the present account, there is nothing else that determines the contents of our beliefs.

This conforms to Kaplan's view in "Quantifying In," though he seems to disagree in his later "Dthat," in *Contemporary Perspectives in the Philosophy of Language*, edited by Peter A. French, et al. (Minneapolis, Minn., 1979), 383-400, at 397. The account also meshes with Keith Donnellan's treatment of Kaplan's example in "The Contingent a Priori and Rigid Designators," in ibid., 45-60, at 53-58.

9. See his response to Jerry Fodor's "Methodological Solipsism," *in Behavioral and Brain Sciences* 3 (1980): 79-80.

10. See, e.g., Bruce Goldberg, "The Correspondence Hypothesis*," Philosophical Review* 77 (1968): 438-54.

11. See esp. Paul M. Churchland, "Eliminative Materialism and Propositional Attitudes," *Journal of Philosophy* 78 (1981): 67-90.

12. Stich, *From Folk Psychology*, chap. 10 and 11.

13. Fodor, Representations; Hartry Field, "Mental Representation," Erkenntnis 13 (1978): 9-61; reprinted in *Readings in the Philosophy of Psychology*, edited by Ned Block, vol. 2 (Cambridge, Mass., 1981), 78-114.

14. Stich, *From Folk Psychology*, chap. 11.

15. Robert Stalnaker, "Propositions," in *Issues in the Philosophy of Language*, edited by Alfred F. MacKay and Daniel D. Merritt (New Haven, Conn., 1976), 76-91, and "Indexical Belief"; David Lewis, "Attitudes *De Dicto* and *De Se,"* Philosophical Review* 87 (1979): 513-43. Neither Stalnaker nor Lewis employs

the notion of a belief state, and Stalnaker at least regards the notion with suspicion (private communication).

16. David Lewis, "Attitudes *De Dicto* and *De Se*"; Roderick Chisholm, *The First Person* (Minneapolis, Minn., 1981).

17. Interestingly, Robert Stalnaker ("Indexical Belief," 143) considers an example similar to the present one and takes something similar to the line of the previous paragraph. (Instead of allowing Sarah-stages to have singular propositions about themselves as immediate objects of their belief, Stalnaker would allow Sarah to have singular propositions about particular thought tokens as something like immediate objects of belief.) But it is not clear whether he realizes that the price of adopting this line is acknowledging that immediate objects of belief do not characterize belief states essentially. I should note here that Stalnaker does not use the notion of an immediate object of belief, and my attempts to cast his views in my terminology may well be misleading.

18. Lewis takes the notion of a centered world from W. V. Quine, "Propositional Objects," in Quine, *Ontological Relativity* (New York, 1969).

19. Compare a passage from Quine's "Propositional Objects." Quine is considering a cat that wants to escape from a dog onto a roof: "One of those possible worlds will have a cat like him on a roof like his, and another cat like him in the dog's jaws; does it belong to both the desired state of affairs and the feared one?" (quoted by Lewis, "Attitudes *De Dicto* and *De Se*," 531). Stalnaker avoids this sort of example by postulating distinct but indiscernible possible worlds ("Indexical Belief," 143-45).

20. I do not mean to imply that the oversimple picture is all there is to procedural semantics. Some lively discussion and criticism of procedural semantics may be found in Fodor, "Tom Swift and His Procedural Grandmother," 204-24 of *Representations*.

21. The Verne-o-scope was invented by David Kaplan. See his "Transworld Heir Lines," in *The Possible and the Actual*, edited by Michael J. Loux (Ithaca, N.Y., 1979), 88-109, at 93.

22. What I have in mind is that my observations through the 'scope are supplemented by something like a written list of information about the world in question. The list is too generous if it includes the sentence 'There are elms in North America', since being able to tell that a sentence S is true in a world on the basis of a list that includes S itself hardly guarantees that one understands S.

the information that there are elms in North America would be admissible if it were not conveyed by the *sentence* 'there are elms in North America'. But how else can it be conveyed?

23. See, among other writings of Donald Davidson, his "Radical Interpretation," Dialectica 27 (1973): 313-28; see also Hilary Putnam's discussion of the "Principle of Benefit of Doubt" in "Language and Reality," 272-90 of Mind, Language and Reality, esp. 274-77.

24. Such principles are offered, along with others, in David Lewis, "Radical Interpretation," *Synthese* 27 (1974): 331-44. There is an extremely interesting discussion of the sort of radical interpretation I have in mind in Daniel C. Dennett, "Beyond Belief," sec. 4.

25. See "Beyond Belief," sec. 4, 40-44.

26. Alvin Goldman, A Theory of Human Action (Princeton, N.J., 1970).

27. I offer the beginnings of such an account in *Beliefs and Their Objects*, chap. 4. [Note added 1999: Now see instead my "Direct and Indirect Belief," *Philosophy and Phenomenological Research* 52 (1992): 289-316. An online version is also available.]