ISSUES IN THE PHILOSOPHY OF LANGUAGE

Proceedings of the 1972 Oberlin Colloquium in Philosophy

Edited by
Alfred F. MacKay and Daniel D. Merrill

New Haven and London, Yale University Press, 1976

successful act of pretended naming? Not I. Nor do I want to say that a faulted name is a pretended one. My way of looking at the matter immediately leads to a familiar analysis of reference failure for names in acts of assertion: the use of faulted names issues in faulted assertions, perfectly meaningful no doubt but unsuccessful in failing to produce statements, true or false, capable of providing answers to questions whether something was so—was Homer really blind?

PROPOSITIONS

Robert Stalnaker

Propositions are things people express when they make predictions or promises, give orders or advice. They are also things people doubt, assume, believe to be very likely, and hope are true. What kind of thing are they? My aim is to present and discuss an account of propositions that appears to have great theoretical promise, but that also is faced with serious philosophical difficulties. I will first give a brief outline of the account I have in mind; then provide this account with some philosophical justification by tying it to an independently plausible account of propositional attitudes; and finally, raise and respond to some of the serious philosophical problems that the account faces. I cannot solve them, but I hope to indicate that they are not insurmountable problems and so are not reasons to reject the account out of hand.

The account of propositions that I have in mind is a byproduct of the semantical treatment of modal logic which defines necessity and possibility in terms of a structure of *possible worlds*. According to this kind of interpretation, the formulas of modal logic are assigned truth values not directly, but relative to possible worlds or possible states of the world. Exactly one possible world, or possible state of the world, is *actual*, and truth itself is just truth in this actual world.

Just as a domain of individuals must be specified in order to interpret sentences in first-order extensional logic, so a domain of possible worlds, each with its domain of individuals, must be specified in order to interpret first-order modal logic. This move allows for a natural interpretation of statements of necessity: "it is necessary that P" means that P is true in all possible worlds in the domain. It also allows for a natural distinction between the intensions and the extensions of singular terms, predicates, and sentences. The extension of an expression is given relative to a possible world; it is what is denoted by that expression in that possible world. The intension of an expression is the rule by which the extension is determined. Thus since the extension or denotation of a singular term is an individual, the intension is a function from possible worlds into individuals (an individual concept). Since

the extension of a one-place predicate is a class of individuals, the intension of a predicate—the property it expresses—is a function from possible worlds into classes of individuals. And if one takes the extension of a sentence to be a truth value, then the intension of a sentence—the *proposition* it expresses—will be a function taking possible worlds into truth values. Equivalently, a proposition may be thought of as a set of possible worlds: the set of worlds in which the sentence expressing the proposition denotes the value true.

Intuitively, this account of propositions suggests that to *understand* what a sentence says is to know in what kinds of situations it would be true and in what kinds it would be false, or to know the rule for determining the truth value of what was said, given the facts. It also means that two sentences express the same proposition in a given context relative to a set of possible worlds just in case they are true together and false together in each of those possible worlds.

Now if propositions are to be the objects of speech acts and propositional attitudes, why should they be understood in this way? Part of the justification requested by this question can be given by pointing to the technical success of the theory of possible worlds in resolving paradoxes concerning referential opacity, in finding and analyzing subtle scope ambiguities, and in providing a formally elegant framework for the representation of the structure of intensional concepts. But a more general philosophical justification for this account can, I believe, be given. This justification rests on an assumption shared by some philosophers who reject the possible-world approach, and it has, I think, independent plausibility.

The assumption is that beliefs, presumptions, and presuppositions, as well as wants, hopes, and desires, are functional states of a rational agent. A functional state is a state which is defined or individuated by its role in determining the behavior of the object said to be in the state. In the case of propositional-attitude concepts, the objects are rational agents, and the relevant kind of behavior is rational behavior. Thus the notions of believing, wanting, and intending, on the assumption I am making, belong to a theory of rationality—a theory which is intended to explain how rational creatures operate when they deliberate, investigate, and communicate; that is, to answer questions about why their actions and reactions are appropriate when they are. A simple theory of this kind goes back at least to Aristotle and is taken for granted by common sense explanations of

behavior. Its most basic concepts are belief and desire (where desire is taken broadly to include long-range dispassionate ends as well as attitudes more naturally called desires). To explain why a person did something, we show that by doing it, he could satisfy his desires in a world in which his beliefs are true. For example, I explain why Sam is turning cartwheels on the front lawn by pointing out that he wants to impress Alice and believes that Alice will be impressed if he turns cartwheels on the front lawn. The notions of belief and desire used in the explanation are correlative dispositions that jointly determine action.

Now if what is *essential* to belief is that it plays this kind of role in determining action, what is essential to the objects of belief? I shall argue that what is essential is given by the possible-world account of propositions sketched above.

First, the functional account, as a theory of rational action, already contains implicitly an intuitive notion of alternative possible courses of events. The picture of a rational agent deliberating is a picture of a man who considers various alternative possible futures, knowing that the one to become actual depends in part on his choice of action. The function of desire is simply to divide these alternative courses of events into the ones to be sought and the ones to be avoided, or in more sophisticated versions of the theory, to provide an ordering or measure of the alternative possibilities with respect to their desirability. The function of belief is simply to determine which are the relevant alternative possible situations, or in more sophisticated versions of the theory, to rank them with respect to their probability under various conditions of becoming actual.

If this is right, then the identity conditions for the objects of desire and belief are correctly determined by the possible-world account of propositions. That is, two sentences P and Q express the same proposition from the point of view of the possible-world theory if and only if a belief or desire that P necessarily functions exactly like a belief or desire that Q in the determination of any rational action. Suppose P and Q express the same proposition in the sense that they are true together and false together in all possible courses of events conceivable to some agent. If any of his attitudes toward the content of P were to differ from his attitudes toward the content of Q, then no coherent division or ranking of the alternative possibilities would be determined, and no straightforward rational explana-

tion of any action could be given. An attitude toward P will be functionally equivalent to the same attitude toward Q, and in a functional theory, functional equivalents should be identified.

Now suppose that P and Q express different propositions in the sense that there is some possible course of events in which they differ in truth value. Then one can always imagine a coherent context of deliberation—one in which the agent's attitudes toward the possibilities that distinguish the two propositions are crucial. In such a context, he will have different attitudes toward P than he has toward Q.

A second reason that the possible-world theory provides a concept of propositions which is appropriate for the functional account of propositional attitudes is that it defines propositions independently of language. If desires and beliefs are to be understood in terms of their role in the rational determination of action, then their objects have nothing essential to do with language. It is conceivable (whether or not it is true) that there are rational creatures who have beliefs and desires, but who do not use language, and who have no internal representations of their attitudes which have a linguistic form. I think this is true of many animals—even some rather stupid ones—but there might be clearer cases. Imagine that we discovered living creatures—perhaps on some other planet—who did not communicate, but whose behavior was predicatble, for the most part, on the hypothesis that they engaged in highly sophisticated theoretical deliberation. Imagine further that we had this indirect evidence supporting our hypothesis: that the beliefs that our hypothesis attributed to these creatures could be causally explained, in many cases, in terms of their sensory inputs; and that the desires attributed to them by the hypothesis were correlated appropriately, for the most part, with the physical requirements for their survival. Finally, imagine that we test the hypothesis by manipulating the environments of these creatures, say by feeding them misleading "evidence" and by satisfying or frustrating some of their alleged desires. If they continued to behave as predicted, I think we would be tempted to attribute to these creatures not just belief and desire analogues, but beliefs and desires themselves. We would not, however, have any reason to hypothesize that they thought in a mental language, or in any language at all.

It is plausible to think that if such creatures were intelligent and adaptable enough, it would almost certainly be in their interest, and within their power, to develop ways of communicating their beliefs and desires. Hence, a community of such sophisticated but inarticulate rational agents would be surprising. But it is not an incoherent hypothesis that there are such creatures, and in any case, on the functional account, the development and use of language is viewed as one pattern of rational behavior among others, and not as something on which the concept of rational behavior is itself dependent. For this reason an account of propositions that treats them as linguistic items of some kind would be inappropriate.

Even if we are concerned only with the behavior of real, language-using rational creatures, we should not treat the objects of propositional attitudes as essentially linguistic. There is no reason, according to the functional theory, a person cannot have a belief that goes beyond the expressive power of the language he speaks or that can be expressed only imperfectly in his language. For example, I may believe of a certain person I saw last week that he is a spy. I may not know his name, or even remember that I saw him last week; I just remember *him*, and I believe that he is a spy. You may attribute the belief to me (for example, by saying, "Stalnaker believes that Ortcott is a spy") in the course of explaining my behavior toward Ortcott without attributing to me either the language in which you express my belief or any translation of it. It should be clear that I may have the belief even if I know of no name or accurate unique description of the person whom my belief concerns. But it would be gratuitous to suppose that in this case there is a private inexpressible name which occurs in my belief. Such a supposition would be required by an account which treated propositions as linguistic things.

There are, of course, several essential features of the objects of propositional attitudes which are also essential features of linguistic items such as statements. Both can be true and false, and can stand in logical relations like implication, independence, and incompatibility. The possible-world account attributes these logical features to propositions without any of the extraneous structure of language. Propositions, according to this account, have no syntax, no "exact words" or word order, no subjects, predicates, or adverbial phrases; nor do they contain semantical analogues to these notions. This accords with the functional account, which assigns no role to such grammatical notions in the explanation of behavior. It also accords with intuitive ideas about belief and other propositional attitudes. We do naturally talk about true and false, incompatible and independent beliefs.

But we do not normally talk about the first word, or the subordinate clause in a belief. For these reasons, it seems plausible to maintain that while beliefs resemble statements in some ways and are often expressible in statements, they are not, as statements are, composed of linguistic elements.

Before looking at the problems with this account that I find difficult, let me dismiss two that I think are not. First, some people find that the possible-world theory troubles their ontological consciences. Since there really are no such things as possible worlds, how can we take seriously a theory that says there are? Second—a closely related worry—some people claim not to understand the notion of a possible world. They say it has no useful intuitive content, and so it cannot play an essential role in an adequate explanation of propositional attitudes.

The first objection seems to me not to be distinct from a general objection that the theory as a whole is not fruitful. If the possible-world theory is useful in clarifying relationships among actions and attitudes, or among the contents of statements and beliefs, and if the basic notions of that theory cannot be analyzed away, then we have as good a reason as we could want for saying that possible worlds *do* exist, at least insofar as it is a consequence of the theory that they do. A simple denial of existence is not a good reason to reject a theory. Rather, one has reason to deny the existence of some alleged theoretical entity only if one has independent reason to reject the theory.

Philosophers with strict ontological scruples often justify their skepticism about some alleged entity by claiming not to understand it. Some people claim just not to know what a possible world could be, not to be able to recognize one or tell that one is different from another. One aim of drawing the connection between the possible-world theory and the functional account of propositional attitudes is to help such people understand possible worlds—to support the claim that the notion does have intuitive content, and to identify one of its sources. The connection suggests that we need at least a rudimentary notion of alternative possible situations in order to understand such notions as belief and rational deliberation. If this is right, then a notion of possible worlds is deeply involved in our ordinary ways of regarding some of our most familiar experiences.

This intuitive notion of an alternative possible state of affairs

or course of events is a very abstract, unstructured one, but that is as it should be. The notion of rationality, as explained by the functional theory, involves a notion of alternative possibilities, but it does not impose any structure on those possibilities. That is, it is no part of the idea of rational deliberation that the agent regard the possible outcomes of his available alternative actions in any particular way. The kind of structure attributed to possible worlds will depend on the application of the theory to a particular kind of rational agent in a particular kind of context.

While the possible-world theory itself is neutral with respect to the form of individual possible worlds, one philosophical application of the theory is as a framework for the articulation of metaphysical theories which may impose some structure on them. One may think of possible worlds as quantities of some undifferentiated matter distributed in alternative ways in a single space-time continuum, or as alternative sets of concrete particular substances dressed in full sets of properties, or as a structure of platonic universals participating together in alternative ways. Those whose inclinations are antimetaphysical may think of possible worlds simply as representations of alternative states of some limited subject matter relevant to some specific deliberation, inquiry, or discussion.

I have not really answered either the ontological skeptic or the philosopher who does not understand what possible worlds are. Rather, I have suggested that these people present not specific objections, but expressions of general skepticism about whether the theory of possible worlds has any fruitful application. A full answer can be given only by developing the theory and by applying it to particular problems.

Let me go on to a more specific and troublesome problem with the possible-world theory as applied to propositional attitudes. The problem is this: if two statements are logically equivalent, then no matter how complex a procedure is necessary to show them equivalent, they express the same proposition. Hence, if propositions are the objects of propositional attitudes, then any set of attitudes which an agent has toward the content of the one statement must be the same as the set of attitudes which he has toward the content of the other. But this is not plausible. If a person does not realize that two statements must have the same truth value, he may believe what the one says while disbelieving what the other says. And in many cases, it may be unrealistic and unreasonable to expect an agent to realize that two statements are equivalent.

The natural first reaction to this problem would be to try to develop finer identity conditions for propositions; that is, to develop a concept of proposition according to which logically equivalent statements sometimes may say different things. But if the intuitive account of propositional attitudes that we are using is right, then this reaction is a mistake. We have previously argued that the identity conditions that our theory imposes on propositions are exactly right from the point of view of the role of beliefs and desires in the rational determination of action. Hence the paradoxical consequence about logically equivalent statements is not just an unfortunate technical consequence of possible-world semantics which demands a technical solution. Rather, it is a consequence of the intuitive picture of belief and desire as determinants of action. In terms of this picture, it is not at all clear what it would mean to say that a person believed that P while disbelieving Q where P and Q are logically equivalent. There is no pattern of behavior, rational or irrational, that the hypothesis could explain. So, because I find this intuitive picture of belief and desire persuasive, I shall not respond to the problem in this way. Instead, I will take the heroic course: embrace the paradoxical consequence and try to make it palatable.

The usual way to make the consequence palatable is to admit that the functional theory of attitudes is an idealization which fits the real world only imperfectly. The ideal notions of belief and desire apply literally only to logically omniscient rational intelligences—agents whose behavior conforms strictly to a certain kind of coherent pattern. Of course no mere mortal rational agent can be expected to have a pattern of behavior which is fully coherent in every detail. The theory cannot plausibly be applied unless certain actions are set aside as actions to be explained not as consequences of some rational process, but in terms of some breakdown or limitation in the rational powers of the agent.

This admission is, I believe, correct, and it is relevant to explaining the possibility of irrational action. But it will not avoid or make palatable the paradoxical consequence for at least two reasons. First, while one might explain the *appearance* of incompatible beliefs, or the failure to believe all the equivalents of one's beliefs, in this way, one could never accept the appearance as reality. No matter how confused or irrational a person may be, one cannot consistently describe his state of mind by saying that he believes that P but fails to believe that Q where P and Q are logically equivalent, since in that case, the

proposition expressed by P just *is* the proposition expressed by Q. A person can be so incoherent in his behavior that one hesitates to apply the notions of belief and desire to him at all, but his incoherence can never justify applying these notions to him in an inconsistent way.

The second reason is that this explanation of the paradoxical consequence seems to rule out too much as a "deviation from the norm." One cannot treat a mathematician's failure to see all the deductive relationships among the propositions that interest him in this way without setting aside all of mathematical inquiry as a deviation from rationality. But this would be absurd. Mathematical inquiry is a paradigm of rational activity, and a theory of rationality which excluded it from consideration would have no plausibility.

Let us look more closely at the paradoxical consequence, which I have expressed in a way that ignores use-mention distinctions. It is that if a person believes that P, then if P is logically equivalent to Q, he believes that Q. In this formulation, the expression "that P" is a schema for a nominalized sentence, which denotes some proposition. The statement "P is logically equivalent to Q," however, is a schema for a claim about the relation between two *sentences*. Hence the letters P and Q here stand in for expressions that denote things that *express* the proposition that P. Now once this is recognized, it should be clear that it is not part of the allegedly paradoxical consequence that a person must know or believe that P is equivalent to Q whenever P *is* equivalent to Q. When a person believes that P but fails to realize that the sentence P is logically equivalent to the sentence Q, he may fail to realize that he believes that Q. That is, he may fail to realize that one of the propositions he believes is expressed by that sentence. In this case, he will still believe that Q, but will not himself express it that way.

Because items of belief and doubt lack grammatical structure, while the formulations asserted and assented to by an agent in expressing his beliefs and doubts have such a structure, there is an inevitable gap between propositions and their expressions. Wherever the structure of sentences is complicated, there will be nontrivial questions about the relation between sentences and the propositions they express, and so there will be room for reasonable doubt about what proposition is expressed by a given sentence. This will happen in any account of propositions which treats them as anything other than sentences or close copies of sentences.

Now if mathematical truths are all necessary, there is no room

for doubt about the propositions themselves. There are only two mathematical propositions, the necessarily true one and the necessarily false one, and we all know that the first is true and the second false. But the functions that determine which of the two propositions is expressed by a given mathematical statement are just the kind that are sufficiently complex to give rise to reasonable doubt about which proposition is expressed by a statement. Hence it seems reasonable to take the objects of belief and doubt in mathematics to be propositions about the relation between statements and what they say.

This suggestion is prima facie more plausible in some cases than in others. To take an easy case, if I do not recognize some complicated truth-functional compound to be a tautology, and so doubt whether what it says is true, this is obviously to be explained by doubt or error about what the sentence says. But in branches of mathematics other than logic it seems less plausible to take the objects of study to be sentences. For these cases we might take the objects of beliefs and doubts to be a common structure shared by many, but not all, of the formulations which express the necessarily true proposition. This common structure would be a kind of intermediate entity between the particular sentences of mathematics and the single, unstructured necessary proposition. In this kind of case, doubt about a mathematical statement would be doubt about whether the statements having a certain structure express the true proposition.

This suggestion for explaining mathematical ignorance and error implies that where a person fails to know some mathematical truth, there is a nonactual possible world compatible with his knowledge in which the mathematical statement says something different from what it says in this world. To develop this suggestion, one would of course have to say much more about what these nonactual possible worlds are like for particular mathematical contexts. Such a development would be a part of a theory of mathematical knowledge. I have no such account in mind, and I do not know if an account that is both plausible and consistent could be constructed. My only aim in presenting the suggestion is to show that there is at least a possibility of reconciling a possible-world theory of propositions and propositional attitudes with the rationality of mathematical inquiry.

There is a closely related problem with a parallel solution. The problem arises for those of us who have been convinced by Saul Kripke's arguments that there are necessary truths that can be

known only a posteriori. That is, there are statements such that empirical evidence is required in order to know that they are true, but nevertheless they are necessarily true, and so true in all possible worlds. The best examples (although not the only ones) are identity statements containing two proper names like "Hesperus is identical to Phosphorus." It is obvious that empirical evidence is required to know that this statement is true, and it is also obvious that the relevant evidence consists of astronomical facts, and not, say, facts about meanings of words or linguistic usage. On the other hand, to see that the proposition is necessary, consider what it could mean to suppose, contrary to fact, that it were false. How could Hesperus not have been Phosphorus? It might have been that other planets—says Mars and Jupiter—were *called* Hesperus and Phosphorus, but this is not relevant. It also might have been that a different planet was seen in a certain place in the evening where Hesperus is in fact seen. But to suppose this is not to suppose that a different planet *was* Hesperus, but to suppose that it was not Hesperus which was seen in the evening. If we mean to suppose, quite literally, that Hesperus itself is distinct from Phosphorus itself, then we are just not supposing anything coherent. The planet could not have been distinct from itself.

My point is not to defend this conclusion, which is adequately done elsewhere, but to reconcile it with the thesis that propositions in the sense explained are the objects of propositional attitudes. The reason reconciliation is needed is this: consider any necessary truth which can be known only a posteriori. Since knowledge of it depends on empirical evidence which one might not have, it is possible for a person—even an ideally rational, logically omniscient person—to be ignorant of that truth. But in the possible-world account of propositional attitudes, this means that there might be a possible world compatible with the person's knowledge in which the proposition is false. But this is impossible, since the proposition is necessary, and hence true in all possible worlds. Thus it would seem that the existence of necessary but a posteriori truths is incompatible with a possible-world account of knowledge.

Let us consider what happens when a person comes to know that Hesperus is identical to Phosphorus after first being in doubt about it. If the possible-world analysis of knowledge is right, then one ought to be able to understand this change in the person's state of knowledge as the elimination of certain epistemically possible worlds. Initially, certain possible worlds

are compatible with the subject's knowledge; that is, initially, they are among the worlds which the person cannot distinguish from the actual world. Then, after the discovery, these worlds are no longer compatible with the subject's knowledge. What would such possible worlds be like? If we can give a clear answer to this question, then we will have found a *contingent* proposition, which is what astronomers learned when they learned that Hesperus was identical to Phosphorus.

If we are right about the necessity of the proposition that Hesperus is identical to Phosphorus, then the possible worlds ruled out in the discovery will not be possible worlds in which Hesperus is distinct from Phosphorus, since there are none of those. Nevertheless, there are some perfectly clear and coherent possible worlds which are compatible with the initial state of knowledge but incompatible with the new one. They are worlds in which the person in question exists (since presumably he knows that he exists), and in which the proposition he would express in *that* world with the sentence "Hesperus is identical to Phosphorus" is false. That proposition will be different from the one expressed by the sentence in the actual world, since it is a contingent fact that the name "Hesperus" picks out the planet that it does pick out. Moreover, a person using the name properly might be in doubt or mistaken about this fact. In such a case, the same sentence, with the same rules of reference which determine its content, will express different propositions in different possible worlds compatible with his knowledge. It is a contingent fact that the proposition expressed is necessarily true, and it is this contingent fact which astronomers discovered.

If this is right, then the relevant object of knowledge or doubt is a proposition—a set of possible worlds—but a different one from the one that is necessarily true. There are two propositions involved, the necessary one and a contingent one. The second is a function of the rules which determine the first.

Now if the person, after finding out that Hesperus is identical to Phosphorus, were to announce his discovery by *asserting* that Hesperus is identical to Phosphorus, what would he be saying? If his assertion is really announcing his discovery, if what he is saying is what he has just come to believe, then it is the contingent proposition that he is asserting. There is generally no point in asserting the necessary proposition, although there is often a point in saying that what some statement says is necessarily true.

I will conclude my defense of the possible-world definition

of propositions by summarizing three points that I have tried to make. First, I argued that this theory is motivated not just by the mathematical elegance of the model-theoretic framework, but by a familiar intuitive picture of propositional attitudes. I suggested that this picture in part explains the heuristic power and intuitive content of the notion of a possible world. Second, I argued that the philosophical problems that this theory faces are deep ones; that is, they spring from essential features of the intuitive picture of propositional attitudes, and not from accidental and removable features of possible-world semantics. Any account of propositional attitudes which explains them in terms of their role in the determination of rational action, and any account which treats the objects of these attitudes neither as linguistic items nor as close copies of linguistic items, will be faced with these or similar problems. Finally, I suggested that there is at least a hope of solving the problems without giving up the basic tenets of the theory if we recognize and exploit the gap between propositions and the linguistic formulations which express them. Ignorance of the truth of statements which seem to express necessary propositions is to be explained as ignorance of the relation between the statement and the proposition. I have not carried out the explanation in the most difficult case of mathematical ignorance, but I hope I have shown that such an explanation might be possible.

COMMENTS

Lawrence Powers

Professor Stalnaker's paper exhibits his usual philosophical wit and elegance, but I cannot help but think the paper is basically wrongheaded throughout.

Before moving to my more substantial criticisms, I should like to indulge myself in railing and fulminating against a part of Stalnaker's paper that I found positively irritating. This part is a remark of his which seems to say that it does not matter if the possible-world theory is committed to the existence of nonexistent entities so long as the theory is fruitful. This remark strikes me as antiphilosophical and, if I may put it this way, offensive to right reason. Let us see how Stalnaker could have been led to such desperation.

It seems, whether rightly or wrongly, that the possible-world theory holds that there are such things as possible worlds. However, it also seems, at least at first glance, that there obviously are no such things as possible worlds. Therefore it seems that the possible-world theory is false.

I am not suggesting that the possible-world theory is not *fruitful* or *useful.* I am suggesting that it appears to be *false.* One can only wait and see whether it is useful, but one's first reaction is that any utility of the possible-world theory will be, as Russell would put it, one of the advantages of robbery over honest labor.

Indeed I know that there are many puzzles and paradoxes which the possible-world theory helps us to resolve. But this fact is irrelevant here. This fact does not show that the theory helps us to resolve the *particular* paradox before us, namely that it seems an incredible idea that there are any such things as possible worlds. In order to resolve this puzzle, we must try to spell out why it seems incredible that there should be such things and then point out those clarifications which will remove this appearance of incredibility. Perhaps Stalnaker's intuitions are so dulled that he does not find the real existence of all these possible worlds a prima facie fantastic idea. Let me try to reawaken his intuitions here.

There were in March 1972 such things as possible Democratic nominees—Muskie, Humphrey, etc. These were actual things

which could become Democratic nominees. Is a possible world then an actual thing which could become or could have been a world? There are no such things. There is only one world. Nor is there something else which could turn into a world. Further, whatever exists exists in this actual world. There is no room for any other worlds.

It might be thought, though, that talk of possible worlds is not talk of actual things that could have been worlds, but rather it is talk of certain merely possible entities.

However, this suggestion runs into two problems. The first is that one of the great advantages of the possible-world theory as usually presented is that it eliminates all reference to merely possible entities. The variables of standard quantified modal logic range only over actual entities. We do not have to accept any merely possible men in order to accept standard modal logic; why then should we have to accept any merely possible entities of the species "world"?

The other difficulty is that there do not seem to be as many possible worlds as the possible-world theory seems to need.

Suppose that standing here next to me is John Doe, who has brown hair. Consider that possible world which is just like this one except that Doe's hair is green. In that world, this man John Doe has green hair. In that possible world, John Doe stands with green hair in this very room, this actual room in which we stand. This actual room is found in that world containing John Doe with green hair. In that possible world, this very planet Earth on which we stand includes this very actual room in which stands this very actual John Doe but with green hair. Similarly, this very actual solar system, this actually existing galaxy in which we actually are is found in that possible world, but there it contains a green-haired John Doe instead of a brown-haired one, although that John Doe is this very one who stands before us with brown hair. Continuing in this way, it seems that in that possible world we find always this very actual John Doe, this very room, this very galaxy, and in fact this very actual world. We do not find another John Doe, another galaxy, or—and this is the point—another world either.

I submit that the ideas that there are such things as possible worlds or that it is necessary for us to refer to nonexistent merely possible worlds—these ideas have a considerable prima facie absurdity. I submit further that these ideas are puzzling in just the ways that we would have expected a good account of modal reasoning to clear up.

I agree with Stalnaker's view that just as a scientific theory is supported by being shown to account for observations, so a philosophical theory is supported in part by being shown to account for paradoxes and puzzles. But one does not show that a theory is able to account for a given apparently adverse observation by insisting that the theory is highly supported by *other* observations. One might justifiably argue that a given apparently adverse observation will probably turn out not to be as damaging as it seems, since one's theory is otherwise so highly supported. But then one is not responding to the apparently adverse observation but only arguing that a really responsive answer will probably eventually be forthcoming. Or one may justifiably admit that a given observation really does refute one's theory and nonetheless urge that an otherwise successful theory must still be close to the truth. But if one does this, one is again not *answering* the adverse observation. One has rather admitted its force and one is now talking about expediencies.

This is just the case in Stalnaker's discussion. The various victories of the possible-world theory do lend support to that theory. But they do nothing to show that that theory is able to respond to that particular puzzle that Stalnaker is supposed to be answering, namely, that the whole idea of possible worlds (perhaps distributed in space like raisins in a pudding) seems ludicrous. To respond to this puzzle it is irrelevant to insist on the fruitfulness of the possible-world theory.

But Stalnaker's confusion about this and certain other subtle methodological points is not what I found irritating. What I found irritating was the view which Stalnaker, befuddled by the above, seemed prepared to fall into. This view is that it does not *matter* that the possible-world theory might be falsely claiming the existence of nonexistent entities so long as the theory is fruitful. On the contrary, the question whether the theory is true is the is the *only* question that philosophically matters.

Here again there is a subtle point that needs to be made clear. It might seem that theories are sometimes accepted, and properly so, because they are useful. Let us however imagine a collection of cruel sadistic nogoodniks who like to kill and torture innocent people. Accepting an empirical theory T, they find that its precepts enable them to wreak unimaginable havoc on innocent people. We know nothing else about theory T. It might be said though that we have some reason to think that

theory T is true. And we do. That fact that T is so useful to these no-goodniks supports the theory T. But notice that we are not inclined to accept theory T for its *utility*. It has no utility for us. Rather the *fact* that it is so useful to those no-goodniks suggests that, unfortunately, it is a true theory. We sometimes accept theories because they are useful in explaining or predicting events or in accounting for philosophical puzzles. But this means not that we accept them for their utility and despite their palpable untruth, but rather that we accept them because *these* utilities are evidence for the only thing that matters here—their truth.

If there are no such things as possible worlds and if the possible-world theory says there are, then the possible-world theory must be rejected, however fruitful it is. Any discussion of Stalnaker's which prepares him to deny this obvious point must be fundamentally confused.

This is the end of my railing and ranting. I have, I suppose, spent a lot of time attacking what amounts to a single not very relevant sentence in Stalnaker's paper. He could easily erase this sentence. However, some ideas seem intrinsically worth stamping out, and that sentence seems to me to be trying to express one such idea.

It might be thought from the previous section that I have great misgivings about possible-world theory. Actually I do not; I just think that the phrase 'possible worlds' is misleading and overly colorful. It might also be thought that Stalnaker does not say anything responsive to misgivings about possible worlds. Actually he does, as it turns out, respond directly to such misgivings. On page 79 he suggests that instead of talking about possible worlds, we might talk of possible states of the world. This suggestion avoids, for instance, my "John Doe-in this room-in this galaxy-in this world" problem.

Yet on page 84 he makes another suggestion which seems to me to be helpful but also overdrawn. He says, "We need at least a rudimentary notion of alternative possible situations in order to understand such notions as belief and rational deliberation." He continues, "If this is right, then a notion of possible worlds is deeply involved in our ordinary ways of regarding some of our most familiar experiences." Apparently he thinks that when we ordinarily talk about alternative possible *situations* we are talking about alternative possible *worlds*. This seems clearly false. If I find myself in a bad situation, I

can act so as to get into a different situation. But it is impossible to get out of the actual world into some other possible world. If I know what situation I am presently in, I can act to get out of it. But if I knew which possible world (total state of affairs, past, present, and future) I was presently in, there would be nothing left to deliberate about. What we ordinarily call a "situation" is *not* a possible world.

Consider the concept of a possible lifetime fruit diet. To specify X's lifetime fruit diet is to specify all the pieces of fruit X eats in his whole lifetime. Now suppose someone offers me a choice between an apple and a banana. I am to choose between two possible (or available) pieces of fruit. Am I now choosing between possible lifetime fruit diets? Maybe the case could be conceptualized this way, but it is false that we ordinarily do conceptualize it this way. It is false that the concept of a possible lifetime fruit diet is deeply involved in our ordinary way of regarding choices between apples and bananas.

Admittedly all this cackle of mine about possible worlds is quite beside the main issues of Stalnaker's paper. I now turn to more substantial comments.

The most striking argument of Stalnaker's paper is the argument that a certain implausible theory of propositions and beliefs fits in surprisingly well with some very plausible ideas about deliberation and the role of beliefs in deliberation. I shall urge that this argument is fraudulent and actually adds no plausibility to the implausible theory about propositions and beliefs.

The implausible theory about propositions and beliefs contains the following claims. If p and q are logically equivalent propositions, they are one and the same proposition, and whoever believes p therefore believes q. Further, if p entails r, then whoever believes p believes r. Perhaps this last claim follows from the first. Whoever believes p believes $p.r$ and, if we assume that whoever believes a conjunction believes its conjuncts, then he believes r. In any case, Stalnaker's theory contains both claims.

Stalnaker says that these implausible ideas fit in well with certain plausible ideas about deliberation.

The first plausible idea about deliberation is that the notion of belief belongs to a theory intended to explain how rational creatures deliberate, and that moreover a belief is a state "which is defined or individuated by its role in determining the behavior"

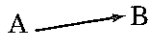
of the organism who has the belief. To Stalnaker this idea suggests that beliefs should be functions from possible worlds to truth values. To me, this idea suggests that, if beliefs are functions, they should be functions from desires to actions or from desires plus other beliefs to actions.

Let us try to work out a partial account of beliefs as functions to actions.

Organism X's belief that it itself is at point A will be the function which takes a desire to be at point A into an act of sitting and smiling and which takes a desire not to be at point A into an act of running away.

Its belief that it is not at point A will take a desire to not be at A into contented sitting and smiling and will take a desire to be at A into running around as if trying to get somewhere.

Now consider the diagram



X's belief that the arrow leads from A to B is a function which takes the belief that it (X) is at A plus a desire to be at B into the act of running along the arrow.

However, the belief that the inverse arrow leads from B to A is a quite different function. It takes X's belief that it is at B plus a desire to get to A into the act of running in the backward direction along the arrow.

Suppose now that X knows how to get from A to B but does not know how to get from B to A. Inferring X's beliefs from his actions (or propensities to act, given certain desires), we conclude that he believes that the arrow goes from A to B but does not believe (does not realize) the logically equivalent proposition that the inverse arrow leads from B to A. This example illustrates that anyone who really infers beliefs from actions will *not* say that whoever believes p believes everything that follows from p, or even everything equivalent to p. Notice further that it is wrong to say that X is not a rational agent in our example. X acts rationally in the light of his beliefs and his beliefs are furthermore consistent. His only failure is that he is not logically omniscient; he does not always know the *consequences* of his beliefs.

The above example was suggested by Jean Piaget's writings. Piaget's attempts to construct the child's conceptions of the world are real efforts to do what Stalnaker only pretends to do—namely, to envisage beliefs as functional states which help to explain actual behavior. According to Piaget, what he calls

reversibility and transitivity are difficult conceptual accomplishments. This means that knowing how to get from A to B does *not* automatically mean knowing how to get back, and that knowing how to get from A to B and knowing how to get from B to C does *not* automatically mean knowing how to get from A (through B) to C. Any serious attempt to define and/or individuate beliefs by their roles in action would lead in the opposite direction from Stalnaker's implausible theory of propositions and beliefs.

But Stalnaker does not even really make any such attempt. Rather he only asks himself what account of beliefs would fit with a certain "simple theory" of action-determination.

This simple theory amounts to the following. One explains why X did A, the theory says, by showing that X's beliefs *entail* that if X does A, his desires (or presumably at least one of them) will be satisfied.

In the first draft of Stalnaker's paper he suggested a different simple theory: namely, that we explain why X did A by showing that X's beliefs entail that if X's desires are satisfied, he does A. Stalnaker has now changed the entailed proposition from "if X's desires are satisfied, X does A" to "if X does A, X's desires are satisfied." In either theory, though, we explain why X did A by citing the entailments of X's beliefs.

Now Stalnaker claims that this simple theory is independently plausible—that is, plausible independently of the assumption that X's beliefs are deductively closed.

But is it? Suppose that X believes p, q, and r, and that p, q, and r entail that if X is to get his desires, he must do A, or that if X does A, he will get his desires. Suppose further that X does not realize that his beliefs have these entailments and therefore does not realize that doing A has any relation to his desires. Is it then plausible to explain his doing A in terms of these entailments? Of course not.

What is "simple" about either of these simple theories is that they both simplify matters by supposing that X's beliefs are deductively closed.

I agree that it is a plausible procedure for someone who wants to think about deliberation to begin by making some simplifying assumptions or idealizations about an agent's belief structure. Scientists quite plausibly begin with frictionless mass points and ideal gases. But it is ludicrous to argue from the plausibility of making idealizations to the plausibility of supposing these idealizations to be accurate realistic truths. It is not plausible

that gases are really ideal, that this ball of cotton is a frictionless mass point, or that the average agent's beliefs are deductively closed.

As a final critical point, I will comment briefly on Stalnaker's suggestion that his account can be squared with obvious facts about mathematical knowledge. I do not see how this is going to work. (I might note here that a certain penciled notation on Stalnaker's first draft suggested to me that he is not unaware of the objection I am about to raise.)

Stalnaker holds that everyone knows every necessary proposition. Therefore it implausibly follows that whoever knows all the axioms of a branch of mathematics knows also all the esoteric theorems. (Apparently it also follows that everyone actually does know the axioms and theorems, but we do not need this.) Thus mathematics really is surprisingly simple stuff! Stalnaker suggests that perhaps the paradox can be muted by supposing that nonetheless people sometimes know the truth of the axiom *sentences* without knowing the truth of the theorem *sentences*.

However, this is not going to work. It will not save us from mathematical omniscience to any interesting degree. Given a formal system, its axiom wffs, and its rules of wff-formation and derivation, the theoremhood or nontheoremhood of given wffs follows logically. Thus if I am logically omniscient, know the axiom sentences and rules of derivation and sentence formation of a given mathematical system, and if I am then given a theorem sentence, I will, as soon as I identify the sentence in question, know that it *is* a *derivable* theorem sentence.

I may sum up so far. The possible-world theory is, I think, true enough though misleadingly explained. Stalnaker's theory of propositions and beliefs seems to me both implausible and false. It cannot account for mathematical knowledge along the lines Stalnaker suggests. Nor does reflection on deliberation really lend any support to this theory. That this false theory may sometimes be a fruitful and useful simplification is no doubt true, but this fact is not enough to persuade me that the theory is true.

Let me now beat around the surrounding bushes. There seem to be three intriguing ideas left kicking around here. One is the idea, suggested to me by Piaget's work, that deepening understanding of a proposition is reflected by a readier interchange

of equivalents and by a faster motion along chains of consequence. Another is Stalnaker's suggestion that propositions are not fundamentally linguistic entities and beliefs not fundamentally propensities to utter sentences. The third is that we might try to give an account of how propositions should be individuated by considering their roles in action.

The second of these three ideas is that propositions should be explained without essential reference to sentences. Notice that my simpleminded account of X getting from A to B and not back again already had this virtue. We may therefore concentrate on the first and third ideas and hope that this second one will take care of itself.

Let us envisage a theory of belief structure that will do justice to the first idea.

Imagine propositions as entities laid out in some logical space. Each proposition is a dot. Each proposition has immediate consequences and is the immediate consequence of others. Furthermore, *sets* of propositions have immediate joint consequences. For simplicity let us ignore sets. Now let each proposition be tied by an arrow to each of its immediate consequences. Therefore it is connected by a path of arrows to its farther consequences, and thus ultimately to all of its derivable consequences. Similarly it is reachable along a path of arrows from any proposition of which it is a derivable consequence. Thus the arrows fix its provable logical relations.

So far I have not mentioned belief. It seems, though, that if q is an immediate consequence of p, then believing that p should entail some readiness to believe that q. Let us postulate that if X believes p and considers p and q together, he will come to believe q. Therefore, if he believes p, he will necessarily have a tendency to come to believe p's farther and farther consequences, if he considers appropriately.

At the same time, suppose he does not believe p. Suppose q is an immediate consequence of p, and r of q. Then if he considers p, q, and r together, r will, we now postulate, become an immediate consequence for him of p. In this way, X's considerings of propositions will lead to a deeper understanding of them, which will exhibit itself in an increasing readiness to substitute logical equivalents and to move quickly along chains of logical consequence.

This picture has the weakness that I needed to postulate an unclarified relation of original immediate consequence. But it has the nice feature that at least the provable logical relations

between propositions are fixed by X's propensities to believe—in other words, X's belief that p entails X's tendency to believe its consequences—without having the bad feature that X's beliefs are deductively closed.

Now let us turn to the idea that beliefs should be individuated by their roles in action determination. The great interest of this idea is that it seems to promise us a way to individuate propositions that would avoid on the one hand the bad idea that they are to be individuated just like sentences, and on the other hand the unwelcome idea that they are to be individuated à la Stalnaker by logical equivalence. As Quine has urged, the failure to give an account of how propositions are to be individuated is a great stumbling block to those of us who like to talk about propositions.

Unfortunately, the intriguing idea before us seems full of snares. In the first place, it does not seem that every proposition is a possible belief. There are many conceivable states of affairs in which I do not exist, but it is hard to imagine a deliberative context in which my action is predicated on my believing that I do not exist. Only an idealist could love a theory that makes my existence essential to every conceivable state of affairs.

More serious difficulties are lurking here. It seems that any action of mine could in principle be explained by saying that I believe that certain contingencies obtain between my so acting on the one hand and my having certain experiences of pleasure and painlessness on the other. Suppose I am running along and, just before slamming into a concrete wall, I skid to a halt. You say I halted because I believed that there was a concrete wall in front of me. But would my action have been any different if I had *only* believed that continuing to run that way would lead to a sudden sharp headache? It seems that the suggested way of individuating propositions threatens to identify the proposition that there is a wall in front of me with some conjunction of propositions about the contingencies between actions and experiential states. Only an idealist could be cheered by such a way of individuating propositions.

Of course it may be objected that not all desires are desires to have pleasures and avoid pains. I might for instance desire that there be a wall in front of me, for instance if I am building a house. Nor does it seem that desiring that p always amounts to desiring pleasures and believing that p's truth will give me these pleasures. I may desire to be buried in a certain place when I die, but I do not thereby think that I shall be happier buried

there. I may desire that an actual living woman and not a clever robot should think me a great fellow, but I suppose that I should be as foolishly happy if I were fooled by a cleverly designed robot which looked, felt, and acted like a beautiful woman and which made noises about my being a great fellow.

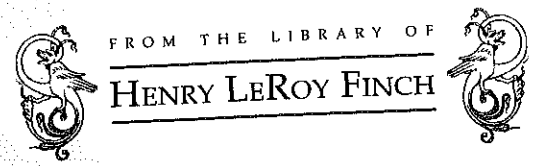
Unfortunately this line of objection makes the project of defining propositions in terms of their possible roles in determining actions seem *less* and not more promising. If I can desire virtually any p, and if I can believe, however falsely, that any given action will lead to that p, the prospects for deducing my beliefs from my actions seems dim indeed.

Alternatively, if basic desires are simple desires for pleasure and painlessness, then the project of defining propositions in terms of their roles in action seems nothing less than a project of describing a constitution of the external world out of experiential contingencies. That Stalnaker's theory of propositions does not face these extraordinarily large difficulties is, I think, simply an indication of the fact that he is not seriously attempting to define and individuate beliefs in the way he suggests. He is just whistling about it.

A FURTHER REMARK

One may wish to object to my A-to-B example along the following lines. X believes that he can move from A to B by running along the arrow but does not believe that he can move from B to A by running backward along it. These propositions are not equivalent. There might be a guard on the arrow who allows only one-way traffic, so that X can get from A to B along the arrow, but cannot get back the same way. However, this still does not explain why X does not at least try going back along the arrow.

It is just this type of difficulty with really working out my example which leads me to think that only action-contingency statements can be plausibly derived from propensities to act. If my example is not an example of failing to realize that the inverse arrow goes from B to A, what on earth would be? Further, note that this kind of objection works only against reversibility problems and does not help against transitivity.


FROM THE LIBRARY OF
HENRY LEROY FINCH