ON THE PHILOSOPHICAL RELEVANCE OF
POSSIBLE—WORLDS SEMANTICS

Robert Kraut

The philosophical community is far from unanimous in its attitude towards possible worlds semantics. One faction, with Jaakko Hintikka as its most creative and prolific spokesman, regards the possible-worlds apparatus as helpful for achieving clarification, illumination, and perhaps solution to various traditional problems — problems about modality, propositional attitudes, and the objects of such attitudes. Yet there is an opposing sentiment, one which is skeptical that anything philosophically valuable can be achieved with the aid of such suspicious entities as worlds, world lines, propositions, and the like. Completeness results and formally adequate truth theories for intensional languages are one thing; but perennial problems are not to be illuminated by supplanting them with even greater problems.

What we need, I think, is some *interpretation* of the possible-worlds apparatus which would give it a more intuitive grounding, thus rendering it intelligible to someone other than a modal logician. I will suggest that possible-worlds talk can be profitably regarded as *metalinguistic* talk, as a codification of data involving the roles played by bits of language, or data involving behavioral dispositions. I am *not* suggesting that we treat model connectives as predicates of sentences. Montague showed in 1963 that such a treatment, if applied to a language which contains elementary arithmetic and some plausible looking modal axioms, would result in inconsistency.[1] For my purposes we can treat modal connectives as propositional operators, and embrace the familiar semantical theories of the sort proposed by Hintikka, Thomason, Kripke, and others.[2] The goal here is to provide an interpretive schema for the model-theoretic apparatus itself, a schema which might, in Thomason's words, "...provide an informally satisfying

explanation of what modality is about...."[3]

The interpretive strategy outlined here might strike some as circular; for I exploit counterfactuals in my explication of possible worlds, yet the most satisfying semantical treatments of counterfactuals available thus far take possible worlds as a primitive notion. To such an accusation of circularity, there are several responses. One is gallantly offered by Hintikka :

> ...a circle of explication need not be a vicious one, provided it is wide enough to enable a logician to uncover nontrivial aspects of the structure of the concepts involved.[4]

Another response is to seek an analysis of counterfactuals which does not make essential use of possible worlds. Alternative treatments have, after all, been proposed : e.g., metalinguistic or "inference ticket" treatments of various sorts.[5] Granted, David Lewis has offered convincing criticisms against such alternative treatments;[6] but the issue is not dead yet, and those who find cogenial the strategy which I suggest here might be provided with additional motivation to attempt a revival of these alternatives.

In Part 1 I will outline and motivate an interpretation of the possible-worlds machinery, with special attention to some typically mysterious sounding locutions which often emerge in the literature. We will then see what light, if any, is thereby shed on the alethic modalities. Part 2 briefly explores the relevance of possible-worlds semantics to discussions about mental states.

<div align="center">1.</div>

We need a specimen possible-worlds semantics to serve as our point of departure; here we mention the salient points of Hintikka's approach, which will serve the purpose admirably.

Let K be a set of possible worlds, and for each world $w \in K$ let D_w be the domain of entities associated with w; a function f is said to be a *world line* iff for each $w \in K$ at which f is defined, $f(w) \in D_w$. Thus a world line is a (possibly partial) function which correlates with each world an entity present in that world. Truth of a sentence of the form $\Box Ft$ amounts to truth of Ft in all members of K. In the process of evaluating the truth of a modal formula, then, we are concerned with the reference of the constituent singular terms in a multiplicity of worlds. Thus in modal contexts we have not a failure of referentiality of singular terms, as some have thought, but rather a *referential multiplicity* :

> Not every free singular term (say, a) picks out from these possible worlds manifestations of one and the same individual. In

fact, usually its references in different worlds are manifestations of different (well-defined) individuals. Such a term does not specify a well-defined individual in the context in question. For this reason we cannot instantiate or generalize with respect to such a singular term.[7]

What matters for the present purpose is that Hintikka speaks of the world lines *themselves* as the individuals with which we are concerned in modal contexts :

...each individual in the full sense of the word is now essentially a function which picks out from several possible worlds a member of their domains as the "embodiment" of that individual in this possible world...[8]

There functions, in a sense, are thus the real individuals we are talking about in our sentences...[9]

Consider, then, the truth conditions for $\Box Ft$. Corresponding to t will be a world line, say f_t. The modal sentence is true iff for each $w \in K$ at which f_t yields a value, $f_t(w) \in V_w(F)$ [where $V_w(F)$ is the extension of F relative to w]. An existential sentence, say $(Ex) \Box Fx$, will be true iff for some world line g defined in the model, $g(w) \in V_w(F)$ for all $w \in K$ at which g yields a value. Relating this to ordinary discourse, consider the ascription of an essential property to Jones, say rationality. On the present semantical approach, the truth of "Jones is essentially (necessarily) rational" demands that the world line correlated with "Jones" yields at each world (at which it yields any value at all) a value which has the property of rationality in that world.

Most entities with which we are concerned are not necessary existents. We can envisage a world, e.g., in which Nixon does not exist; the formal representation of this is the failure of the world line correlated with "Nixon" to yield a value at that world.

Though we have here the outline of what might appear to be an adequate semantics for alethic contexts, when suitably spelled out, certain considerations constrain Hintikka to throw in the towel and claim that "quantification makes no sense in a context of logical modalities; ...Quantifiers and *logical* modalities simply do not mix."[10] But why ?

On Hintikka's approach, quantification into modal contexts is dependent upon a system of world lines; he feels that, in the case of the alethic modalities, "...we cannot answer the questions that have to be answered in order for quantification to make sense";[11] specifically, we need data which enables us to extend a world line through an arbitrarily selected world, or which enables us to determine that the line yields no value at that world — data which we

very often do not have. So the slogan behind the pessimism is "World lines cannot always be continued". As Hintikka puts it, "...when trying to extend a world line of an individual i to a new world w," we run into cases in which "It makes no sense to ask whether i exists in w or not. (The candidates for the role of i are not narrowed down to one at most, or are not well defined at all.)"[12]

Possible-worlds theorists are fairly comfortable with such talk. But traditionally oriented philosophers are often inclined to ask "What does all of this *mean* ? And how, if at all, does it illuminate the "meaning" of modal claims ? " One strategy at this point is to ignore them, or to accuse them of lacking formal sophistication. Another strategy is to try to answer them.

World lines are functions from worlds to extensions. Perhaps something further can be said about them. Imagine a behaviorally minded jungle linguist, interested in charting the stimulus meaning of some expression, say "Cayster". He keeps a running list of the stimulus conditions which prompt assent to "Cayster ? ", and of those which don't. On the basis of such evidence, he offers a hypothesis as to which conditions *would* prompt assent, and which would not. With any luck (and some skill at projecting general hypotheses) he would construct a reliable picture of the stimuli in response to which the expression is tokened, given various possible conditions. What the linguist wants to determine, we might say, is the *role* or *function* of the expression in the native's community. By analogy, we might depict the role of a piece of wood in a certain game by indicating how it would be proper to move the piece, given various ways the game board could be arranged. But when the game involves moving pieces of language around in response to the environment, how should we describe the game board ? Notoriously, the stimulus conditions surrounding a speaker can be characterized in terms of whatever theoretical framework seems most useful; our decision as to what counts as an input to the language user is as subject to revision and confirmation as is any other part of our semantical theory. The native can be described as tokening expressions in response to mass-point aggregates, local curvatures in space-time, or trees and rabbits. For the present let us describe his environment in terms of ordinary macro-objects. Then we can depict the stimulus meaning of an expression by speaking of a multiplicity of possible stimulus conditions and the region(s), if any, in each such condition to which the term is properly applied. Our final diagrammatic representation of such data takes the form of a function from possible stimulus conditions, represented as sets of

entities with relations defined on them, to local regions within those stimulus conditions, represented as subsets of the domains associated with each situation.

Let us call one of these possible stimulus conditions envisaged by the linguist a *possible world;* call the portion of one of these scenarios to which the speaker applies some expression t, the *denotation of t in that world.* And call the function which maps each possible scenario w onto that portion of w to which t would be applied by our speaker, were he there, the *world line* corresponding to t. Such a function codifies the range of phenomena to which t would be a proper response. The world line thus represents the stimulus meaning, or what Sellars would call the *language entry role,* of t.[13]

Questions might arise as to what the denizens of our worlds ought to be; this is but the question of how to characterize the domain to which the speaker is responding. If it behooves us to describe the input to the native in terms of, say, quarks, then each world will consist of a set of quarks with relations defined on them.

Questions might arise as to whether a denizen of one world is identical with a denizen of another. But when would such a question arise ? Surely in the initial stages of the translation enterprise the linguist can tell, in a relatively unproblematic way, whether the entity which ostensibly elicits assent to the query "t ? " is the same the entity which did so on another occasion. But he also asks whether there might be situations in which the native *would* apply the expression to either a numerically distinct entity, or to the same entity distinct in certain respects, from that which has prompted assent to the query thus far. At most, the linguist wonders whether there is a possible stimulus condition (one which might well confront the native before the linguist is finished with him) in which an expression is applied to an entity of such-and-such characteristics; how far would the native be willing to go in his application of the expression; given his present speech dispositions ? This is a question about the role played by the expression, about whether the rules governing its use are such as to allow it to be applied under certain specific circumstances.

Questions might arise as to whether a certain entity exists in a certain possible world. Let us see how such a question might profitably be construed.

Puzzles involving the behavior of the good old predicate 'exists' are well known. A careless semantical treatment of "Socrates does not exist" might involve locating a member of the domain, the referent of 'Socrates', and determining whether or not it posesses the

property of existence. But what if it doesn't ? What an odd object it must be, there on the one hand but non-existent on the other. Is there something whose existence we are denying ? Very odd. "This is the old Platonic riddle of nonbeing. Nonbeing must in some sense be, otherwise what is it that there is not? "(Quine) Such puzzles are probably misguided, generated by a bad semantics for existence talk. Several recent discussions of these issues share a common insight, an insight which might illuminate talk about existence in a world.

Consider the use of the expression 'Socrates'; there are circumstances under which it would be proper to apply it to some portion of the passing show. Given the role that the expression plays, it simply has no proper application as things stand. Perhaps we can say that "Socrates doesn't exist" is a claim about the role played by the expression 'Socrates', that any expression which plays that role has no proper application.

This is a strategy occasionally hinted at by Wilfrid Sellars.[14] Call the conditions on the proper applicability of the expression to the environment the *language entry rules* governing the expression. Of course, there may be items in other languages which are applicable where and only where the expression 'Socrates' would be applicable. Following Sellars, call any item which plays in its own language the role played by 'Socrates' in our language a ·Socrates·. Thus ·Socrates· is a general term applicable to any linguistic item which is functionally equivalent to 'Socrates'. Call this role an *individual sense;* and when a ·Socrates· is properly applicable to some portion of the environment, let us say that the individual   sense ·Socrates· is realized.

On such an approach, the assertion that Socrates doesn't exist is an assertion that no expression which is a ·Socrates· is properly applicable. Existence contexts thus emerge as metalinguistic, insofar as the analysis takes the form

> "a exists" is true iff the individual sense ·a· is realized

A similar approach, couched in a slightly different idiom, is taken by Pavel Tichy.[15] He suggests that we view intensions or senses as search procedures, or, better, identification procedures, which are applied to a multiplicity of individuals and which select one of them (if it is an individual sense) or several of them (if it is a general sense) as its output. On this view, to say that Socrates exists is but to say that the Socrates search procedure yields an output. We can conceive of circumstances under which it would yield an output, and those under which it would not. The symmetry between such an approach

and the previous one is patent, for the Socrates search procedure yields an output where and only where a 'Socrates' is properly tokenable.

Consider now a third approach, which has unfortunately received far less attention than it deserves. Suppose we were to read "Socrates does not exist" thusly : "You know what it would be like to encounter Socrates; you would stumble across a person who at some time past had stood as a link in an appropriate causal chain leading to our use of 'Socrates'. Let us play the Socrates-finding game, given that we know the rules. It is fairly certain that we will not win. No winning strategy is available, as things stand, for that game." Such intuitions are closely related to Hintikka's suggestions about a game-theoretic interpretation for first-order quantification theory, an interpretation which serves to tie the concept of *existence* to the concepts of *searching for* and *finding*.[16] Since his work in this area is relatively unfamiliar to philosophers, a few words about its rationale are in order. The proposed game-theoretic semantics in a partial response to Quine's claims about the indeterminacy of ontology. Truth functional connectives, according to Quine, admit of a reasonably determinate translation from a first order canonical idiom to an alien language. Quantificational expressions, however, admit of a less direct, and hence less determinate, translation. What accounts for this asymmetry, Hintikka suspects, is Quine's having selected assent and dissent as the priveleged modes of behavior in terms of which to gain information about the roles played by the natives' expressions. Hintikka finds such a choice highly restrictive, since (as Quine acknowledges) other modes of behavior, no more difficult to recognize than assent behavior, bear upon what is empirically determinable in language. Specifically, Hintikka wants to exploit the behavior which is characteristic of searching for and finding, and spends some time arguing that the role of quantificational expressions can best be understood against the backdrop of such activities. The connection is made precise by constructing a semantics for first order languages in terms of such search and find games : each quantified formula is associated with a two-person, zero-sum game played with perfect information. The formula is true iff there is a winning strategy available in the associated game, false otherwise. An existential quantifier marks my move; I am to select from the environment an object which satisfies the open matrix. Thus the force of "(Ex)Fx" is, intuitively, "There is a winning strategy available in the F-finding game". Such a claim demands that I know the rules of the relevant game, since I must know what is to count as a winning move. The end point of the game, the

culmination of the search, is signaled by a kind of recognizable behavior; it may be accompanied by words ("Now I've got one ! ") or by some other functionally equivalent behavior. The resulting treatment, if ultimately satisfactory, results in a conceptual tie between the concept of existence and that of "being able to be found". As Hintikka says, "A first order sentence becomes, as it were, a kind of prediction of what may happen in the game associated with it."[17]

This is not the place to attempt a thorough examination of Hintikka's proposal. Aside from its *prima facie* circularity (existence claims are explicated in terms of the existence of winning strategies)[18] we would have to see precisely what consequences the proposal has for the Quinean program. But my goal here is to note the strong affinity between Hintikka's strategy and that pursued by Sellars. The production or location of an individual qualifies as a winning move in the Socrates-finding game iff a 'Socrates' is applicable to it; and the tokening of a 'Socrates' is surely a paradigmatic end point in a Socrates search. Similarly, "There is a golden mountain" is true just in case there is a winning strategy available for the "golden mountain-finding game". But what are the rules of this game ? What are the circumstances under which we can properly say "Now I have found one ! " ? They are precisely the circumstances under which the expression 'golden mountain' is properly applicable to the environment. Thus an intimate tie emerges between finding behavior and language-entry move behavior (that is, the production of linguistic activity in response to environmental input).

Let us say, then, returning to our initial quarry and summing things up, that "Socrates exists in a world w" can be construed as

    a) a 'Socrates' is properly applicable to some portion of stimulus condition w (in the spirit of Sellars)

    b) the Socrates-finding game has a winning strategy in the field of search represented by w (in the spirit of Hintikka)

    c) the Socrates search procedure yields an output when applied to w (in the spirit of Tichy)

We noted earlier that stimulus meaning is cashed out in terms of a multiplicity of possible stimulus conditions. And we use an analogous method for the characterization of a particular game of search and find. We want to approximate the set of all conditions

which the native would treat as winning moves, the set of all stimuli which would elicit finding behavior. Take, for example, the game of woman finding; it is distinct from the game of Sarah Jones finding : in the former game, any ostension of any woman would qualify as a winning move; in the latter game, the only way to win is by ostending Sarah. Thus the set of stimulus conditions which would prompt finding behavior is distinct for the two games.

Earlier we interpreted a world line as representing the stimulus meaning of an expression. What emerges here is that we can also regard a world line as representing the rules of a specific search and find game, where the *values* of the line at a world represent the winning strategy (or strategies) in that game relative to the stimulus condition in question.

The cash value of these intuitions regarding the interpretation of our semantical apparatus is straightforward. Corresponding to each search and find game, or to each linguistic role that we might be concerned with, will be a function or world line with possible stimulus conditions as its domain; the existence of Socrates at w, e.g., will be represented in terms of the corresponding world line (that which codifies the stimulus meaning of 'Socrates', or likewise the rules of the Socrates-finding game) yielding a value at w. These world lines are partial functions — there are worlds at which they yield no value. And this corresponds to the simple fact that there might be circumstances in which certain lexical items are inapplicable to the environment, given the role that they play.

A world line which yields no value at the actual world (as happens, e.g., regarding the function correlated with the 'golden mountain') is occasionally rolled into a ball and called an *unactualized possible*. Philosophers are left to ponder the states and legitimacy of such "entities". Yet, mobilizing our interpretive resources, it emerges that

the golden mountain is an unactualized possible

which corresponds to the model-theoretic

the world line denoted by 'the golden mountain' yields no value at the actual world, though it yields values at alternative worlds

has as its "interpreted" counterpart the more intuitive

the golden mountain· is not realized, though it could be

or, from the equivalent game-theoretic standpoint,

the golden mountain-finding game has no winning strategy available in the actual field of search, though it has in another field of search.

Whether this simply replaces oddities with oddities I leave to the reader to decide. I, for one, would rather ponder the dynamics of a loosing game than wonder about the size, or lack thereof, of an ethereal entity which I can't quite get my hands on. And there is no model-theoretic price to pay for this preference, for we can continue to think of constants and variables in modal contexts as Hintikka does, as designating world-lines.

Consider now the "problem of identifying individuals across worlds"; as much current sentiment has it, the problem constitutes adequate reason to steer clear of possible–worlds semantics. But perhaps the problem has its source elsewhere, and is merely being brought to a sharper focus by the apparatus in question. This, at any ~ate, is what I will suggest.

Suppose that the reference of t in a world is simply that hunk of the passing show to which t would be properly applicable, were that show to take place. The problem of identifying Jones in a given world then becomes the problem of determining whether it would be proper, given the stimulus condition in question, to apply a ·Jones· in that situation, and, if so, the precise portion of the environment to which the expression could be applied. Such a question demands our resolving questions about the role of the expression, about its proper use. This is often very difficult; often it calls for decisions (as certain embarassing questions about persistence tend to suggest.) When one suggests, as Kripke occasionally does, that something in a world is Nixon iff it has the same origins as does the thing we call 'Nixon', one can be construed as stressing that origin and continuity conditions enter into the rules for properly tokening 'Nixon'.[19]

Of course, Kripke has inveighed against the "telescope view" of possible worlds, according to which worlds are something akin to distant cities, described solely in terms of properties, and we must look through our telescopes and determine, given the data about how properties are distributed, which person, if any, is Nixon.[20] This is the wrong picture, Kripke says — possible worlds are *stipulated,* not discovered, and we just specify that *there* is Nixon in that world, never having been President, yet Nixon nonetheless. So Kripke suggests that the problem of trans-world identity arises from an incorrect way of looking at possible-worlds talk.

However, Kripke's observation does not eliminate the traditional

problem. *Now* the puzzle becomes "What are the *constraints* governing the stipulative process ? " Can I stipulate a world in which the number seven is three feet wide ? Can I stipulate a world in which Nixon is born a woman, of completely different parents, with completely different genetic structure, and born at a radically different point in space-time, without doing injustice to our linguistic conventions governing the expression 'Nixon' (as that expression is presently used) ? My goal here is not to answer such questions, but to suggest that they are a notational variant of the "telescope man's" questions. And, stated *either* way, the questions are resolved by considering the roles played by the singular terms in question. The more difficult problems of trans-world identity, then, emerge as problems generated by the open texture of singular terms. Is it correct to token a ˙seven˙ in response to a spatially extended and piano shaped object (thus allowing for a possible world in which the number seven is a piano) ? Or to token a ˙Nixon˙ in response to a person of toally different origin than the man who in fact preceeded Ford as President ? Just how far do the rules of the language enable us to go in tokening a ˙Nixon˙ ? I am suggesting that it is this very question which is misleadingly stated (in the "material mode") as "Which things in alternative worlds are Nixon ? " It is often a very difficult question. But the possible-worlds enthusiast is no worse off than anyone else in this regard.

Let us now see what light our metalinguistic interpretation of the possible-worlds idiom can shed upon the modalities. Suppose that someone who countenances de-re modalities claims that Walter is essentially rational. What can this mean ? Given standard treatments, we might propose truth conditions such as

(1) Walter is rational in every possible world in which he exists.

But our interpretive strategy enables us to go further. A world in which Walter exists is but a situation in which the Walter-finding game has a winning strategy, i.e. a situation in which a 'Walter' is properly tokenable. There are, of course, conditions on the proper use of a ˙Walter˙; these conditions correspond to the rules of the Walter-finding game. (Since there are many people named 'Walter', there would be a multiplicity of such games; we would have to specify which one we intend. Thus for one such game, a condition on winning might be that we locate the person haptized 'Walter' at time t and place p.) (1) indicates that any portion of the environment to which a ˙Walter˙ is properly applied is also one to which a ˙rational˙ is

properly applied. Hence, claims about the essential properties of b invite us to consider the rules governing the proper application of a ·b·. Game-theoretically, (1) indicates that a win at the rational-finding game is prerequisite for a win at the Walter-finding game. Ascriptions of essential properties invite us to consider the rules of specific search and find games.

The conception of modal claims which emerges in the light of this interpretive framework is not radically new. Goodman e.g., says of the distinction between "natural kinds" and other classes

> Philosophers often hold that members of a favored class share some real attribute or essence, or bear some absolute resemblance to each other. I think the distinction depends rather upon linguistic habit.[2][1]

As it turns out, attributions of essential properties do indeed emerge as being claims about "linguistic habits". Consider a law-like claim such as

(2) Whales are essentially mammals

(2) is true iff any situation licensing the tokening of a ·whale· would also license the tokening of a ·mammal·; or iff a win at the mammal-finding game is prerequisite for a win at the whale-finding game. Recall that the usual semantical treatment of (2) dictates the truth conditions

(3) Any whale is mammalian in every world in which it exists.

There is nothing wrong with such a semantical treatment; I have not offered any alternatives. But by further analysing the notions of *existence* and *possible world,* we have transfigured the likes of (1) and (3) into claims which deal with the roles played by bits of language, or with the rules of the search and find games people play. All we have done is to provide a further interpretation of the model-theoretic concepts, so that we can continue to use the possible-worlds apparatus without falling prey to misguided ontological objections. Crudely put, the roles played by lexical items tend to *overlap;* in many cases an item which plays one role (e.g., that played by 'triangle') is applicable only when an item playing another role (e.g., that played by 'three sided') is applicable. It is this overlap of roles which is called attention to in claims about essential properties, as is reflected in the possible-worlds apparatus.

Earlier we spoke of Hintikka's pessimism about quantified alethic modal logic; it is interesting to see what his qualms come to in the light of the present framework. His major misgiving stems from the claim that "World lines cannot always be continued". More specifically,

> ...there may eventually be no trans world hier lines [i.e., world lines] left when all sorts of weird worlds are admitted into our class of possible worlds. In less metaphorical terms, if the world lines are drawn on the basis of similarities between worlds and of their intrinsic regularities, then this whole enterprise will break down when very dissimilar and irregular worlds have to be considered.[22]

This is still stated in fairly metaphorical terms. As it turns out, Hintikka's criticism does not appear all that cogent when further interpreted.

On our present approach, the world line of the individual Caesar, e.g., is a representation of the role played by the individual designator 'Caesar'; the "inability to extend Caesar's world line into a world w" is thus tantamount to an inability to say how, if at all, a ·Caesar· would be properly applied in stimulus condition w. This is not merely an *epistemic* inability, one which could be remedied by a good does of omniscience. Hintikka seems to be envisaging situations which are so *very* peculiar that the rules governing the proper use of an expression (or, if you prefer, the rules according to which a certain search and find game is played) are totally inapplicable; thus

> ...we run into possible worlds that simply are so irregular that our customary methods of cross-identifying individuals...may simply fail.[23]

Perhaps this can be made clearer with an example. Consider the conventions governing the tokening of a ·Caesar· — or, equivalently, the rules of the Caesar-finding game; let us simplify things here and say that the game is won by locating a member of the field of search which is space-time continuous with the hunk of matter-time which triggered off the original "baptismal tokening" of 'Caesar'. This is, of course, not an *epistemic* criterion, because space-time continuity is not the sort of property which is readily observable (not only do individuals "not carry their names on their foreheads", they do not carry their birthdates or birthplaces either). But there is some *fact* of the matter as to whether a certain candidate does stand in the requisite continuity and similarity relations. Hintikka is calling attention to cases in which these continuity relations are no longer defined (worlds in which "...our laws of nature [which] serve to guarantee those continuity properties" simply are not in effect — a

phenomenon which we will encounter in certain "distant" worlds, since those laws do not hold as a matter of logical necessity).[24] In such cases, the ordinary rules for proper application of certain singular terms break down; this is in turn represented as a breakdown in the world line of the corresponding individual. Such considerations, Hintikka thinks, ultimately suffice to vindicate Quine's criticisms of quantified modal logic, at least as applied to the alethic modalities. But let us evaluate these considerations.

The first thing to note is that such "breakdowns" occur even in connection with the temporal modalities, about which Hintikka is more optimistic. Take, for example, the familiar puzzle about the reconstructed ship of Theseus, recently discussed by Roderick Chisholm (who directs his attention to the U.S.S. South Dakota);[25] each day we remove an old wooden plank from our ship, replacing it with an aluminum one. All of our discarded wooden planks are gradually gathered up by some prankster, and gradually glued back together in the form of the original ship. When our handiwork (and his) is done, which of these vessels is *really* the U.S.S. South Dakota ? The aluminum atrocity ? The reconstructed wooden ship ? And is there really any fact of the matter ? We have watched the process from the very beginning; no amount of continuous observation, no resort to causal considerations, seems especially helpful. So how do we answer the question ?

What we need is a *decision,* a ruling as to which structure the expression "U.S.S. South Dakota" ought to be applied to, which structure the captain ought to go down with, which structure is the property of the original owner of the U.S.S. South Dakota, and the like. The conventions governing the application of a singular designating expression have failed us, given the odd state of affairs which has come to pass. So we must examine the conventions very carefully, and extend them if need be.

If we portray this situation formally, with our "possible worlds" representing successive momentary stages of the actual world, we would reach an embarassing moment at which "The candidates for the role of i [let i be the U.S.S. South Dakota] are not narrowed down to one at most, or are not well defined at all."[26] But this simply reminds us that our conceptual machinery is not equipped to deal, in an unproblematic way, with every circumstance which might develop; the question here is the extent to which this vitiates quantified alethic logic.

Granted, the alethic cases which Hintikka hints at are a bit more bizarre; not only can we not find the U.S.S. South Dakota; we cannot find the dock, or the water, or anything else, since all the

laws of physics have somehow broken down. But the difference between this and our temporal case seems simply one of degree, and at any rate not sufficiently clear cut to sanction Hintikka's consequential verdict.

Possible asymmetries between temporal and alethic modalities notwithstanding, Hintikka's criticisms are still not all that compelling. Consider again the imputation of essential rationality to Walter; our interpretation leads to truth conditions of the form

(1') Every possible situation in which a ·Walter· is properly tokenable is one in which a ·rational· is properly tokenable

or, equivalently,

(1″) No member of the field of search can be a win at the Walter-finding game unless it is rational.

Why does it matter that there are possible situations in which the applicability of a 'Walter' to the environment is problematic ? The claim is simply that *once* you have found Walter, you have *ipso-facto* found an instance of rationality. The essential attribution is a claim about the rules governing a certain search and find game; those rules can be clear cut despite the existence of possible situations in which the game cannot sensibly be played.

The game theoretic conceptualization suggests a certain analogy : consider a situation in which Armageddon has occurred, thus leaving everything in a vaporous state. Were two persons to be dropped into such conditions, they would have great difficulty playing chess. It would not be clear how to proceed, given that there is nothing available to serve as pieces or a board. But it is nonetheless true, even there, that checkmating one's opponent is prerequisite for a win. That is simply how the game works. And, similarly, if our essential attribution concerning Walter is true, it is simply a requirement for a win at the Walter-finding game that the candidate be rational.

It is no problem for quantified alethic modal logic, then, that "Worldlines cannot always be continued"; it is no problem formally, for even if our world lines are undefined at certain bizarre worlds, an adequate semantic theory would demand that we concern ourselves with only those worlds at which the world lines yield clear cut values. And it is no problem philosophically, for the limitations and the "open texture" of our conceptual apparatus is something we have all learned to live with by this time.

2

Hintikka's work in the semantics of propositional attitude ascriptions has not been properly appreciated. It has touched off countless discussions about "virtual knowledge" and "logical saints" (issues involving the closure of epistemic alternatives under logical consequence), and discussions about "knowing that one knows" (issues involving the appropriateness of a transitive accessibility relation in models for psychological language). But its relevance to discussions about the proper philosophical analysis of psychological states has not been explored. Many writers do not expect Hintikka's work (and other work along similar model-theoretic lines) to be very informative here — thus Føllesdal :

> Of course, this interpretation in terms of 'possible worlds' does not tell us what a knows, nor give us much of the full meaning of the word 'knows'. It is merely an algebraic, or model-theoretic device, which helps to clarify some of the formal properties of the verb 'knows'.[27]

It strikes me, however, that Hintikka's remarks about the semantics for psychological contexts bear on less formal considerations about the mental — as offered, e.g., by Sellars, Putnam, and others.[28] Let me briefly indicate the kind of relation I have in mind.

Suppose that some variant of functionalism, e.g., that outlined by Putnam or Lewis, is true. In saying that Pedro has a certain fear, then, we are saying among other things that Pedro will produce certain outputs, *ceteris paribus,* when presented with certain inputs. Taking a cue from the behaviorists (a cue which every functionalist seems willing to take), we can isolate certain output as characteristic fear output. If we view possible worlds as possible stimulus conditions, we can select out those specific stimulus conditions which generate fear behavior in Pedro — call such stimulus conditions *fear worlds.* Then perhaps we can say : "Pedro fears Russell" is true iff every world in which Russell is present and causally affecting Pedro is *(ceteris paribus)* a fear world. (Note here that spelling out such a *ceteris paribus* clause is something which would have to be done whether or not we were concerned with formal semantics).[29] At any rate, this simple truth condition requires further commentary : Pedro might *fear philosophers* in general, and this is a quite distinct state from his fearing Russell, though of course the states are intimately related. If Pedro fears philosophers then every world in which Russell is present (assuming Pedro knows Russell to be a philosopher) is a fear world; but then what is the difference between fearing Russell and fearing philosophers ? Perhaps in the

latter case there must also be fear worlds in which philosophers other than Russell are present.

Considerable juggling would be required to come up with an intuitively sound truth definition. Each new modality would require different considerations. But note the *tie* that emerges — for in specifying the truth clause for "P fears x", e.g., we would in effect be indicating the machine table, or functional state, which corresponds to the psychological state of fearing x. Thus those concerned with the actual details of a functionalist theory of mind, and those concerned with a possible-worlds semantics for propositional attitudes, would emerge as chasing a similar quary.

Let us apply these intuitions to an intentional attitude with which we were concerned before, viz. *looking for* or *searching*. Suppose that Jones is looking for an honest man — any one will suffice — and that Smith is looking for Father Murphey, who just happens to be the epitome of honesty. How can we characterize the distinction between their attitudes without invoking any peculiarly *psychological* entities (contents of mental acts, immanent objects, and the like) ? We might think of our friends as involved in the playing of certain games; in the former game, any ostension of any honest man would qualify as a winning move; in the latter game, the only way to win is by ostending Father Murphey. As before, we characterize the games in question by spelling out the stimulus conditions which would qualify as end points of search — and note that the set of stimulus conditions which *would* prompt finding behavior *(ceteris paribus)* is distinct for the two games. The point is that the *object* of search ("what is being searched for") can be characterized without immediate reference to any *mental* notions — all we need speak of is the range of phenomena which would trigger the behavior which is conceptually related to the attitude of searching (i.e., finding behavior). More bluntly, we can just as well speak of what would qualify as a successful end point of a game which is being played, as speak of what someone is searching for.

But what is to be gained from such conceptualizations ? A few transformations might make it more evident :

(1) Jones is seeking an honest man (any one will do)

(2) Jones is playing a game which has as a successful culmination any stimulus to which a ·honest man· is properly applicable

(3) All possible situations in which there is an honest man present to Jones are situations in which Jones wins.

Call a stimulus condition which would elicit finding behavior in Jones a *find alternative* to the actual world. If Jones were seeking only one thing in his life, we could propose as truth conditions for (1) something like

1T. Some honest man or other is present in each of Jones' find alternatives.

But perhaps we want to allow for the possibility that he has other goals besides, thus demanding that some of his find alternatives have no honest man present in them — this suggests

1T'. Every world in which an honest man is causally affecting Jones is a find altern   ve.

Notice what has happened. Arguments about the proper analysis of a certain psychological state, and about the kind of behavior symptomatically related to that state, have been transformed into questions about inclusion relations among worlds (finding alternatives, worlds in which honest men are present to Jones, etc.) I have not yet offered any final, rigorous analysis. Nor will I do so here.[30] For my methodological point has already been made. A traditional philosophical inquiry has been recaste into one about relations between various kinds of worlds. Nothing has been lost — we have simply siezed upon apparatus which enables us to discuss dispositions. And much has been gained, for our discussion is now mobilizing resources which can be used to provide a rigorous semantical treatment for various kinds of psychological contexts.

Of course this bears upon some traditional ontological issues. Suppose that Jones is looking for a unicorn. There are no such things. Must we therefore postulate some suspicious "content of a mental act" to which Jones is related, in the process of spelling out the truth conditions for our description of Jones ? No. All we need say is that each possible situation in which a unicorn is present to Jones is a find world (relative to Jones and the actual world). Moreover, extensional equivalence no longer poses a problem — for while Jones seeks unicorns, Smith seeks a centaur. If we refuse to postulate mental entities, must we conclude that our friends are embarked upon the same enterprise (given that 'unicorn' and 'centaur' are extensionally equivalent) ? Clearly not; for Smith's finding behavior would be elicited in situations in which Jones' would not. Some possible stimulus condition, e.g. one containing a unicorn but no centaur, would make one a winner and the other not. So the difference between their enterprises is semantically

representable in terms of the distinct sets of find worlds to which they are related. Subsistent unicorns and subsistent contaurs, the putative entities traditionally invoked to account for our friends' disparate hunting patterns, lose their hold upon us, as well as their passport to even second class status ("subsistence"). All that is required is that we take dispositions seriously. And invoking the possible-worlds apparatus does exactly that.

The Ohio State University

NOTES

*Thanks are due to Joseph Camp and Tom Vinci, for helping me to avoid some fairly serious blunders; and to Bill Lycan and Robert Turnbull, for reminding me that the use of possible-worlds semantics for philosophical purposes is far from universally congenial.

[1] Richard Montague, "Syntactical Treatments of Modality, with Corollaries on Reflexion Principle and Finite Axiomatizability", in *Acta Philosophica Fennica* 16 (1963) : 153-167.

[2] See, e.g., Jaakko Hintikka, "Modality and Quantification"; Jaakko Hintikka, "Existential Presuppositions and Uniqueness Presuppositions"; both reprinted in *Models for Modalities* (Dordrecht, D. Reidel, 1969); Richmond Thomason, "Modal Logic and Metaphysics", in *The Logical Way of Doing Things,* ed. Karel Lambert (New Haven, Yale University Press, 1969) : 119-146; Saul Kripke, "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica* 16 (1963) : 83-94.

[3] Richmond Thomason, "Modal Logic and Metaphysics", p. 120.

[4] Jaakko Hintikka, "Quantifying In," in *The Intentions of Intentionality and Other New Models for Modalities* (Dordrecht, D. Reidel, 1975) : 135 fn. 41.

[5] See, e.g., Wilfrid Sellars, "Counterfactuals, Dispositions and the Causal Modalities," in Vol. II of *Minnesota Studies in the Philosophy of Science* (ed. H. Feigl, G. Maxwell, and M. Scriven), Minneapolis 1957; Roderick Chisholm, "The Contrary-to-fact Conditional," *Mind* 55 (1946) : 289-307; Nelson Goodman, "The Problem of Counterfactual Conditionals," *Journal of Philosophy* 44 (1947) : 113-128.

[6] See David Lewis, *Counterfactuals* (Harvard University Press, 1973) : Ch. 3.

[7] Jaakko Hintikka, "The Semantics of Model Notions and the Indeterminacy of Ontology", in *Semantics of Natural Language,* ed. D. Davidson and G. Harman (Dordrecht, D. Reidel, 1972) : p. 403.

[8] *Ibid.,* p. 402.

[9] Jaakko Hintikka, "Existential Presuppositions and Uniqueness Presuppositions," p. 136.

[10] Jaakko Hintikka, "The Semantics of Modal Notions and the Indeterminacy of Ontology," p. 409.

[11] *Ibid.*

[12] Jaakko Hintikka, "On the Proper Treatment of Quantifiers in Montague Semantics," in *Logical Theory and Semantic Analysis,* ed. Soren Stenlund (Dordrecht; D. Reidel, 1974) : p. 49.

[13] See, e.g., Wilfrid Sellars, *Science, Perception, and Reality* (London : Routledge and Kegan Paul, 1963) : esp. Ch. 11.

[14] See Wilfrid Sellars, *Science and Metaphysics* (London, Routledge and Kegan Paul, 1968) : Ch. 3.

[15] Pavel Tichy, "An Approach to Intensional Analysis," *Nous* 5 (1971) : 273-297.

[16] see Jaakko Hintikka, "Language Games for Quantifiers"; Jaakko Hintikka, "Behavioral Criteria of Radical Translation : A Comment on W.V. Quine's *Word and Object"*; Jaakko Hintikka, "Quantifiers, Language-Games, and Transcendental Arguments"; all reprinted in *Logic, Language-Games and Information* (Oxford, Clarendon Press, 1973).

[17] Jaakko Hintikka, "Quantifiers, Language-Games, and Trans-cendental Arguments," p. 103.

[18] Hintikka comments upon this; see "Language Games for Quantifiers," fn. 12

[19] See Saul Kripke, "Naming and Necessity," in *Semantics of Natural Language,* ed. D. Davidson and G. Harman (Dordrecht, D. Reidel, 1972) : 253-355.

[20] *Ibid.,* p. 266-268.

[21] Nelson Goodman, *Languages of Art* (N.Y., Bobbs-Merrill, 1968) : p. 202 fn. 15.

[22] Jaakko Hintikka, "The Semantics of Modal Notions and the Indeterminacy of Ontology," p. 408.

[23] Jaakko Hintikka, "Existential Presuppositions and Uniqueness

Presuppositions," p. 145.

[24] Jaakko Hintikka, "The Semantics of Modal Notions and the Indeterminacy of Ontology," p. 408.

[25] Roderick M. Chisholm, "The Loose and Popular and the Strict and Philosophical Senses of Identity," in *Perception and Personal Identity,* ed. Norman Care and Robert Grimm (Cleveland, Case Western Press, 1969) : 82-106.

[26] Jaakko Hintikka, "The Proper Treatment of Quantifiers in Montague Semantics," p. 49.

[27] Dagfinn Follesdal, "Knowledge, Identity and Existence," *Theoria* 33 (1967) : p. 7.

[28] See Hilary Putnam, "Minds and Machines," in *Minds and Machines* ed. Alan Ross Anderson (N.J., Prentice-Hall, 1964) : 72-97; Wilfrid Sellars, "Language as Thought and as Communication," *Philosophy and Phenomenological Research* 29 (1969) : 506-527; David Lewis, "An Argument for the Identity Theory," in *Materialism and the Mind-Body Problem,* ed. David Rosenthal (N.J., Prentice-Hall, 1971) : 162-171.

[29] Putnam addresses himself to this issue in "The Mental Life of Some Machines," in *Intentionality, Minds, and Perception,* ed. Hector-Neri Castaneda (Detroit, Wayne State University Press, 1967) ; see especially pp. 180-181.

[30] Some detailed proposals along these lines are offered in my dissertation *Objects* (University of Pittsburgh, 1976), Chapter V.