4

Transworld Heir Lines

DAVID KAPLAN

University of California, Los Angeles

A certain kind of linguistic context has come in for increasing attention over the past several years. The occurrences of the word "nine" in

> It is provable in arithmetic that nine is the square of three;
> It is possible that the number of planets is nine;
> It is permissible that the number of occupants exceed nine;
> It is probable that the number of enrolled students will be less than nine;

 This paper was presented at a joint symposium of the American Philosophical Association and the Association for Symbolic Logic in Chicago in May 1967. The commentators were Jaakko Hintikka and Terence Parsons. The paper was written to be heard, not read, which accounts for its stylistic deviation from the stately prose of my other publications. This version is essentially unaltered except for the addition of the footnotes, all of which are 1978 postscripts. The appearance of the paper at this time may seem anachronistic. It is anachronistic. Most of its truths are now well known, and most of its errors are now rarely repeated. Furthermore, I no longer champion the antihaecceitist viewpoint (as I called it in "How to Russell a Frege-Church," *Journal of Philosophy*, 72 [1975]; Chapter 11 of this anthology). Why, then, allow it to appear at this time? Primarily because it has been called a "classic" by the leading modal logician of our time, and I would not wish to be accused of suppressing classical sources. (To be more exact, the paper is described as a "*locus classicus*" of a certain philosophical mistake. I suppose one must learn to take the bitter with the better.) A second reason for permitting publication at this time is that in its easy, naive way, the paper does focus on what I still take to be the fundamental metaphysical ideas that underlie various common logical constructions in intensional logic. The research was supported by the National Science Foundation.

> It is desirable that symposia be limited to nine;

and,

> It is believable that nine is prime

illustrate such contexts. These occurrences of the word "nine" are neither so vulgar as that in

> Nine is larger than five

nor so accidental as that in

> Canines are larger than felines.

Presumably there are no logical or semantical problems concerned with vulgar or accidental occurrences. Vulgar occurrences of "nine" denote a certain number, are open to substitution and generalization, and contribute to the meaning of the containing sentence. Accidental occurrences are irrelevant to all such concerns. But analysis of the intermediate contexts produced by "provable," "possible," "permissible," "probable," "believable," and "desirable" is neither trivial nor pointless.

Gottlob Frege, who tried to assimilate such intermediate occurrences to the vulgar ones by means of a doctrine about ambiguity and indirect denotation, called such contexts "oblique" or "indirect." W. V. O. Quine, who often seems to want to assimilate the intermediate occurrences to accidental ones by means of a doctrine of indissolubility, calls such contexts opaque.

Frege's way is the more sanguine, for it suggests the possibility of developing a nontrivial logic for such contexts. And indeed our intuitions suggest that such arguments as

> It is probable that the number of enrolled students will exceed nine.
> Therefore it is probable that the number of enrolled students will exceed six

are valid, and valid in view of logical form.

Frege did not link his doctrine of indirect denotation with any particular kind of entity indirectly denoted. But he emphasized cases in which what was indirectly denoted was the ordinary sense.

A number of different proposals have been advanced for conditions under which two sentences, say, would be said to have the same ordinary sense. The most liberal of these, proposed by both Rudolf Carnap and

Alonzo Church, is that the sentences should be logically equivalent. The most restrictive of these, proposed by Benson Mates, is that distinct sentences never have the same sense. In between lie two alternatives, discussed by Church, and Carnap's Intensional Isomorphism. I believe it best to think of these proposals as suggesting different senses of the word "sense". In terms of these proposals, we can form a useful classification of oblique contexts by seeking the most liberal sense of "sense" according to which we can interchange the component sentences in a given oblique context without affecting the truth value of the whole compound.

Following Carnap, let us call the sense of "sense" according to which logically equivalent sentences have the same sense *intension*. And let us call the oblique contexts within which sentences with the same intension can be interchanged *intensional contexts*. These contexts form a large and important class of oblique contexts. Of our original examples, those associated with "provable", "possible", "permissible", "probable", and "desirable" are all rather clearly intensional, and "believable" (or better just "believe") is thought by some to be intensional in at least one of its senses. But in their primary senses the psychological oblique contexts, what Bertrand Russell called *propositional attitudes,* form a second important group whose specific logic has been little investigated.

For the remainder of the paper I will discuss only intensional contexts.

So far I have spoken of intensions only by way of the phrase "have the same intension". But, as Frege noted in the case of "have the same number", to determine conditions under which two classes have the same number is not yet to say what *the number of a class* is, let alone what numbers in general are. Frege, of course, thought of senses as definite entities of a certain kind which could be combined and decomposed, and if we are to develop an intensional logic it would certainly be helpful to have some notion of at least the structure of these entities.

A very natural and simple proposal about the nature of intensions has been advanced by Carnap. He proposes to understand the category of intensions appropriate to sentences (which intensions he calls *propositions*) as sets of possible worlds. The intension of a particular sentence, i.e., the proposition expressed by the sentence, is then taken as the set of all possible worlds in which the sentence is true. This immediately yields the desired consequence that two sentences will express the same proposition if and only if they are logically equivalent.[1]

1. At least if all logically possible states are represented by possible worlds.

I prefer to think of propositions as what might be called characteristic functions of sets of possible worlds, that is, as functions which assign to each possible world one of the two truth values. I prefer this way of thinking about the intensions of sentences because one sees quickly and easily how to generalize the idea. We have rather general agreement now as to what kind of entity the *extension* of an expression of a given grammatical category should be. Thus, the extension of a term is the individual named or described by that term (if there is one; otherwise it has no extension), the extension of a one-place predicate is the class of individuals to which the predicate applies, the extension of a sentence is its truth value, the extension of a truth-functional sentential connective is a certain truth function, etc. We can even provide an analogous extension for variable binding operators. Without going into the matter at this time, let me just mention that it is possible to give arguments showing that in each of these cases the expression bears a similar relation to its extension.[2] So the general notion of the extension of an expression is not just an arbitrary union of some semantical property of terms, some semantical property of predicates, some semantical property of sentences, etc., but really has a kind of validity (as a notion) of its own.

Carnap's simple idea for constructing intensions of arbitrary expressions is just this. Let the intension of an expression be that function which assigns to each possible world the extension of the expression in that world. Intensions, so understood, are independent of expressions, in that we can identify (i.e., define) the class of intensions independently of their being expressed by any particular expression. In fact, if there are at least ω possible worlds, there are at least 2^{ω} propositions, and for most languages this would exceed the number of sentences available to express them.

We will now have a one-minute quiz to make sure that you have been paying attention. Let us call the intension of a name an *individual concept*. What kind of an entity is an individual concept? All those who mumbled something like "a function which assigns to each possible world an individual in the universe of that world", pass. But there is a little problem here. Suppose we have a name which has an extension in some worlds but not in all. "Hamlet" has no extension in the actual world, and I like to think that there are other worlds in which "Reagan" has no extension. You

2. I had in mind here arguments like those given in my dissertation to show that it is natural, if extending the notion of *denotation* from names to sentences, to take the denotation of a sentence to be its truth value.

can probably think of a number of ways of handling this problem. One very simple way is to treat such names as if their extension in such a world were the whole world itself or some other entity so chosen that could not be in the universe of that world, which is the prime desideratum. A slightly more flexible method is to imagine the universe of a world divided into two parts: the individuals which *exist* in the world and those which do not. Then we can stick to the notion of an individual concept as a function which assigns to each possible world an element of its universe, without the consequence that the function always assign something which exists in the possible world.

So to determine the intension of an expression it suffices to determine its extension in each possible world. And if we are given for each possible world the extension of each *atomic* expression in that world and the range of each style of variable in that world, we should be able to determine for each world the extension of an arbitrary compound expression by using the familiar method developed by Alfred Tarski. But there is a hitch.

A special problem arises when the expression in question contains free variables. Suppose we ask for the intension of "x is bald". One way of treating this is to assume implicit universal closure and take "x is bald" as synonymous with "For all x, x is bald." But this avoids the real question. The notion we need is not just the *intension of* Φ but *the intension of* Φ *with respect to a given assignment of values to the free variables of* Φ. Another way of putting it is to say that what we need is not just the proposition expressed by a closed sentence, but the function expressed by an open sentence. Where such a function would yield a proposition for every set of values of the free variables and for different values of the variables, we might get different propositions. Russell called such functions, from individuals to propositions, *propositional functions*. He drove poor Frege to despair by also calling *expressions* like "x is bald" propositional functions. But let's forget that and adopt the terminology "propositional function" to describe the intension of "x is bald". For us, the importance of propositional functions appears in connection with expressions in intensional contexts containing free variables bound to quantifiers outside the intensional contexts. For in such situations the truth conditions of the whole depend in part upon how the proposition expressed by the formula changes as the variable takes different values.

The more subtle among you will have noted that this coining of the phrase "propositional function" does nothing to clarify the notion. Let me review: we are clear on *the proposition expressed by the closed sentence* Γ—it is that function which assigns to each world W, the truth value of Γ in W, and we are also clear on the simpler notion of a *proposition*—any function from worlds to truth values. We are clear on the simple notion of a *propositional function*—a function from individuals to propositions (and we know what propositions are). But we are *unclear* on the notion *the propositional function expressed by the open formula* Φ.

When we hit this problem about propositional functions, I was saying that to determine the intension of an expression it suffices to determine the extension in all possible worlds. So to determine the propositional function expressed by "x is bald" it suffices to determine for each value of the variable "x" the extension of "x is bald" in each possible world. That of course means each actual value, each individual of the actual world.

Let us try in a simple case. Let the value of "x" be Bobby Dylan. Now we must determine the truth value of "x is bald" in each possible world. That is, for each world W does "x is bald" hold of Bobby Dylan in W? First, consider the actual world at the present moment[3] (I'll come to change over time shortly). Clearly "x is bald" fails of Bobby Dylan in this world. Now consider another possible world, the one which would have currently obtained if I had taken my black pen to Chicago and left my red pen in Los Angeles rather than taking my red pen and leaving my black pen as I in fact did. I am fairly confident that "x is bald" fails of Bobby Dylan in that world too. Now how about the world in which the Germans win the Second World War and immediately issue an edict that the only public musical performances allowed will be Wagner and polkas. Call this world "G". I don't want you to focus on the counterfactuals, "would Bobby Dylan have shaved his head if . . . " That is not the problem. I'll even let you peep in at this other world through my Jules Verne-o-scope. Carefully examine each individual, check his fingerprints, etc. The problem is: which one, if any, is Bobby Dylan? That is, which one is *our* Bobby Dylan—of course he may be somewhat changed, just as he will be in our world in a few years. In that possible world which ours will become in, say, thirty years, someone may ask "What ever happened to Bobby Dylan?" and set out to locate him. Our problem is to similarly locate him in G (if he exists there). Although I will continue to speak of *identification,* there are reasons, to which I will return, for claiming that the Bobby Dylan in G is not strictly identical with our Bobby Dylan but related to him in a way something like descendant to ancestor, what Kurt Lewin called *gen-identity*. So I call the

3. That is, May 1967.

task of locating individuals in other worlds the problem of determining transworld heir lines.

I will flatly assert that this problem is the central problem of philosophical interest in the development of intensional logic. The other problems are all technical. (For the general treatment of oblique contexts, there remains the pressing problem of a development, corresponding to that given for the notion of intension, of a sense of "sense" appropriate to the propositional attitudes.)

I know that I have not yet sorted out in detail all our different intuitions related to the transworld identification problem. But I would like to outline three kinds of response. I call them (1) the skeptical, (2) the metaphysical, and (3) the relativistic.

The skeptical response is that it just can't be done. Everyone to his own world. This position may be elaborated by an attempt to show how the idea that we can locate Bobby Dylan in another world arises from confusion of mention and use. Thus the skeptic may claim that it is perfectly reasonable to attempt to locate an individual-under-a-description in another world. So we may try to find Bobby Dylan-under-the-description-"Bobby Dylan" in some other world by, say, looking him up in the telephone book. Or we may seek Bobby Dylan-under-the-description-"the composer of 'Blowin' in the Wind'" in another world by looking him up in the ASCAP registry. But these endeavors may well lead to different results. (We are all aware of the consequence of looking up nine-under-the-description-"the number of planets.") And why take one description rather than another? It would be more enlightening to break the "x-under-the-description-α" nomenclature down into its two components: (a) the intension of α (this is some individual concept, and it does the transworld identification), and (b) the fact that α actually describes x (in our world). The skeptic feels that such talk as

> Bobby Dylan-under-the-description-"the composer of 'Blowin' in the Wind'" is necessarily a musician

is like saying that

> Professor Marcus-under-the-description-"Ruth" is monosyllabic.

Being monosyllabic attaches directly to the name "Ruth" and only in a most remote and indirect way to Professor Marcus. And, similarly, being *necessarily* a musician attaches directly to the individual concept expressed by "the composer of 'Blowin' in the Wind'", and only indirectly, by way of that particular concept, to Bobby Dylan. Another way of putting the skeptic's point about "why select one description rather than another" is to say that he sees no favored way of making the identification. Now you may feel that at least we can eliminate some possibilities, namely, any individual of another world who shares *no* description with our Bobby Dylan. But if there is any sentence true in the other world and false in our own and any description whatsoever which applies to the other individual in his world, then by a logician's trick we can construct a description which describes Bobby Dylan in our world and the given individual in his.[4]

To retrace briefly: the skeptic says there are no favored transworld heir lines; we wanted the transworld heir lines in order to make sense of quantification into intensional contexts, that is, quantification over *individuals* into intensional contexts. Back at the problem of quantifying-in, the skeptic might offer a kind of antiseptic (perhaps "sterile" is a better word) version as follows: Replace such formulations as

> Necessarily x is bald

with free "x", by either

> For *all* descriptions α which in fact describe x the proposition expressed by $\ulcorner\alpha$ is bald$\urcorner$ is necessary.

or

> For *some* description α which in fact describes x the proposition expressed by $\ulcorner\alpha$ is bald$\urcorner$ is necessary.

These two make sense, according to the skeptic, but nothing in between. In view of the logician's trick, however, the first virtually reduces to

> Necessarily everything is bald,

and the second reduces to

> Necessarily something is bald

Among proponents of the skeptical view, I would count William Kneale, Church, half of Quine, and myself at times. I count only half of Quine because, although he is skeptical about quantifying into modal contexts, and I believe for pretty much the reasons I have indicated, he appears

4. Let α describe the other individual in the world in which Φ is true. Then, $\ulcorner$The $x((x =$ Bobby Dylan $\wedge\sim\phi)$ v $(x = \alpha\wedge\phi))\urcorner$ is the required description.

to have no qualms about quantifying into epistemological contexts such as "Wyman believes that x is bald". Of course, his discussions don't reach the point of talking about possible worlds and transworld identifications so he might just reject our whole way of looking at these problems.[5]

I turn now to the second response, the metaphysical position. This is the typical view of logicians. Let me tell the philosophers how logicians think about possible worlds. They identify a possible world with a certain construct which they call a *model*. This construct will depend in an essential way on some language L which might be used to describe the possible world. The model is usually a function or an ordered n-tuple or something of that sort. It has two main features. First, for each style of variable in L, the model provides a set (usually nonempty) as universe of discourse for that style of variable. Distinct styles of variables represent distinct grammatical categories, so we might say that the model provides a set of entities which forms the ontology for each grammatical category of L. Second, the model provides an extension for each nonlogical constant of L, an extension chosen from the ontology corresponding to the grammatical category of the constant. The model, so to speak, gives us the extension of all the atomic expressions of L, and from this we can obtain the extension of any compound expression of L, in particular, the truth value of any sentence of L.

There are two interesting reasons for calling these things "models." First, the primary use of logicians, is in the sense of "exemplar" as in "he is the very model of a modern major general". In this sense we speak of a model as being a model *of* any set of sentences which correctly describe the possible world represented by the model. The second reason for calling these things models is that they *represent* possible worlds. It is reasonable to say that every possible world is represented by some model and that, assuming that there are no unexpressed logical dependencies among the nonlogical constants of L, every model represents some (logically) possible world.

As long as we relied only on a syntactical criterion—being able to derive an explicit contradiction—to tell us when a sentence was logically consistent, we ran the danger that a sentence which under the intended interpretation could not be true in any possible circumstances still would not yield an

5. In 1968, I undertook a more detailed examination of Quine's views on quantifying into epistemological contexts ("Quantifying In," *Synthese* [December 1968]). It, too, I fear, is a *locus classicus* of a philosophical mistake. At least it is a different mistake.

explicit contradiction. This could come about if axioms or rules were not formulated in a sufficiently strong way. And without some independent notion of truth in a possible world there was little to say about the adequacy of such formulations. Insofar as we are satisfied that models represent all possible worlds, and no impossible worlds, they provide the independent criterion against which to check our formulations of axioms and rules. This idea of using models to represent possible worlds was really quite clever. I think it is fair to say that it is due mainly to Tarski and that about 37 percent of the work in logic over the past thirty years is based on it.

In fact, the use of models as representatives of possible worlds has become so natural to logicians that they sometimes take seriously what are really only artifacts of the model. In particular, they are led almost unconsciously to adopt a *bare particular metaphysics*. Why? Because the model so nicely separates the bare particular from its clothing. The elements of the universe of discourse of a model have an existence which is quite independent of whatever properties the model happens to tack onto them. Suppose we want a model for the sentence of L which asserts that there is exactly one thing and that it is a unicorn. A model for such a sentence must have a universe with only one element, and the extension assigned to the predicate "is a unicorn" must be the set consisting of that single element. And that is *all* that is required of the model. It is certainly *not* required that the single element of the universe of the model really be a unicorn. That would make the whole idea of the models unworkable (since there are no unicorns). The single element of the universe of the model may be Jaakko Hintikka, or more likely, because logicians like their entities to exhibit a maximum degree of purity, it may be the null set, or singleton null. But, at any rate, it will be some definite entity which, in this model, is dressed as a unicorn.

Let me refer in a loose way to this kind of situation by saying that the entity in question is *intrinsically* Professor Hintikka or the null set or whatever, and *extrinsically* a unicorn. Most of the time logicians recognize a certain lack of significance in the intrinsic nature of the elements of the universe of a model (except, of course, with respect to identity and difference, i.e., how *many* of them there are) and focus their attention on isomorphism classes of models (two models being isomorphic if their universes can be put into a one-to-one correspondence in such a way that corresponding elements differ only intrinsically). But there is a kind of confusion as to whether we should think about isomorphic models as distinctive representatives of the same possible world, or as representatives

of distinct possible worlds which differ only as to individuals (i.e., bare particulars). The last is in violation of the law of the identity of indiscernibles (that law becomes interesting primarily in these transworld situations).[6]

I hope you see how taking the intrinsic nature of the elements of the universe of a model seriously is connected with the bare particular metaphysics. If we adopt this metaphysical view, we have a simple solution to the transworld identification problem: we identify by bare particulars. If our worlds are represented by models and we take the elements of the universe of the models to be (or represent) bare particulars, individuated by their intrinsic characteristics, then there is no difficulty from the point of view of the metalanguage in making such identifications. A metalanguage in which we talk explicitly about models provides us with a kind of meta-x-ray of the bare particular lurking beneath each individual. For example, let us take our earlier model, call it M_1, of "There is exactly one thing and it's a unicorn". To identify that unicorn in some other model M_2, say, of "Everything wears sandals", we ignore the extrinsic characteristics of the unicorn and the hippies and check instead to see if Professor Hintikka who appeared in M_1 as a unicorn reappears in M_2 as, say, Timothy Leary.[7]

This metaphysical conception is at the heart of much of the extremely interesting work that has been done on the semantics of modal logic in the last twenty years. It appears most explicitly in Saul Kripke's work, but occurs also in Carnap's pioneering article of 1946, "Modalities and Quantification."

The distinction between intrinsic and extrinsic characteristics is somewhat blurred in Carnap's article because his systems were intended to have *abstract* entities such as space-time positions as the values of the variables. Thus he was not faced with the problem of finding entities to appear as unicorns. He did have to find space-time positions to appear as occupied-by-a-unicorn, but the conception of a space-time position separable from its occupant still seems to me some distance from that of a bare particular separable from all its properties. (Maybe not.) I think this is an interesting and even relevant problem, but I won't pursue it here.

6. I wish I hadn't said this.

7. That is, reappears in M_2 "clothed" in the predicates known to apply to Timothy Leary, a notorious sandal wearer of the sixties.

Now I want to move on to the third position on transworld identifications: the relativistic. This position might be associated with the bundle-of-qualities metaphysics insofar as it is associated with any particular metaphysical view. We so-to-speak look only at the clothes, and we identify individuals in terms of their strikingly similar manner of dress, i.e., their sharing of a large number of prominent qualities. Our view of individuals in different worlds is through the Jules Verne-o-scope, which, you recall, enables us to compare fingerprints, ASCAP registries, and even CIA files, but does not allow us to see into any underlying bare particular as does the meta-x-ray machine of the previous position. Now let us go back to some of the cases considered earlier; first, the world in which I brought my other pen to Chicago, call that world "p". [Note that I have actually introduced this world in terms of a transworld identification of myself, but let's neglect that. Given sufficient time or technical skill I might have introduced it by reading its complete book of history or showing you transcriptions from the Verne-o-scope.] I think we would have no difficulty in deciding with what individual of that world to connect Terence Parsons. Let's call that individual "Parsons-in-p". In the terminology of my colleague David K. Lewis, Parsons-in-p is a *counterpart* of our Parsons.[8] He is not *identical* with our Parsons, because *he* is sitting next to a man carrying a black pen, whereas our Parsons is sitting next to a man without one.[6] And whatever you may think about the identity of indiscernibles, *no* sensible person would deny the indiscernibility of identicals. We call Parsons-in-p and our Parsons counterparts, in spite of this difference, because we regard the difference as inessential, especially in view of the overwhelming similarities. Our world and p are very much alike, and that makes the transworld identifications quite easy. But as the disparity grows, these identifications become more difficult. And we find ourselves forced to make finer and finer discriminations between what is essential to Professor Parsons (being named "Terence Parsons"? being a philosopher? being rational?) and what is only accidentally true of him (sitting next to a man carrying a red pen? having lived in California? being bearded?). What we are searching for is his *essence*, that which identifies an individual of any possible world W as Parsons-in-W, or more exactly: as being the counterpart in w of our Professor Parsons.

8. David K. Lewis, "Counterpart Theory and Quantified Modal Logic," *Journal of Philosophy*, 65 (1968); Chapter 5 of this anthology.

Don't get confused here. I'm not talking about anything so prosaic as the intension of some particular name, say, the name "Professor Terence Parsons"; I'm talking about a counterpart of *him* (pointing).

I prefer to think of an essence in this way (as a transworld heir line) rather than in the more familiar way (as a collection of properties) because the more familiar way too much suggests the idea of a fixed and final essential *description,* and that the essence should somehow be expressible, whereas my way of thinking of essences seems to me to accord better with our intuitions and the ordinary practices of scientists. When geographers decide whether the Missouri-Mississippi is one river or two, demographers whether Los Angeles and Ventura form one metropolitan area or two, and jurists whether bludgeoning the victim with the spent rifle constitutes a second attempted murder or a continuation of the first, they are neither searching for a metaphysical oneness (a common bare particular) nor are they applying a previously fixed formula for, say, the individuation of rivers and the Mississippi in particular. Instead they make a determination based on a careful examination of all the facts of the case and aimed at the discovery of especially prominent characteristics relevant to the particular science. Typically, each case is judged on its merits (within certain broad guidelines) and at no point is a fixed and final principle of individuation, or essence, offered.

From here on out, when I speak of "transworld identifications" you should understand that I am not speaking of identities in the strict sense, but of counterparts. Since each transworld heir line corresponds to an individual concept, the essences are some subset of the individual concepts, namely, those which link counterparts. Remember that we admit individual concepts that are not the intension of any singular term, and some essences may be of this form.

I believe we all have at least partial intuitions about many essences. If you will allow a momentary indulgence in the subjective, I might remark that I have the feeling that some things are so unremarkable that they have no essence, which is to say more than that they have no counterparts. (The latter might hold of an extremely vivid person whose remarkable qualities were somehow specific to this world.) And that other persons, for example Da Vinci, seem to me to have more than one essence.

It may help to clarify this idea of an essence if I make explicit what is implicit in the analogy to the decisions of geographers, demographers, and jurists, namely, that any choice of essences (from among *all* the individual concepts) is *relative* to certain interests. Thus the constant appearance in

my earlier formulations of such adjectives as "relevant", "prominent", etc. We can acknowledge the skeptic's point regarding the plentitude of definite descriptions of a single object, where each description expresses a different individual concept (i.e., a different candidate for the object's essence), and still maintain that relative to certain interests, one description may be more relevant or pertinent than another. If we inquire after the unique essence of Sir Lancelot, Lady Guinevere, King Arthur, and Lancelot's mother would no doubt each answer differently to the question: "What makes this knight different from all other knights?"

Suppose I spend twelve hours in the hospital, go home, spend a few days feeling lousy, and return to the hospital for another twelve hours. How many periods of confinement and of what length have I spent? My insurance company, which only wants to pay for certain medical charges when connected with a *serious* illness, restricts such payments to cases involving a single period of hospital confinement lasting at least eighteen hours. But they count interrupted confinements as single periods provided that I do not return to work during the interruption. So from their point of view I was confined to the hospital for a single period of twenty-four hours. But the hospital accountant, who is interested in determining appropriate hospital rates, must distinguish the variable costs of maintaining me in the hospital from the fixed costs of admitting, discharging, and billing me, because these latter costs are independent of the *length* of the period of confinement. So from his point of view I was confined to the hospital for two separate periods of twelve hours each.

If my imagination were not exhausted at this point, I could probably think of someone whose interests are such that the color of the pen in my pocket would be of such importance that he would not make what we earlier thought were the natural connections between individuals in our world and those in p.

Let me point out here that the results and the technical constructions of logicians who might be described as using the metaphysical method could also be described in terms of the present method. As conceived in accordance with the metaphysical position, the transworld identification of Hintikka appearing as a unicorn and Hintikka appearing as Timothy Leary was based on the previous isolation of Hintikka as common bare particular of the unicorn and Timothy Leary. This is the substance-before-accident point of view. But suppose instead that for some reason we wish to identify Timothy Leary and the unicorn solely in view of their extrinsic characteristics; possibly both exhibit a certain unique dreamy expression in which we

are very interested. We might now conceive of the logician's representation of the two possible worlds by models in which a single entity, Professor Hintikka, appears as the unicorn in one model and as Timothy Leary in the other, as being simply a technical device to carry out the preconceived identification. You see we have reversed the whole procedure. This is the accident-before-substance point of view. First we decide how we want to connect individuals and then we treat those individuals that we want to connect as identical.

I rather like this last way of conceiving of models in connection with the transworld identification problem. The only remaining difficulty in representing our choice of essences in terms of the intrinsic characteristics of the elements of the universe of the model is that it lacks flexibility. It lacks flexibility in that it reduces the problem of transworld identification to a form of ordinary identity within the metalanguage, and so does not permit representation of essences which may merge in one world and divide in another. But once we adopt the relativistic method, such a restriction is plainly undesirable. In many cases, our interests in a person are solely in terms of his office. He is the butcher, or the baker, or the candlestick-maker. It may easily happen that the butcher and the baker are in fact one and the same person (possibly unbeknown to us). Or, though distinct in our world, they may be one in another.[9] In such cases I would say that two essences characterize a single individual in one world, but different individuals in another.

To move to cases where delineation of the essences is less clear-cut, imagine a world like ours through 1842 but in which the counterparts of the parents of William and Henry James have but a single son who is educated at (the counterpart of) Harvard and later accepts a chair in philosophy there, but takes frequent leaves, spent in England writing such novels as *The American, The Bostonians,* etc. His academic work consists of *The Will To Believe, Principles of Psychology* (2 vols.) etc. You fill in the details. If you doubt the possibility of such a combination of essences realized in a single person, just think of our own Bertrand Russell, who is clearly the counterpart of at least three distinct persons in some more plausible world.

There is a wider use for the intensional notions we have been considering than in application to what we would think of, properly speaking, as possible worlds. In considering the foundations of a logic of tenses, or a

9. This "they" is short for the phrase "the butcher and the baker"; it does not refer to the two gentlemen.

logic of token-reflexive words such as "I", "he", "now", and "here" we often come upon a situation in which we have a number of distinct *frames of reference,* which from an abstract logical standpoint function very much like possible worlds. Richard Montague has especially emphasized this idea by attempting to show how the logic of a wide variety of seemingly special theories can be unified within a single logical system which he calls *Pragmatics.*[10] A typical example is the case of temporal logic—a logic of change. Here the frames of reference, or possible worlds, are temporal slices of a world. Now what are the *things* of these possible worlds? Well, just as in the case of the elements of the universe of a model, there are two ways of looking at these things. From what we might call the *meta* point of view (metalanguage or metaphysical), we have some single individual appearing at different times in different roles: first as an infant, next as a child, etc. From this point of view, the infant of 1933 and the adolescent of 1948 are artificially constructed slices of a single person, slices which occur wholly within a single temporal stage. Looked at from the point of view of the temporal slices, i.e., the point of view of the possible world itself, we have as basic entities an infant in one time slice and an adolescent in another. From this point of view, we may undertake to construct the whole person in terms of some means for connecting an infant in one world with a child in another, with an adolescent in another, and so on. But whichever way we look at it, whether we start from the individual slices and construct persons like sandwiches, or start with persons and construct the person stages like slices of baloney, there are two kinds of entities involved: the entity specific to a frame of reference and the superentities which run across frames. If our logical system is viewed in this general way, it again seems undesirable to disallow the possibility of distinct superentities fusing in a single local entity at some particular frame of reference and dividing into two at another frame. Maybe I can express the kind of generality that I want to allow by saying that we should permit what appears as a single thing from one frame of reference to appear as distinct things from another frame. Such fusion and division through time is a feature of the ordinary behavior of corporations, iris, and the amoeba, not to mention pathological behavior in personalities and maple trees. Fusion and subsequent division through space is an ordinary feature of highways. Of course, we could insist that upon division the career of an amoeba ends and two new ones are born (or at least that the original

10. "Pragmatics" in Richard Montague, *Formal Philosophy,* ed. by Richmond Thomason (New Haven: Yale University Press, 1974).

amoeba is identified with one of the pair and one new one is born) and similarly for schizophrenia, corporate mergers, and the state and federal highway systems. But why not let the amoebaeologists, psychologists, economists, and traffic planners settle the question in their own way—in the way that best suits *their* interests.

Let me add here a note of caution to the overly enthusiastic transfer of these general notions back and forth between intensional logic proper (a theory of possible worlds) and, say, time logic (the theory of stages of a single world). I find on introspection that in most cases (though not all) it is the superentity that I think of as basic when considering temporal stages as the possible worlds (and similarly for spatial stages) and the slices, the entities specific to a stage, that I think of as somehow artificial. But when I think about different possible worlds proper (i.e., what might be but is not), the entities specific to a stage seem to be basic and naturally determined and the superentities (the transworld heir lines as I earlier called them) seem to me somehow artificial and determined only relative to certain interests. And I am not sure that further consideration of these intuitions would not lead to the discovery of a logical difference between the two kinds of frame of reference.

I would like to conclude by briefly outlining a formal system which embodies some of the ideas I've been talking about. I call the theory *Essentialism*[6] since it is a theory of the kind of transworld heir lines that I have identified with essences.

ESSENTIALISM

The Language

1. Logical Signs
   - (i) Two styles of variables, V^e (the set of variables ranging over essences) and V^i (the set of variables ranging over individuals).
   - (ii) $=$, $\forall$, $\exists$, $\rightarrow$, —, v, $\wedge$, $\longleftrightarrow$, (,).
   - (iii) δ (for $\alpha \in V^e$, $\delta(\alpha)$ denotes the individual determined in the given world by the essence denoted by α).
   - (iv) $\Box$ (a kind of logical necessity).
2. Nonlogical Signs
   - (i) n-place predicates of individuals and n-place operation symbols applying to individuals.

   - (ii) m-n-place propositional operators (these relate individuals and propositions).
   - (iii) necessity operators.
3. Terms
   - (i) The only terms whose values are essences are the variables in V^e.
   - (ii) The *individual* terms (T^i) are given as follows:
     - (a) if $x \in V^i$, then $x \in T^i$,
     - (b) if $\alpha \in V^e$, then $\delta(\alpha) \in T^i$,
     - (c) if $t_1, \ldots, t_n$ are in T^i and η is an n-place operator symbols, then $\eta(t_1, \ldots, t_n) \in T^i$.
4. Formulas
   - (i) If $\alpha, \beta \in V^e$, $\alpha = \beta$ is a formula,
   - (ii) If $t_1, t_2 \in T^i$, $t_1 = t_2$ is a formula.
   - (iii) If $t_1, \ldots, t_n \in T^i$ and π is an n-place predicate, then $\pi(t_1, \ldots, t_n)$ is a formula.
   - (iv) If ϕ, ψ are formulas, $-\phi$, $(\phi \wedge \psi)$, $(\phi \text{v} \psi)$, $(\phi \rightarrow \psi)$, and $(\phi \longleftrightarrow \psi)$ are formulas,
   - (v) If ϕ is a formula and $v \in (V^i \cup V^e)$, then $\forall v\phi$, $\forall v\phi$ are formulas.
   - (vi) If $t_1, \ldots, t_m \in T^i$, $\phi_1, \ldots, \phi_n$ are formulas, and ¶ is an m-n-place propositional operator, then $(t_1, \ldots, t_m \text{ ¶ } \phi_1, \ldots, \phi_n)$ is a formula.

For grammatical purposes, $\Box$ is a necessity operator, and all necessity operators are 0-1-place propositional operators.

Model Structures, Models, and Assignments

$\mathcal{A}$ is a model structure iff there are W, U, X, I, P, R, E such that $\mathcal{A} = \langle W, U, X, I, P, R, E \rangle$ and

1. W is a nonempty set (the set of possible worlds w)
2. for $w \in W$, U_w is a nonempty set (the set of individuals existing and otherwise, of w)
3. for $w \in W$, $X_w \in U_w$ (the set of existing individuals of w, X_w may be empty)
4. I is a function which assigns an appropriate intension to each predicate and operation symbol (e.g., if π is a n-place predicate, $n \rangle o$, then $I(\pi)$ is a function whose domain is W, and for $w \in W$, $I(\pi)(w)$ is

a set of n-tuples of individuals drawn from U_w. Note the treatment of O-place predicates in 5 below.

5. P is a function which assigns an appropriate intension to each m-n-place propositional operator (other than the necessity operators). Thus, if ¶ is a m-n-place propositional operator $P(¶)$ is a function whose domain is W, and for $w \epsilon W$, $P(¶)\ (w)$ is a set of m-n-tuples $\langle i_1, \ldots, i_m, P_1, \ldots, P_n \rangle$ where $i_1, \ldots, i_m \epsilon U_w$ and $P_1, \ldots, P_n$ are propositions (a proposition, the sense of an O-place predicate, is here thought of as a subset of W, rather than a characteristic function).
6. R is a function which assigns an appropriate intension to each necessity operator (other than $\Box$). Here the extension of a necessity operator is thought of in the way popularized by Kripke, as the set of possible worlds accessible from the given worlds. Thus if $\mathbb{N}$ is a necessity operator and $w \epsilon W$, the extension of $\mathbb{N}$ in w is a subset of W. Hence $R(\mathbb{N})$ is a function whose domain is W and for $w \epsilon W$, $R(\mathbb{N})\ (w) \subseteq W$. (An equivalent formulation is to let $R(\mathbb{N}) \subseteq (W \times W)$).
7. E is included in the set of all individual concepts of $\mathscr{A}$. An individual concept of $\mathscr{A}$ is a function which assigns to each $w \epsilon W$, an element of U_w. (E is, intuitively, the set of essences of W).

If $\mathscr{A}$ is as above and $w \epsilon W$, then $\langle w\mathscr{A} \rangle$ is a model.

If $f \epsilon E^{Ve}$, then f is an $\mathscr{A}$-assignment to essence variables.

If $g \epsilon U_w^{vi}$, then g is a w-assignment to individual variables.

If f and g are as above, $(f \cup g)$ is a $\langle w\mathscr{A} \rangle$-assignment. (Every $\langle w\mathscr{A} \rangle$ assignment can be uniquely decomposed into the components f and g).

If h is an assignment and v is a variable in its domain $h_b^v = (h \sim \{\langle v, h(v) \rangle\}) \cup \{\langle v, b \rangle\}$.

Value and Satisfaction

Let $\mathscr{A}$ be as above, $w \epsilon W$, and h $(=(f \cup g))$ be a $\langle w\mathscr{A} \rangle$ assignment.

The value (extension, denotation) of a term in a model is given as follows: (We treat $\mathscr{A}$ as fixed and write "h-val$_w$" for "h-val$\langle_w \mathscr{A}\rangle$")

1. if $v \epsilon (V^i \cup V^e)$, $h\text{-val}_w(v) = h(v)$
2. if $\alpha \epsilon v^e$, $h\text{-val}_w(\delta(\alpha)) = h(\alpha)\ (w)$
3. if $t_1, \ldots, t_n \epsilon T^i$, and η is a n-place operation symbol, $h\text{-val}_w(\eta(t_1, \ldots, t_m)) = I(\eta)\ (w)\ (h\text{-val}_w(t_1), \ldots, h\text{-val}_w(t_n))$.

Satisfaction is given as follows: (Again "$\langle w\mathscr{A} \rangle$" is written as "$w$")

1. if $\alpha, \beta \epsilon V^e$, h sat$_w$ $(\alpha = \beta)$ iff $h(\alpha) = h(\beta)$
2. if $t_1, t_2 \epsilon T^i$, h sat$_w(t_1 = t_2)$ iff $h\text{-val}_w(t_1) = h\text{-val}_w(t_2)$
3. if $t_1, \ldots, t_n \epsilon T^i$ and π is an n-place predicate h sat$_w$ $(\pi(t_1, \ldots, t_n))$ iff $\langle h\text{-val}_w(t_1), \ldots, h\text{-val}_w(t_n) \rangle \epsilon I(\pi)\ (w)$
4. if ϕ, ψ are formulas, h sat$_w(\phi \wedge \psi)$ iff h sat$_w \phi$ and h sat$_w \phi$.
5. Similarly for other sentential connectives
6. if ϕ is a formula, and $v \epsilon V^i$ $[v \epsilon V^e]$ h sat$_w$ $(\forall v \phi)$ iff for all $b \epsilon X_w$ [for all $b \epsilon \exists$] h_b^v sat $_w \phi$.
7. Similarly for existential quantifier.

Auxiliary notion: Given an assignment f of values to V^e, the proposition expressed by a formula Φ, is the one which consists of just those worlds w' in which Φ is true. Since we intend to connect individuals across worlds *only* by way of essences, *individual* variables free in Φ are treated as if bound by a universal quantifier ranging over all of $U_{w'}$ (note that the quantifier in our language allows the individual variables to range only over $X_{w'}$).

Let $f \epsilon \exists^{Ve}$, Φ a formula, $w' \epsilon W$, then $w' \epsilon f\text{-prop}(\phi)$ iff for all $g \epsilon U_{w'}{}^{vi}$ $(f \cup g)$ sat$_{w'}$.

8. If $t_1, \ldots, t_m \epsilon T^i$, $\phi_1, \ldots, \phi_n$ are formulas, and ¶ is an m-n-place propositional operator, h sat$_w(t_1 \ldots t_m$ ¶ $\phi_1 \ldots \phi_n)$ iff $h\text{-val}^w(t_1), \ldots, h\text{-val}_w(t_m)$, $f\text{-prop}\ \phi_1, \ldots f\text{-prop}\ (\phi_m)) \epsilon P(¶)\ (w)$ [recall $h = (f \cup g)$].
9. if $\mathbb{N}$ is a necessity operator and Φ is a formula h sat$_w$ $(\mathbb{N}\ \phi)$ iff $R(\mathbb{N})\ (w)\ c\ f\text{-propr}(\phi)$
10. if ϕ is a formula, h sat$_w(\Box \phi)$ iff $f\text{-prop}(\phi) = W$

If ϕ is a formula, and $\langle w\mathscr{A} \rangle$ is a model, ϕ is *true* in $\langle w\mathscr{A} \rangle$ iff for all $\langle w\mathscr{A} \rangle$-assignments h, h sat$_w \phi$

If ϕ is a formula, ϕ is *valid* iff for every model $\langle w\mathscr{A} \rangle$, ϕ is true in $\langle w\mathscr{A} \rangle$.

Axiomatizability

The valid formulas are axiomatizable by means of a finite set of axiom schemes and rules.[11]

11. I have here made a deletion. In the original version I claimed that if it were required that E consist of *all* individual concepts of $\mathscr{A}$, the valid formulas would not be axiomatizable. It now seems, in view of recent results by Saul Kripke and independently by Hans Kamp, that this was in error.

If quantifiable variables ranging over all propositions were added, the valid formulas would not be axiomatizable.

Representation of Other Theories

Many known intensional logics, in particular modal logics, are interpretable in Essentialism or one of its extensions obtained by adding some special axioms for one of the necessity operators plus some special axioms about essences.

At least four different interesting propositional functions (in Russell's sense) expressed by "x is bald" can be explicitly represented in Essentialism.

The four propositional functions may be loosely expressed as follows: to the individual x, assign the proposition which includes all those possible worlds in which

(1) every counterpart of x which exists is bald,
(2) some counterpart of x exists and is bald,
(3) every counterpart of x exists and is bald,
(4) some counterpart of x is bald if it exists.

Let ¶ stand for the O-1 place propositional operator "It was asserted that." Then corresponding to the four propositional functions, we have four translations of "It was asserted that x is bald" (with free "x").

(5) $\forall\alpha(\delta(\alpha) = x \rightarrow \P[\forall y(\delta(\alpha) = y) \rightarrow y \text{ is bald})])$,
(6) $\exists\alpha(\delta(\alpha) = x \wedge \P[\exists y(\delta(\alpha) = y) \wedge y \text{ is bald})])$,
(7) $\forall\alpha(\delta(\alpha) = x \rightarrow \P[\exists y(\delta(\alpha) = y) \wedge y \text{ is bald})])$,
(8) $\exists\alpha(\delta(\alpha) = x \wedge \P[\forall y(\delta(\alpha) = y) \rightarrow y \text{ is bald})])$.

Among special axioms on essences the following are of particular interest:

(9) $\forall x\exists\alpha(\delta(\alpha) = x)$,
(10) $\forall\alpha\forall\beta(\exists x(\delta(\alpha) = x \wedge \delta(\beta) = x) \rightarrow \alpha{=}\beta)$,
[Note that $[\alpha = \beta \longleftrightarrow \Box(\delta(\alpha) = \delta(\beta))]$ is already an axiom.
(11) $\forall\alpha(\exists x(\delta(\alpha) = x) \rightarrow \Box\exists x(\delta(\alpha) = x))$.

(9) says that everything which exists has at least one essence. (10) says that essences of existing things do not fuse and divide, or equivalently that

everything which exists has at most one essence. (11) says that every counterpart of an existing thing exists. Taken together (9)–(11) say that the "same" things exist in all possible worlds. (9) and (10) taken together say that each existing thing has exactly one counterpart (in each world) which may or may not exist. Note that in the presence of (9) and (10), (5) and (8) become equivalent as do (6) and (7) and that if (11) is added (5) and (6) become equivalent, thus reducing the four propositional functions to one.

If it is desired to treat some or all individual constants t as proper names expressing essences, this can be done by adding axioms of the form:

(12) $\exists\alpha\Box(\delta(\alpha) = t)$.

In addition to the four translations (5)–(8), iterated intensional operators provide another dimension of translation in which alternatives are available. Using the method of (5), we may translate "It was asserted that it was asserted that x is bald" in either of the following ways:

(13) $\forall\alpha(\delta(\alpha) = x \rightarrow \P\,\P[\exists y(\delta(\alpha) = y) \rightarrow \delta(\alpha) \text{ is bald}])$
(14) $\forall\beta(\delta(\beta) = x \rightarrow \P[\exists z(\delta(\beta) = z) \rightarrow \forall\alpha(\delta(\alpha) = \delta(\beta) \rightarrow \P[\exists y(\delta(\alpha) = y) \rightarrow \delta(\alpha) \text{ is bald}])])$.

The first might be thought of as obtaining by applying a compound intensional operation (¶ ¶) to a proposition, whereas the second results from iterative application of a single operator. The distinction is connected with an argument by Church in footnote 22 of "A Formulation of the Logic of Sense and Denotation" in *Structure, Method, and Meaning: Essays in Honor of Henry M. Sheffer,* ed. by Pául Henle, H. M. Kallen, and S. K. Langer (New York: Liberal Arts Press, 1951).

A formal system of slightly greater flexibility can be obtained by allowing the class of essences to vary from one possible world to another.

The Possible and the Actual

READINGS IN THE METAPHYSICS OF MODALITY

EDITED BY

Michael J. Loux

CORNELL UNIVERSITY PRESS
Ithaca and London

First published 1979 by Cornell University Press.
Published in the United Kingdom by Cornell University Press Ltd.,
2-4 Brook Street, London W1Y 1AA.

International Standard Book Number (cloth) 0-8014-1238-2
International Standard Book Number (paper) 0-8014-9178-9
Library of Congress Catalog Card Number 79-7618
Printed in the United States of America
Librarians: Library of Congress cataloging information appears on the last page of the book.

TO NEIL DELANEY
BEST OF ALL POSSIBLE CHAIRMEN