DANA SCOTT

ADVICE ON MODAL LOGIC

Everyone knows how much more pleasant it is to give advice than to take it. Everyone knows how little heed is taken of all the good advice he has to offer. Nevertheless, this knowledge seldom restrains anyone, least of all the present author. He has been noting the confusions, misdirections of emphasis, and duplications of effort current in studies of modal logic and is, by now, anxious to disseminate all kinds of valuable advice on the subject. Thus he is very happy that the Irving meeting has provided such a suitable and timely forum and hopes that all this advice can provoke some useful discussion – at least in self-defense. The time really seems to be ripe for a fruitful development of modal logic, if only we take care to purify and simplify the foundations. A quite flexible framework is indeed possible: the old puzzles can be brushed aside, and one can begin to provide meaningful applications.

Before embarking on details, here is one general piece of advice. One often hears that modal (or some other) logic is pointless because it can be translated into some simpler language in a first-order way. Take no notice of such arguments. There is no weight to the claim that the original system must therefore be replaced by the new one. What is essential is to single out important concepts and to investigate their properties. The fact that the real numbers can be defined in terms of sets is no argument for being interested in *arbitrary* sets. One must look among the sets for the significant ones and cannot be censured if one finds the intrinsic properties of the reals more interesting than any of their formulations in set theory. Of course if we can argue that set theory provides other significant concepts, then we may find some reason for going beyond the real numbers (and it is not hard to find the reasons!). But clearly this discussion cannot proceed on purely formal grounds alone.

This essay is divided into six sections. Section 1 discusses individuals: actual, possible, and virtual, Section 2 introduces possible worlds and their generalizations, Section 3 combines possible worlds and individuals to produce the fundamental notion of *individual concept*. Section 4 dis-

K.Lambert (ed.), Philosophical Problems in Logic.

cusses general intensional operators. Section 5 completes the discussion of the previous section by showing how to handle variable-binding operators. The final Section 6 brings out the distinction between incidence and equality needed for the problems of cross-world identifications. (The final discussion is not as satisfactory as the author hoped, and he looks forward to future developments.) The plan of writing was to set out the semantical framework for modal logic gradually – with (hopefully) adequate explanation – so, that all the good advice would not seem too authoritarian. In any case the various critical remarks were meant in the spirit of rational discussion. We want a synthesis of all good ideas and not just a triumph of one man's approach over another's.

I. INDIVIDUALS

What is an individual? A very good question. So good, in fact, that we should not even try to answer it. We could assume that being an individual is a primitive concept – that is harmless: any sufficiently clear concept can be made primitive. But maybe we do not want the individuals themselves but only some *constructs* or *tokens* representing the individuals. For the moment it does not matter: what is important is to agree that the individuals (or their legal representatives) can be collected together into one domain, a set that we may call D.

I feel it is important to be thinking of D as *fixed in advance*. This does not mean that one knows all the elements of D in any constructive sense, for one may only know some property specifying D. But much confusion results if one considers D vague. Any set can be enlarged by the introduction of new elements; but then a new set $D' \neq D$ obtains, and various results may undergo modification when interpreted relative to the new domain. Sometimes this change may take on the aspect of paradox if one forgets that $D' \neq D$. Maybe in the future we shall understand the logic of *potential totalities* (through intuitionism possibly?) but for the present our simple two-valued logic demands this idealization. Be not discouraged that from the first this idealism hangs in the air: ask yourself if you fully appreciate how powerful and flexible a tool two-valued logic really is. I feel that it is better to work out the idealized situation first – and then retrench later when one can value thoroughly what is gained and lost.

How big is D? Well, it should be rather large (at least *non-empty*)

because we want D to be the domain of *possible individuals*. Here 'possible' means possible with respect to some a priori conception. The term 'conceivable' might be more to the point. In any case let me emphasize again that it is a relative notion. Of course we can conceive of individuals not in D, but that is neither here nor there. My only demand is that we conceive of at least *one* possible individual, because in general logic on the non-empty domain is simpler. Quantifiers over possibly empty domains will come in later in a more reasonable context.

It took me a very long time to concede the point that the notion of possible individual was reasonable. It was only after many discussions with the UCLA group (Montague, Kaplan, Cocchiarella and Kamp) that I finally saw the light. I was glad to note at the present meeting that others are coming to the same conclusions (van Fraassen and Lambert in particular). There are many examples to make the idea plausible: consider the following sentences. 'No two presidents of the United States ever looked alike.' 'All Nobel prize winners were equally deserving.' 'All readers of the novel will be as deeply moved as I.' We are playing here with the passage of time. Individuals (in this case persons) come into and go out of existence. Nevertheless it is meaningful to compare two of them existing at different times: not always two specific persons so that quantification over all the possible individuals is actually required. That is the difference between the first two of these sentences (in past tense) and the third (future). This is only one line of examples; once the problem is appreciated hundreds of examples come to mind.

We shall return to the distinction between *actual* and *possible* individuals below in Section 5. In this section we wish to emphasize another distinction: *possible* vs. *virtual*. It has been Quine who has recently popularized virtual entities. Forgive the word 'entity', since Quine would seem to be happier to treat them merely as *façon de parler* with all references thereto avoidable by contextual definition. That seems to me to be a mistake of emphases or maybe an error of shortsightedness. The possibility of introducing virtual entities is unlimited and for the most part relatively unexplored. They are not to be regarded as ghosts but rather as *ideal objects* introduced to enhance the regularity of our language. By using the names of these entities we often find a simple formulation that avoids a confusing proliferation of cases. In the ordinary theory of real numbers $\pm\infty$ are excellent examples of useful ideal points; in projective geometry, the points at infinity; in set theory, the virtual classes. Of course each of the above mentioned notions can be eliminated (at the cost of much longer statements) by contextual definitions. We must now examine my reasons for advising not making contextual elimination the central feature of these concepts.

My most outstanding reservation centers on the feeling that I do not know what the general theory of contextual definitions is. I even have some doubt that there is such a theory. Maybe I am just being stubborn as I have often been in the past, but I do not know of any really serious study of the problem. This is not to say that Quine has not given us some important examples of the use of contextual definitions. Rather I would propose a somewhat more neutral course that will allow us to think about the question in a clear way before deciding on the final answer. The point is that the case by case presentation of the contextual eliminations (the cases being governed by the possible contexts) always makes me worry that some type of discontinuity may creep in, that is to say, in one context the virtual object may behave as one kind of thing whereas in another place it may be quite different. Well, that is not unreasonable: no object stands in all the same relationships to all other objects. What disturbs me rather is that the contextual method makes it too easy to allow the entities to be fickle.

What is to be done? Simply this: let us think of the expressions for virtual objects as actually denoting (abstract) entities. Instead of *definitions* we take *axioms* (maybe just by using '$\leftrightarrow$' in place of 'for' – working in the object language of course!). It is a heavier commitment, no doubt about it. Thus if something goes wrong, we have to face the fact that our axioms are inconsistent – we cannot so easily confuse the issue by changing some clause of a 'definition' and calling it harmless. That is a negative advantage. A positive advantage may also be possible: in the case of virtual classes I have shown in my paper on descriptions that the axioms can be made simpler and more elegant than the definition. I feel that this makes virtual classes easier to understand. Of course an elimination metatheorem can be justified on the basis of the axioms; this may often be the case. Thus my advice is to leave the elimination problems to the *development* of the theory rather than having them complicate the *formalization.*

The next question is: should we quantify over virtual objects? I think the answer should be a firm no. Quantification (or the ranges of the bound variables) should only be over the possible individuals. If we have come to value the virtual entities so highly that we want to quantify over them, then we have passed to a *new* theory with a *new* ontology (and with new virtuals also!) The role of the virtual entities is to make clear the structure of the basic domain D, not to introduce a whole new collection of problems. That is why we are happier when references to the virtuals can be eliminated. But maybe virtuals can also be used to simplify the introduction of the *primitive* structure into D, so elimination may not always be possible. Still they must take second place to the individuals in D.

The kind of quantification just rejected was quantification over *all* the virtuals. It may very well be that quantification over *part* of them is useful. Let us take the theory of actual and virtual classes as an example. Every actual class has its *cardinal number*. If a, b are actual classes (sets), let us write $a \approx b$ to mean that a one-one correspondence exists (as an actual relation). Then the cardinal number of a set a can be defined by the equation

$$|a| = \{x : x \approx a\}.$$

In the usual theories of classes $|a|$ is generally virtual but *not* actual.

Among the virtual classes we can define what it means to be a cardinal in the obvious way:

$$\mathrm{Card}\,[\tau] \leftrightarrow \exists x\,[\tau = |x|],$$

where τ is a term and x is not free in τ.

Would it not be rather pleasant to be able to have *variables* ranging over cardinals, say German letters $\mathfrak{m}$, $\mathfrak{n}$..., so that various existence theorems could be stated in the usual way? The answer is simple: yes, you may do so and even not compromise your views on virtual objects. Obviously we simply let

$$\forall \mathfrak{m}\, \Phi(\mathfrak{m}) \leftrightarrow \forall x\, \Phi(|x|),$$

and

$$\exists \mathfrak{m}\, \Phi(\mathfrak{m}) \leftrightarrow \exists x\, \Phi(|x|).$$

The meaning is clearly what was intended, and we see that the quantification is no worse than the usual quantification over sets. The same approach can clearly be applied to quantification over any portion of the virtuals that can be *enumerated* or *indexed* by the sets with the aid of some definable operation like $|a|$. By this method one never quantifies over all the virtuals, however, for no matter what operation $\varphi(a)$ one takes the virtual class $\tau=\{x:x\notin\varphi(x)\}$ is such that $\neg\exists y[\tau=\varphi(y)]$. Still the method seems useful and natural.

In summary then we are distinguishing between *virtual*, *possible*, and *actual*. Let us use V for the domain of virtual objects, D for possible, A for actual, where we assume for simplicity that

$$A\subseteq D\subseteq V.$$

The distinction between A and D does not arise in the usual systems but becomes very critical in modal logic. (In fact, A will have to be replaced by a whole family of domains $A_i\subseteq D$ for $i\in I$). However, even in non-modal logic it can be useful to make the separation (cf. the work of Cocchiarella and of Lambert).

Note that we have already been using the relation of *equality* $(=)$ between individuals – even on V. This is just and proper. Equality *is* a logical notion. Just because people have been in the past confused about the properties of equality does not mean that the notion is basically unclear. We shall be discussing it fully in connection with the modal notions.

II. POSSIBLE WORLDS

This is not the place to discuss where the idea of possible worlds came from, it is sufficient to remark that the recent works of Kanger, Montague, Hintikka, and Kripke have clearly established the usefulness of the notion. Indeed the idea seems so useful that I wish to advocate keeping in mind an extension of the concept, that makes the method much more flexible, or so it seems to me.

My advice is to use the principle of *indexical expressions* of Carnap-Bar-Hillel. The possible worlds thought of as particular collections of individuals with or without additional structure give only one aspect of the idea. Any system of structures can be indexed by the elements of some suitable set, usually in many different ways. Thus we are going to take a fixed set I of these indices and in the first instance index the system of actual individuals by having possibly distinct $A_i\subseteq D$ for each $i\in I$. It is

important not to assume a one-one correspondence between the A_i and the i, for the elements of the set I may possess significant structure not at all reflected in the change from one A_i to another. This is why I reject taking either the models of Montague-Kripke (earlier writings) or the model sets of Hintikka as fundamental. This point has to be argued a little more closely. It is not sufficient to point out that with the models or model sets we can assume enough distinctions to gain the *effect* of any index set (by using several distinct copies of a set or by introducing special predicates into the language) for that is only to admit that the method of indices is in fact more fundamental.

Let us see how the indices work. The point is that in the expressions of our language we make some indefinite reference to the indices – hence the terminology. Already we can give one example, though better examples come soon. This example has to do with quantification. We have the one quantifier $\forall x$ which is interpreted as ranging over D. We now introduce the 'actual' quantifier $\forall.x$, where the dot is meant to suggest an indefinite index. Thus even if we know the meaning of the predicate P we cannot say whether

$$\forall.x\,P(x)$$

is true or false. However, if we specify an $i\in I$, then *relative* to this index the sentence assumes a truth value: namely it is *true* if $P(a)$ is true for all $a\in A_i$. Until we specify the $i\in I$, the range of the quantified variable is not known.

The above example supplies us with a statement whose truth-value is not constant but varies as a function of $i\in I$. This situation is easily appreciated in the context of time-dependent statements; that is, in the case where I represents the instants of time. Obviously the same statement can be true at one moment and false at another. For more general situations one must not think of the $i\in I$ as anything as simple as instants of time or even possible worlds. In general we will have

$$i=(w,t,p,a,\ldots)$$

where the index i has many *coordinates*: for example, w is a *world*, t is a *time*, $p=(x,y,z)$ is a (3-dimensional) *position* in the world, a is an *agent*, etc. All these coordinates can be varied, possibly independently, and thus affect the truth values of statements which have indirect references to these coordinates.

The question arises as to why we cannot simply make the references direct and reduce the logic of indexical expressions to ordinary logic. Well, the answer is that we just do not speak that way. All kinds of statements have indirect references to the here, the now, the I (first person). And these statements have a logic; it is possible to say that some of them are true by virtue of their syntactical form. Thus in view of the commonness and simplicity of these statements, we are certainly obliged to investigate their intrinsic logic. Of course a translation into a more elementary language may help for certain results, but that does not mean that the more involved language has been eliminated or made uninteresting. It is somewhat like the idea of *typical ambiguity*: we need not specify the exact type of variables – only the relative distances between levels. Many statements are true at all levels, and it is only painful to force oneself to carry along the exact levels in the notation.

One could call the $i \in I$ *points of reference* because to determine the truth of an expression the point of reference must be established. Previously I had suggested that terminology as more neutral and more suggestive of the proper generality than *possible worlds*. Maybe *index* is just as good a term, though it seems to me to make them sound rather insignificant. In any case in making up interpretations of the language the set I, along with $A_i \subseteq D \subseteq V$, is to be fixed in advance. Remember it is not fair to shift the size of I in the middle of a discussion (or to invoke unsuspected coordinates) and then deride the logic as inconsistent or counter-intuitive. Such arguments may, if they have any serious content, show an interpretation to be inadequate for representing a complicated situation. That is to say a *particular* interpretation may be discredited, but that does not discredit the *method* of constructing interpretations.

In summary then we conclude that the truth values of statements vary with the $i \in I$. In order to state in a convenient way the connections between statements and their parts some notation is in order. Let us first make the truth-values visible: we write 1 for *true* and 0 for *false*. The reason for this choice of notation is that $2 = \{0, 1\}$ is a simple and readily available symbol for the set of the two truth values. Next associated with a statement Φ will be a *function*, call it $\|\Phi\|$, the *value* of Φ in the interpretation, defined on I with values in 2. In other words we shall write the

equation

$$\|\Phi\|_i = 1$$

to mean that Φ *is true at i*. Other notations are possible, and some variants are discussed later. Note that we use the subscript notation f_i interchangeably with the function-value notation $f(i)$. Sometimes we may wish to refer to the set of all functions from I into 2; it is denoted by 2^I, and we may write

$$\|\Phi\| \in 2^I$$

whereas

$$\|\Phi\|_i \in 2$$

for $i \in I$.

We have already in effect agreed that our language will contain the usual propositional connectives $\neg$, $\vee$, $\wedge$, $\rightarrow$, $\leftrightarrow$ and the quantifiers $\forall$, $\forall.$, $\exists$, and $\exists.$ with other operators to be discussed later. In explaining the meaning of these symbols we give the usual kinds of semantical definitions as follows:

$$(\neg) \qquad \|\neg\,\Phi\|_i = 1 \quad \text{iff} \quad \|\Phi\|_i = 0,$$
$$(\vee) \qquad \|[\Phi \vee \Psi]\|_i = 1 \quad \text{iff} \quad \|\Phi\|_i = 1 \quad \text{or} \quad \|\Psi\|_i = 1,$$
$$(\wedge) \qquad \|[\Phi \wedge \Psi]\|_i = 1 \quad \text{iff} \quad \|\Phi\|_i = 1 \quad \text{and} \quad \|\Psi\|_i = 1,$$

and similarly for the clauses of the truth definition $(\rightarrow)$ and $(\leftrightarrow)$. Note that we do not have to say $\|\neg\Phi\|_i = 0$ otherwise in $(\neg)$ because we have already agreed to the two-valued character of the logic.

To be able to interpret the quantifiers I would advise introducing *constants* $\bar{a}$ into the language corresponding to *all* the $a \in V$. Of course the correspondence between a and $\bar{a}$ must be one-one. We can then write:

$$(\forall) \qquad \|\forall x\, \Phi(x)\|_i = 1 \quad \text{iff} \quad \|\Phi(\bar{a})\|_i = 1 \quad \text{for all} \quad a \in D$$
$$(\forall.) \qquad \|\forall.x\, \Phi(x)\|_i = 1 \quad \text{iff} \quad \|\Phi(\bar{a})\|_i = 1 \quad \text{for all} \quad a \in A_i.$$

Similarly for $(\exists)$ and $(\exists.)$. Now I know that some people do not like to have their languages made uncountable with too heavy a 'telephone book' of individual names. That was one of the reasons Tarski introduced his version of the definition of satisfaction. I have worked with both kinds of definitions and have found the style with individual constants *much easier* to *communicate* to those unpracticed in set theory. For one thing the clause of the truth definition is shorter.

So far we have only discussed the connections between and quantification of propositions. The nature of the interpretation of atomic formulas and terms (like descriptions) deserves a section of its own.

III. INDIVIDUAL CONCEPTS

Why should we not throw ourselves into the midst of the battle at once by discussing *definite descriptions*. They illustrate all the problems and lead directly to the general operators. So first of all what is the notation? We have had $\imath$, T, U to suggest *the* and *unique*, but I do not care for any of these. It is a matter of taste, but since I am using inverted sans-serif $\forall$ and $\exists$ for the quantifiers, I have finally decided to use an (inverted?) sans-serif capital I to suggest the word *individual*. (Actually T is not bad, but then what about T and F, or $\top$ and $\bot$, for the truth values?) The semantics is rather obvious no matter what the notation, for we have:

$$(\mathsf{I}) \qquad \|\mathsf{I}x\, \Phi(x)\|_i = a \quad \text{iff} \quad \{a\} = \{b \in D : \|\Phi(\bar{b})\|_i = 1\}.$$

That is to say, given $i \in I$ so that we know how to evaluate $\|\Phi(\bar{b})\|_i$ for each $b \in D$ we can ask whether there is a *unique* b making the formula true. If so, call it the individual a and that will be the *value* of the descriptive phase *relative to i*. Hence, $\|\mathsf{I}x\Phi(x)\|$ is being defined as a *function* from I into D. But what kind of a function? What of the *improper* descriptive phrases? Answer: the function is most naturally taken as a PARTIAL FUNCTION; improper descriptions have 'undefined' values.

The opinion just presented is a change of mind over what I have been advocating previously (even in my original Irvine lecture!). It was only when I started to set down in an informal, conversational way my accumulated suggestions that I saw how to take full advantage of the idea (yet to be explained here in detail) of *individual concepts*. The scheme will become clearer as we go along.

Note that along with I we also have the operator $\mathsf{I}.$ which picks out actual individuals and is defined in the obvious way by:

$$(\mathsf{I}.) \qquad \|\mathsf{I}.x\, \Phi(x)\|_i = a \quad \text{iff} \quad \{a\} = \{b \in A_i : \|\Phi(\bar{b})\|_i = 1\},$$

where we take the value to be *undefined* in case there is no a satisfying the condition on the right. Of course this operator is not really fundamental because it is so easily defined in terms of the other notions:

$$\mathsf{I}.x\, \Phi(x) = \mathsf{I}x\exists.y\, [y = x \wedge \Phi(y)].$$

This last remark reminds us that we have not discussed equality. As usual our language has *terms* and *formulas* combined by the usual particles as well as by some unusual ones yet to be introduced. In particular two terms τ and σ can be combined into an equation $\tau = \sigma$ which is a formula. We are going to be able to evaluate every well-formed expression without free variables, so we may assume that we know what $\|\tau\|$ and $\|\sigma\|$ are: namely partial functions from I into —— into what? So far for descriptions the values have only been in D but in general the values should clearly lie in V. Let us write $V^{(I)}$ for the set of *partial* functions whereas V^I would denote the set of *total* functions, so that

$$\|\tau\|,\ \|\sigma\| \in V^{(I)}.$$

Now for equality:

$(=)$ $\quad \|\tau = \sigma\|_i = 1$ iff either $\|\tau\|_i$ and $\|\sigma\|_i$ are both defined and equal or neither are defined.

Yes, yes, I can hear the objections being shouted from all corners. If one is going to use undefined terms why not undefined truth values? Is not that more natural? Maybe so, but I have yet to see a really workable three-valued logic. I know it can be defined, and at least four times a year someone comes up with the idea anew, but it has *not* really been developed to the point where one could say it is pleasant to work with. Maybe the day will come, but I have yet to be convinced. So my advice is to continue with the two-valued logic because it is easy to understand and easy to use in applications; then when someone has made the other logic workable a switch should be reasonably painless.

On the basis of $(=)$ it is obvious that the following two formulas are always true for all $i \in I$:

$$\tau = \tau, \quad \text{and} \quad [\tau = \sigma \wedge \tau = \theta \rightarrow \sigma = \theta],$$

which was the reason for the choice of the definition. As will become apparent as we go along the following generally *fails*:

$$\tau = \sigma \wedge \Phi(\tau) \rightarrow \Phi(\sigma),$$

which seems disturbing. However, an important special case does hold. First we must say what to do with (virtual) individual constants:

$$\|\bar{a}\|_i = a$$

for all $a \in V$. That is surely the intended meaning of $\bar{a}$. Now it is easy to check that

$$\bar{a} = \bar{b} \wedge \Phi(\bar{a}) \rightarrow \Phi(\bar{b})$$

is always true. In particular this slightly weaker statement also holds:

$$\forall x\, \forall y\, [x = y \wedge \Phi(x) \rightarrow \Phi(y)].$$

To understand better how the semantics goes let us take over (more or less) the following terminology from Carnap: we call the elements of 2^I *propositional concepts* and those of $V^{(I)}$ *individual concepts*. Note that

$$A_i^{(I)} \subseteq D^{(I)} \subseteq V^{(I)}$$

so that we have a classification of our individual concepts parallel to that for individuals. We call these things *concepts* because they have values depending on the index or point of reference in I and together form a coherent range of values, that is to say a *function*. Such a stringing together of individuals (or truth values) gives us, not a single individual, but a concept of an individual relative to the point of reference. Everyone is familiar with this from such old descriptive phrases as 'the present king of France'. Note that V is constructed *first* and $V^{(I)}$ constructed *second* to be able to evaluate terms. That is to say the individual concept is a semantical construct rather than a primitive ontological notion.

To repeat: for terms τ and formulas Φ:

$$\|\tau\| \in V^{(I)} \quad \text{and} \quad \|\Phi\| \in 2^I.$$

Thus $\|\tau\|$ is somewhat like Frege's *sense* while $\|\tau\|_i$ is like *denotation*. But not exactly. Let us not get into that discussion here. Maybe it would be better to be a little more like Carnap and call $\|\tau\|$ the *intension* of τ and $\|\tau\|_i$ the *extension* of τ at $i \in I$. The basic principle of our semantics is:

> *The intension of a whole expression is determined by the intensions of its parts.*

The same may very well *not* be true of the extension. This is how we explain the 'paradox' of equality: the natural reading of $\tau = \sigma$ is only that

τ and σ have the same extension (if any) relative to the current point of reference. This does *not* in general imply equality of intensions. Hence we cannot have a general substitutivity of equals – unless like true individual constants the intension is indeed a *constant* function.

Returning now for a moment to descriptions we note the validity of:

$$\forall y\, [y = \text{\i} x\, \Phi(x) \leftrightarrow \forall x\, [x = y \leftrightarrow \Phi(x)]].$$

This is the basic property of descriptions that has been advocated by many of us (Hintikka, Van Fraassen, Lambert, the present author, to name only a few), and by now it seems natural. According to the way I have treated individual concepts we also have a definite result for the improper descriptions. We see that $\|\text{\i} x[x \neq x]\|$ is the *totally undefined* function in $V^{(I)}$; let us call it $*$. This concept $* \in V^{(I)}$ works very much like Frege's null-entity and it now seems to me to be *more natural* than an *arbitrary* choice of a null element in $V \sim D$. For one thing it is uniquely determined: it can be called a semantical construct, which seems to put it in its proper place. We find, however, the same valid sentence we had before when $* \in V \sim D$:

$$\neg \exists y\, [y = \text{\i} x\, \Phi(x)] \rightarrow \text{\i} x\, \Phi(x) = \text{\i} x\, [x \neq x].$$

IV. INTENSIONAL OPERATORS

So far we have been moving along on a very pure logical level: all the notions (except for individual constants) have been logical notions. It is now time to discuss the *non-logical* notions and to be led by this discussion even to further logical notions.

What about (atomic) predicates? Let us discuss binary relations as a prime example. Our only example so far has been equality and it is somewhat special. Note first, however, that just as with equality we should allow for expressions

$$\tau\, \mathbf{R}\, \sigma$$

where τ and σ may be arbitrary terms. (We certainly want to be able to say such things as 'Yvonne is the wife of the present king of France'.) Now this brings up the question as how we are to interpret $\mathbf{R}$; call the interpretation $\|\mathbf{R}\|$. Is it to be a relation between individuals *or* individual concepts? Answer: the latter. Why? Because even if we only start with individual relations, we will be able to define an intensional relation that depends on the whole intension of each of the arguments. (This will be obvious as soon as we have the modal operators.) Therefore, in general we should take

$$\|\mathbf{R}\| \in (2^I)^{V(I) \times V(I)}$$

so that in the corresponding clause in the truth definition we have

$$(\mathbf{R}) \qquad \|\tau\, \mathbf{R}\, \sigma\| = \|\mathbf{R}\|\,(\|\tau\|\,,\, \|\sigma\|)\,.$$

Similarly for predicates with more places. Note that the subscript i is unnecessary here because we took the values of $\|\mathbf{R}\|$ to belong to 2^I.

The main reason for making general predicates so complicated was not simply to torture the reader but to make our logic *general*. When we know that a formula is *logically valid* (that is, true in all interpretations relative to all points of reference of the interpretation), then it should remain valid when *formulas* are substituted for *predicate letters*. There is no other convention that is reasonable. But that does not keep us from discussing special kinds of predicates, as is clear.

One special kind of predicate is *extensional*. The best examples come from ordinary relations $R \subseteq V \times V$. (Remember: $2 < +\infty$, so that our relations must be defined over all of V.) Let us write $\bar{R}$ for the constant in the language that corresponds to R in analogy with individuals. Then

$$(\bar{R}) \qquad \|\tau\, \bar{R}\, \sigma\|_i = 1 \quad \text{iff} \quad (\|\tau\|_i\,,\, \|\sigma\|_i) \in R\,.$$

Note for this atomic formula to be true, both $\|\tau\|_i$ and $\|\sigma\|_i$ must be defined. This convention was not what we chose to do for equality – because the special role of equality favors a different convention making for more regularity among the valid formulas. (An interesting sidelight: if all our atomic formulas were either made with $=$ or various $\bar{R}$'s and if our only terms were either variables or descriptions, then descriptions could be entirely eliminated. In general this is not so.) For example the *extensionality principle* (for $\bar{R}$) reads

$$\tau = \tau' \wedge \sigma = \sigma' \wedge \tau\, \bar{R}\, \sigma \rightarrow \tau'\, \bar{R}\, \sigma'\,.$$

Not every extensional predicate in this sense comes from an ordinary R. The reader can easily work out for himself other examples, and we shall return to this question below.

Intensional relations do not spring to mind as easily as do intensional propositional operators. The simplest example is $\Box$, *necessity*, defined by

$$(\Box) \qquad \|\Box\Phi\|_i = 1 \quad \text{iff} \quad \|\Phi\|_j = 1 \quad \text{for all} \quad j \in I$$

This means nothing more or less than true in all possible worlds. What should we call it? *Logical necessity* does not seem correct since it depends on I (as do the values $\|\mathbf{R}\|$). How about *universal necessity*? Because we are in effect quantifying over our universe I. Whatever we call it we find that it is an S5 modal operator, and a very useful one. Of course possibility is defined in the dual way and we do not need to write down here the clause ($\Diamond$) of the truth definition.

If we want we can say that $\Box$ has a value $\|\Box\|$ just as we did for $\|R\|$. Indeed $\|\Box\|$ is of the same logical type as $\|\neg\|$:

$$\|\Box\|\,,\ \|\neg\| \in (2^I)^{2I}$$

They are mappings from propositional concepts to propositional concepts. The difference is that one is extensional and the other is not. Note that we can write

$$\|\Box\Phi\| = \|\Box\|\,(\|\Phi\|)$$

(without subscripts!) making our dogma about intensions quite evident.

Now that our logical notation is becoming more interesting it is useful to examine certain compound phases: for example $\Box\tau = \sigma$. By definition we have:

$$\|\Box\tau = \sigma\|_i = 1 \quad \text{iff} \quad \|\tau\|_j = \|\sigma\|_j \quad \text{for all} \quad j \in I\,,$$

where we agree that the equation on the right holds in case *both* sides are undefined. That is to say $\|\Box\tau = \sigma\|$ is the concept of the true proposition iff $\|\tau\| = \|\sigma\|$ which means $\|\tau\|$ and $\|\sigma\|$ are the *same* individual concept. Let us introduce a symbol for this notion:

$$\tau \equiv \sigma \leftrightarrow \Box\tau = \sigma\,.$$

Thus we are distinguishing between extensional equality ($\tau = \sigma$) and intensional equality ($\tau \equiv \sigma$). Let us call the stronger notion *identity* for short.

In view of our dogma of intensions (which can be formally justified from the semantics) we find that if τ and σ have no free variables, then:

$$\tau \equiv \sigma \wedge \Phi(\tau) \rightarrow \Phi(\sigma)$$

is valid for identity; while the corresponding principle for equality fails. Similarly if Φ and Ψ are without free variables and Θ' results from Θ by replacing occurrences of Φ by Ψ, then

$$\Box[\Phi \leftrightarrow \Psi] \wedge \Theta \rightarrow \Theta'$$

also holds. Thus our logic has a certain degree of extensionality: *logically equivalent* formulas are substitutible. Some people feel that a truly intensional logic should reject even this principle. I have not seen a clear semantical discussion of such a logic. Usually there is a confusion between propositional operators and predicates of sentences; that is, the object language must formalize a part of syntax. Even then I have not seen any reasonable general system. My advice is to work for a while on this intermediate intensional level, because I shall try to argue in the final section that there is much to do involving notions of interest.

The next most interesting combination involves $\Box$ and $\forall$. This is valid:

$$\Box\forall x\, \Phi(x) \leftrightarrow \forall x\, \Box\Phi(x),$$

This is *not*:

$$\Box\forall.x\, \Phi(x) \leftrightarrow \forall.x\, \Box\Phi(x),$$

indeed, *neither direction* is valid for $\forall.$. The counterexamples are easy to come by. There has been so much fuss over these principles in the last few years, and it was all unnecessary: there *are* two kinds of quantifiers with *different* properties. All we needed to do was to make the semantics clear. (This point has been made by Kripke and others – but it does not hurt to make it again.)

To make the counterexample more obvious for the quantifier problem just mentioned it is useful to discuss extensional predicates again. Suppose that to each $i \in I$ we attach a relation $R_i \subseteq V \times V$ with possibly different R_i's for different i's. (We can treat R as a relation-valued function). Then we can introduce into our language a constant $\bar{R}$ (the bar $^{-}$ suggests a *constant* relation, the tilde $^{\sim}$ a *variable* relation) where the meaning is given by

$(\bar{R})$ $\quad \|\tau\, \bar{R}\, \sigma\|_i = 1 \quad \text{iff} \quad (\|\tau\|_i, \|\sigma\|_i) \in R_i$

Indeed $A_i \subseteq D$ is just such a variable one-place predicate so we can write

$(\tilde{A})$ $\quad \|\tilde{A}\tau\|_i = 1 \quad \text{iff} \quad \|\tau\|_i \in A_i.$

We then find the following valid:

$$\forall.x\, \Phi(x) \leftrightarrow \forall x\, [\tilde{A}x \rightarrow \Phi(x)],$$

that is to say $\forall.$ is a *restricted quantifier*. Now it is easy to see why

$$[\tilde{A}x \rightarrow \Box\Phi(x)]$$

and

$$\Box[\tilde{A}x \rightarrow \Phi(x)]$$

are independent, which accounts for the phenomenon we noted.

The use of variable predicates in the way just indicated seems to me to be preferable to the manner in which Van Fraassen suggested using what he calls *logical space*. If I understand his method, he treats D (which may as well be equal to V for the moment) as a set of abstract place-holders for individuals. I imagine something like a giant egg crate. Defined on this space are certain *fixed* predicates. The actual individuals are thought of as being mapped (or stored) into the set D. Next a family of transformation $T = \{t_i : i \in I\}$ is given, and we can follow an individual around D by seeing how $a \in D$ is transformed into $t_i(a)$. Of course, a and $t_i(a)$ have different properties with respect to the fixed predicates. There is just something about this idea that strikes me as being inconvenient – or at least not general enough. I would imagine we can obtain the same effect by replacing the fixed relation R by the family R_i of relations defined by

$$a\, R_i\, b \leftrightarrow t_i(a)\, R\, t_i(b)$$

and by keeping the individuals fixed in place. The point is that I feel that Van Fraassen is mixing the variability of the predicates together with the interpretation of his modal operator. These are things that I feel should be kept separate, and so my advice is to adopt the kind of model structure presented here where the modal operators can be treated in isolation.

For example we may assume that I is a system with a binary mode of composition $\circ$, where, if we like the transformational approach, we may assume

$$t_j(t_i(a)) = t_{i \circ j}(a).$$

We can then define a propositional operator $\boxed{\rightarrow}$ by:

$(\boxed{\rightarrow})$ $\quad \|\boxed{\rightarrow}\, \Phi\|_i = 1 \quad \text{iff} \quad \|\Phi\|_{i \circ j} = 1 \quad \text{for all } j \in I.$

That will give a fairly interesting operator, but of course it is one of a rather special kind – it would come up in tense logic for example. It is only one of many different types, however.

More generally let $\rho \subseteq I \times I$ be a binary relation on I, corresponding to ρ is the operator $\boxed{\varrho}$ defined by

$$(\boxed{\rho}) \qquad \|\boxed{\rho}\Phi\|_i = 1 \quad \text{iff} \quad \|\Phi\|_j = 1 \quad \text{for all} \quad j \in I \quad \text{where } i\,\rho\,j\,.$$

In other words, ϱ is simply an *alternative relation* on possible worlds. We are by now very familiar with all the work done on this type of operator. It must be stressed, however, that this is not the only way to obtain interesting operators; though it does include the previous example.

The simplest example of an operator which is not naturally specified by an alternative relation occurs in tense logic. Suppose I is taken as time (i.e. the set of real numbers). We define the *progressive tense operator* $\boxed{\leftrightarrow}$ by:

$$(\boxed{\leftrightarrow}) \qquad \|\boxed{\leftrightarrow}\,\Phi\|_i = 1 \text{ iff there is an open interval } J \subseteq I \text{ with } i \in J \text{ such that } \|\Phi\|_j = 1 \quad \text{for all} \quad j \in J\,.$$

Long ago McKinsey and Tarski showed that the logic of $\boxed{\leftrightarrow}$ is exactly S4. Now on some other index set I we can capture S4 with the help of an alternative relation – but on the ordinary reals there is no way of obtaining the same $\boxed{\leftrightarrow}$.

It is by now only a small step to the most general monadic propositional operator. Associate with each $i \in I$ a family $\mathscr{I}_i$ of *subsets* of I. We then can define the corresponding operator:

$$(\boxed{\mathscr{I}}) \qquad \|\boxed{\mathscr{I}}\Phi\|_i = 1 \quad \text{iff} \quad \{j \in I : \|\Phi\|_j = 1\} \in \mathscr{I}_i,$$

which we see is only a trivial reformulation of the statement that

$$\|\boxed{\mathscr{I}}\| \in (2^I)^{2^I}.$$

But somehow it reads differently in terms of sets. We note that in the example of $\boxed{\leftrightarrow}$, the family $\mathscr{I}_i$ should be taken as the collection of all subsets of I containing the point $\mathscr{I}_i$ in their interiors; the neighborhoods of i in other terminology.

Among all the various operators $\boxed{\mathscr{I}}$ it is clear that only very few should be called *logical operators* – that is those that do not depend on the structure of I. With a proper definition (invariance under permutation of I, for example) we can no doubt show that the only logical operators are $\neg$, $\neg\neg$, $\Box$, and $\Diamond$. This of course does not mean that the others are uninteresting – far from it. Here is what I consider one of the biggest mistakes of all in modal logic: concentration on a system with just *one* modal operator. The only way to have any philosophically significant results in deontic logic or epistemic logic is to combine those operators with: tense operators (otherwise how can you formulate principles of change?); the logical operators (otherwise how can you compare the relative with the absolute?); the operators like *historical* or *physical* necessity (otherwise how can you relate the agent to his environment?); and so on and so on. But where to stop? This list can be extended further and further. One must stop somewhere, but to stop the list at *one* is obviously missing out on something important. The point I am trying to make is that the semantics being explained here allows for several operators side by side in a simple convenient, and natural way: one has only to think what coordinates $i = (w, t, p, a, \ldots)$ one wants. Furthermore one should not forget the logical operators.

This last point is one on which I am somewhat critical of Hintikka. He uses only one operator (which I would write here as $\boxed{\rho}$) and tries to express the various conditions for substitutivity and existence in terms of equations $\boxed{\varrho}^n\tau = \sigma$ with iterated modalities. I feel that this is making the alternative relation do too much work: the fundamental condition for substitutivity is $\Box\tau = \sigma$. The conditions involving the weaker relation should be presented as *metatheorems* concerning *special* contexts (special formulas $\Phi(\tau)$) rather than being emphasized as part of the basic semantics. It seems to me that the present approach provides quite a suitable place for Hintikka's useful and important insights.

V. VARIABLE BINDING OPERATORS

Up to this point we have been discussing at most intensional relations or intensional operators on propositions. Though it is painful, a certain greater degree of generality is called for. In particular the extensional operators like the quantifiers and the discriptive operators have intensional counterparts. Let us consider first an operator \$ which binds one variable and operates on one term and one formula: we can employ it in the language with the following format:

$$\$x\,[\tau(x),\,\Phi(x)]\,.$$

Now we have our choice: the result is either a *term* or a *formula*.

Let us take the first case where \$ makes terms. Now the values of this compound term must be allowed to be virtual (cf. the expression

$$\bigcup_{x\in A}\tau(x)$$

in class theory: its being a set depends on A and $\tau(x)$ in an essential way). But we have an intensional logic so (virtual) individual concepts are called for as values. Thus

$$\|\$x\,[\tau(x),\,\Phi(x)]\| \in V^{(I)}$$

is required. Now how does this value depend on $\tau(x)$ and $\Phi(x)$? Well, these expressions represent *functions* of x and *all our bound variables* range over D. This leads us to the statement that

$$\|\$\| \in (V^{(I)})U, \text{ where } U = (V^{(I)D}) \times (2^I)^D.$$

because $(V^{(I)})^D$ is the set of all concept-valued functions and $(2^I)^D$ is the set of all propositional functions on D. Thus

$$(\$)\qquad \|\$x\,[\tau(x),\,\Phi(x)]\|_i = \|\$\|\,(f,F)_i\,,$$

where $f\in(V^{(I)})^D$ and $F\in(2^I)^D$ are defined by:

$$f(a)_i = \|\tau(\bar{a})\|_i$$

and

$$F(a)_i = \|\Phi(\bar{a})\|_i\,, \quad \text{for all } a\in D.$$

It is a little hard to give really dramatic examples of such operators which stand by themselves: usually they are compounded out of other, simpler operators. For example

$$\text{I}y\,\forall x\,\Box\,[\Phi(x)\to y=\tau(x)]$$

would be a likely compound. (Do not ask for a poetic reading of that descriptive phrase, please). It might be possible to always reduce an operator to a compound where the only variable bindings came in the quantifiers and descriptions; but then again maybe not. The matter bears

some further thought. Note that in case \$ were a formula maker, the only change would be to make its values propositional concepts.

To emphasize again the kind of interchangeability inherent in this logic we have the following as valid:

$$\forall x\,\Box\tau(x)=\tau'(x)\wedge\forall x\,\Box\,[\Phi(x)\leftrightarrow\Phi'(x)]\to$$
$$\$x\,[\tau(x),\,\Phi(x)] = \$x\,[\tau'(x),\,\Phi'(x)]\,.$$

And at risk of repetition, let it be noted that $\Box$ is the *only* modal operator that is suitable for this purpose: that is one of the main reasons for including it among the *logical* constants.

An important consideration that has been neglected up to this point concerns the use of various interpretations. Indeed we have not actually said what an interpretation *is*. This must be remedied at once. An *interpretation* assigns values to the symbols according to their logical types. To specify an interpretation one must first specify V, D, I, and the A_i for $i\in I$. Next for each symbol \$ one must specify its value. What are the symbols? Well at a minimum they are the $\bar{a}$ for $a\in V$; the $\neg$, $\to$, $\forall$, I, $=$; the special $\bar{A}$; the $\Box$; and finally all the non-logical symbols \$ which include all the terms and formula makers, with and without bound variables, in whatever number we choose. The first listed are logical (at least they are logical with respect to the given domains of virtual, possible, actual individuals and the indices.) The remainder are non-logical. The values of the logical symbols are *fixed*; the values of the non-logical symbols are open to *variation*. So far I have *not* indicated this freedom in my notation. Let us then call the given interpretation $\mathfrak{A}$ (a German capital A). It is some God-awful multi-tuple, which has hidden in it all the information about the domains, indices, and non-logical symbols. Once we know $\mathfrak{A}$ we know everything. All our semantical rules (like ($\forall$), ($=$), ($\Box$), (\$)) are energized and we can find the value of any well-formed expression (without free variables). The notation I would use is:

$$\|\tau\|^{\mathfrak{A}} \quad\text{and}\quad \|\Phi\|^{\mathfrak{A}},$$

whereas Montague has recommended:

$$\tau_{\mathfrak{A}} \quad\text{and}\quad \Phi_{\mathfrak{A}}.$$

His is shorter – too short it seems to me. The notation leaves us nothing to write when mention of $\mathfrak{A}$ is suppressed. (Of course to some, suppression is evil, and they would never consider doing it.) Thus I prefer the writing of the double bars as forcing me to remember the distinction between the *expression* and its *value*. That clearly is the kind of advice that one can either take or leave: all I ask it that you be reasonably clear about what you are doing. I only hope I have made the *content* of the present approach definite enough.

Having admitted the variability of interpretations, the definition of *logical validity* becomes more definite. Let us state it for *formulas* $\Phi(x, y, \ldots)$ with *free* variables. Such a formula is *valid* if and only if for all interpretations $\mathfrak{A}$ and all individuals $a, b, \ldots \in D$ (the possible individuals of $\mathfrak{A}$) we have

$$\|\Phi(\bar{a}, \bar{b}, \ldots)\|_i^{\mathfrak{A}} = 1$$

for all $i \in I$ (the index set of $\mathfrak{A}$). Since our logic has been first-order logic (only bound individual variables) it will be no surprise to hear that the valid formulas are axiomatizable. The axioms and rules are in fact well-known: predicate logic with an S5 modal logic for $\Box$; in addition the axiom schema of replacement of *necessary* equalities (and biconditionals) is required. David Kaplan and I will present his completeness proof (a standard type of Henkin argument) in our joint paper devoted to the more technical details. Thomason has also made a detailed study of the completeness proofs in the paper he read here at Irvine. His systems are, however, only fragments of the general system described in this paper – though the methods he developed will apply. One bit of advice seems needed at this juncture: the aim of logic is not solely to provide completeness proofs. The real aim is conceptual clarification. Completeness proofs are needed but are not ends in themselves.

One small point concerning the proper formalizations can be mentioned here. It will be noted in the definition of logical validity that the free variables were replaced by constants for elements $a \in D$ and *not* $a \in V$. The reason being that individual variables are for the individuals over which one quantifies. The Greek letters τ, σ, ... in the *metalanguage* are the ones that can be replaced by expressions for virtual entities. In foundational studies I advise *not* making the individual variables do double duty. (For pedagogical purposes I found that the double use *is* better because students dislike too heavy use of Greek letters!)

The next step is to push on to the higher-type logic. For the present

I can only refer to the latest papers of Montague for a discussion. It is interesting to note that in his Helsinki lecture Lemmon discussed modal set theory *without* giving semantics. That can now possibly be done, but I have not worked out the details in full.

VI. INCIDENCE VS. EQUALITY

We have already been led to the distinction between equality and identity. It took me a very long time to understand what was involved inasmuch as I was confused about a *weaker* relation I shall write as

$$\tau \approx \sigma$$

and call *incidence*. I was pleased (in a sense) to find that the same confusion has been the crux, at least in my opinion, of the recent interchanges between Hintikka and Føllesdal. It would seem to be finally the proper time to clear up the distinction through the use of the semantics advocated here. In fact, I feel that the present approach is better suited to providing the needed clarification than are Hintikka's model sets – though he can claim other advantages for his method. The point is that we must clearly distinguish between individuals and individual concepts – which our semantics does. I do not feel that the situation is as transparent with Hintikka, but maybe it can be made so. I am completely indebted to Montague and Kaplan for showing me the way to the present formulations.

The idea intuitively is that two individuals that are generally distinct might share all the same properties (of a certain kind!) with respect to the present world (present index). Hence they are equivalent or *incident* at the moment. Relative to other points of reference they may cease to be incident. This will allow for branching and merging of individuals – at least as far as distinctions concerned with certain kinds of properties permit. My feeling is that the confusion about merging of individuals rests on this question of the *allowed properties*: if the class of properties is vague and changeable, then the equivalence of individuals will seem paradoxial. We have to agree to a class of properties given in advance – or at least agree that the notion is relative to a fixed class of properties that we wish to investigate more closely. I am reasonably certain that this is what Hintikka has in mind. Unfortunately he uses the equality symbol = for this incidence relation, which leads to unnecessary arguments when others read $=$ in a different way. We shall see why it is tempting to use $=$ in this way in a moment.

How can we represent the idea of incidence in our semantics? We could of course add a non-logical constant $\approx$ and the axioms of an equivalence relation. This is unsatisfactory because no particular analysis is involved. A better approach is the specialization of the class of models (interpretations) to those with a more defined structure, the features of which can be interpreted in an intuitive and suggestive way. So we begin by supposing that to each individual $a \in V$ and relative to each $i \in I$ we can attach certain *particulars* that specify or individuate a with respect to the properties involved. We take these particulars as forming some kind of abstract object, call it the *state* of a at i, and for convenience let S be a set having all these states of all the individuals as members. *Now this next step may be an oversimplification*: we then IDENTIFY the individual a with this individuating function. Thus we make $V \subseteq S^I$, and for each $a \in V$ and $i \in I$ write a_i for the state of a at i. The meaning of incidence is now determined:

$$(\approx_\circ) \qquad \|\bar{a} \approx \bar{b}\|_i = 1 \quad \text{iff} \quad a_i = b_i .$$

(I have written $\approx_\circ$ because this clause involves only individual constants; the full version for compound terms will be given later.)

It is seen that we are assuming that, to know an individual, it is sufficient to know all particulars about him from all points of reference. In words that does not sound too bad, but it is an assumption that there are *enough* points of reference and sufficiently detailed particulars. I can imagine in later development that we may drop this condition; but for present purpose of trying to see how a coherent incidence relation can be introduced, the convention does not seem too unreasonable.

Note that equality of individuals, under our convention, can be defined in terms of incidence because the following is now valid:

$$\bar{a} = \bar{b} \leftrightarrow \Box \bar{a} \approx \bar{b}$$

It is this feature that undoubtedly led Hintikka to take $\approx$ as more fundamental than what I write as $=$. The relationship between the two notions is somewhat obscured by Hintikka, however, because he uses an operator $\boxed{P}$ instead of the stronger $\Box$, and this introduces some complications. In any case, I want to argue now that no matter how the modalities are treated we *cannot* take $\approx$ and define $=$.

The trouble comes in with the individual concepts. I hope that I have finally made this distinction between individuals and individual concepts clear to everyone. At least for me it was the semantic approach and the suggestions of Montague and Kaplan that made the clarity possible. You see, even if we treat individuals as *functions* in S^I, the concepts enter in a secondary way (on top of the individuals, as it were) and have values which are functions with the *individuals* as values. It seems to me now that this is really the only proper arrangement. Thus for a term τ we will find that

$$\|\tau\| \in (S^I)^{(I)} .$$

(The value of a term, remember, is a *partial* function. It seems better to take the $a \in V$ as *total* functions $a \in S^I$, because an individual can always be given *some* abstract entity as its state no matter which $i \in I$ is chosen.) This requires the following semantics:

$$\|\tau \approx \sigma\|_i = 1 \quad \text{iff} \quad (\|\tau\|_i)_i = (\|\sigma\|_i)_i .$$

That is, find out which individuals τ and σ denote ($\|\tau\|_i$ and $\|\sigma\|_i$), and *then* see if they have the same state. Everything is fine (i.e. the truth value is well-determined) if both $\|\tau\|_i$ and $\|\sigma\|_i$ are defined; otherwise our convention for *truth* is that they *both* be undefined.

Now for the trouble: what does $\tau \approx \sigma$ mean in general? Formally we have:

$$\|\Box\tau \approx \sigma\|_i = 1 \quad \text{iff} \quad (\|\tau\|_j)_j = (\|\sigma\|_j)_j \quad \text{all} \quad j \in I .$$

Clearly this is *not* the same as $\|\tau\| = \|\sigma\|$ *unless* $\|\tau\|$ and $\|\sigma\|$ are *constant* functions in $V^{(I)}$. Let us try a homely example: take τ to be 'the President of the United States' and σ to be 'the biggest crook in the world'. (In our interpretation any similarity to persons living or dead is purely coincidental.) For simplicity we assume I to involve *time*. Thus for each $i \in I$, the individual $\|\tau\|_i$ is the (whole) person who is President at time i. If we like we can think of $(\|\tau\|_i)_i$ as an instantaneous portrait of the man which captures on very sensitive paper his essential qualities at time i. As i varies in I we run through the (rather extensive) official portrait gallery dedicated to the office of president. Meanwhile the FBI, acting quite on its own, has been photographing at each instant the biggest crook of the moment. One can imagine the consternation when a comparison reveals the truth of $\Box\tau\approx\sigma$. One can charitably excuse the President himself, however, for it is only by virtue of his high office that he must be put in that other distasteful category. The statement $\Box\tau\approx\sigma$, though unpleasant, is perfectly meaningful and does not entail $\tau=\sigma$.

Why does not it? As I write the above I notice a certain weakness of argument. Thinking of I as simple time may not be enough. Our elementary view of individuals is such that if $a_i=b_i$ at *one* time i, then $a=b$ (all times!). We have to consider not only of the flow of time but also of *alternative* courses of events. – No, come to think of it, that is not the answer either, for that only makes the individual concept $\|\tau\|$ fatter but not more amusing. – Or maybe it does. (Oh my, I see that much more thought and experimentation are needed to make the ideas into something useful. In any case I feel that a precise and general semantical framework is essential, and that is, as I have been trying to indicate, now available.) Let us attempt to save this example. We can imagine as a possible individual the stirling public servant, possessor of mountains of good qualities from every point of reference. Next imagine a master crook. These are not individual concepts but possible individuals – distinct individuals. The public servant is not always president; the crook is not always successful. But at that moment of triumph when both reach the high points of their careers we find that they are the same in all particulars.

No, it just will not do. I am twisting the language to fit an abstractly constructed semantics! Nevertheless I still think that the relation $\approx$ has a sensible meaning. The trouble is that I have been trying to defend the principle:

$$\bar{a}=\bar{b}\leftrightarrow\Box\bar{a}\approx\bar{b},$$

and the defense has gone wrong as I have tried to write it down. It seems to me that one must read $\bar{a}\approx\bar{b}$ as '$\bar{a}$ and $\bar{b}$ are equivalent (indistinguishable) with respect to certain (fixed) properties'. Indeed, as Hintikka has pointed out, there may be several different ways of individuating our individuals (different perceptions) which would provide for several equivalence relations $\approx_1, \approx_2, \dots, \approx_n$, in my framework. Whether any of these should satisfy the above biconditional is a moot point. But let me again stress that we must take care not to confuse equivalence and equality. We must, on my view, make the distinctions which can be summarized in the following diagrams where the arrows indicate implications. Note

that the diagrams are *different* for individuals and individual concepts. The biconditionals are all firmly justified; the single arrows are clearly not reversible – except for the questionable $\dashrightarrow$. I guess we will come to reject that also.

$$\begin{tikzcd}
\bar{a}\equiv\bar{b} \arrow[d, leftrightarrow] \\
\Box\bar{a}=\bar{b} \arrow[d, leftrightarrow] \\
\bar{a}=\bar{b} \arrow[d, shift right] \\
\Box\bar{a}\approx\bar{b} \arrow[u, dashed, shift right] \arrow[d] \\
\bar{a}\approx\bar{b}
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
 & \tau\equiv\sigma \arrow[d, leftrightarrow] & \\
 & \Box\tau=\sigma \arrow[dl] \arrow[dr] & \\
\Box\tau\approx\sigma \arrow[dr] & & \tau=\sigma \arrow[dl] \\
 & \tau\approx\sigma &
\end{tikzcd}$$

The puzzle of cross indentifications of individuals has been treated by David Lewis from another point of view. He thinks of

$$D=\bigcup_{i\in I}A_i$$

and of the A_i's as being disjoint:

$$A_i\cap A_j=0 \quad \text{for} \quad i\neq j.$$

That is, each world $i\in I$ has its own separate set of tokens representing individuals. Having made the individuals separate, we now join some of them together again by the *counterpart* relation C. If $a\in A_i$ and $b\in A_j$, then aCb means that a is the *counterpart* of b in A_i: the person b would be if he lived in A_i, or at least one who is very much like him – we will allow for several. In case $i=j$, we agree that the reading of the relationship demands $a=b$.

We can easily put Lewis' relation into our language by using the constant $\bar{C}$ as with our previous notation. These principles will then be satisfied:

$$\forall x, y\, [x\, \bar{C}\, y \rightarrow \tilde{A}x],$$
$$\forall x, y\, [x\, \bar{C}\, y \wedge \tilde{A}y \rightarrow x = y],$$
$$\forall x\, [\tilde{A}x \rightarrow x\, \bar{C}\, x].$$

(If one wants, he can rewrite these using the $\forall.$ and $\exists.$ quantifiers.) Lewis then uses his counterpart relation to introduce a kind of modal operator which does not please me very much. My reading of his operator gives for a formula $\Phi(x)$ with one free variable:

$$\forall y\, [y\, \bar{C}\, x \rightarrow \Phi(y)]$$

for the meaning of *necessarily* $\Phi(x)$. I feel that this operator is too complicated to be taken as fundamental – but that is a discussion for another time and place.

The reason for bringing up the counterpart relation here is that it does have intuitive appeal, and it does lead to an incidence relation that may be natural. Let us define

$$\tau \approx \sigma \leftrightarrow \forall . x\, [x\, \bar{C}\, \tau \leftrightarrow x\, \bar{C}\, \sigma]$$

We then find that:

$$\forall x, y\, [\tilde{A}x \wedge x \approx y \rightarrow x\, \bar{C}\, y],$$

though the converse does not hold in general. Thus the incidence relation is too weak to be used to define counterparts, but it does have a very close connection with them. We can read $\bar{a} \approx \bar{b}$ as saying that $\bar{a}$ and $\bar{b}$ have the same counterparts in the current world, which means that they will have a large number of properties in common.

To summarize the foregoing rather unsatisfactory discussion we can make these points: (1) *at least three relations* $\approx$, $=$, $\equiv$ (which are easily confused) *do arise naturally*; (2) *semantical discussions are available that make the distinctions precise*; (3) *one needs to keep in mind that individual concepts behave differently from individuals*. Though David Kaplan has spoken at length on these distinction in several public lectures, I find in rereading his notes that I am not really convinced that we have a final solution to cross-world identifications. But I do have some advise to give (surprise!): we need to make more experiments on the construction of

models. Up to this point our indices of possible worlds have been too vague, too abstract. The idea of using $V \subseteq S^I$ was to treat an individual as a *process* (in the mathematical sense of the word.) That is, an analysis of the individual must be attempted. The method is not yet discredited, because it has not really been carried out in sufficient detail. Just to write S^I is only to indicate a skeleton but not to give it any flesh. One essential step to take in this analysis is to make the elements more specific. That must be done first before we can decide what to put in S. I do not feel that Hintikka's method of model sets would give the right solution, because I feel that one must have a concept of the *kind* of possible worlds there are *before* one can specify what is true about them. Well, that is a very philosophical question, and it would seem best discussed after one has tried to construct interesting models within the framework of this semantics. My advice is to work on this problem.

Princeton University

POSTSCRIPT (DECEMBER, 1969)

This paper was written very hastily in the latter part of May, 1968. The haste is apparent and the style intolerable; I find it now very painful reading. The point of writing the essay was to stimulate discussion, but only two of my colleagues took the time to respond (to me) in writing: Kaplan and Montague. Their detailed criticisms showed me that the paper should be completely rewritten – a task which I have unfortunately not had the time to undertake. Nevertheless the editor was kind enough to encourage publication in this very imperfect form. I would like to thank him for this and hope the obvious flaws will provoke others to do a better job.

Though I was only able to take account of a few of the many points made to me by Kaplan and Montague it will be helpful to quote some of their remarks so that my paper will not make entirely the wrong impression. First from a letter by David Kaplan (July 28, 1968):

I am repelled by your calling $\equiv$ 'identity'. This stems, I believe, from a rejection of what Carnap calls 'the method of the relation' [see Meaning & Necessity] and the attendant talk of denotation. There are just these two entities associated with each term: an extension and an intension. $=$ indicates same extension and $\equiv$ indicates same intension. But this is to ignore the primacy of extensional contexts. It is technically more convenient to treat all constants as intensional simply because we can represent the extensional ones as a subset. But one of the main aims of doing intensional logic in the way we do it is to describe intensional object languages in extensional meta-languages. If one is overly impressed by the simplifications that come from treating all constants as (at most) intensional, they may even start thinking of the intension as a kind of denotatum and talk, as you do at places, of the principles of intensional interchange as supplying a degree of 'extensionality' to the language. Look at the values of the variables to find the individuals and then keep individuals and individual concepts forever distinct. Although it is in a technical sense correct to say that extensional logic is a special case of the more general intensional logic, it is also in an important sense correct to say that extensional logic can be taken as the most general form. That is, that we can keep the sense and denotation distinction (or the name relation) between *what* we are talking about and *how* we talk about it. It was this that Frege saw and Church assumed. I suppose the situation here is very much like the question of the primacy of two valued logic and can be argued on both sides. The important thing is to be clear; I don't think it helps to call $\equiv$ 'identity' or to pretend that the displayed formulas on p. 121 are like Leibniz' law.

Concerning incidence, Kaplan had written to me earlier (July 10, 1968):

It appears that around page 132 you lost track of some entities. I think there are two problems here. (1) can $=$ be defined using $\approx$?, (2) can $\tau=\sigma$ be defined by $\Box(\tau\approx\sigma)$? The answer to (1) is yes but only if there is no merging of individuals. So in your example of people through time (where, as you remark, merging is implausible) we can use the definition:

$$\tau=\sigma\leftrightarrow\tau\approx\sigma$$

But of course if merging is allowed, as for highways through space this definition won't work, since $\overline{\text{U.S. }60}\approx\overline{\text{U.S. }70}$ is true at West Covina but $\overline{\text{U.S. }60}=\overline{\text{U.S. }70}$ is false everywhere. (Here we can even use *essential* names of the highways, $\overline{\text{U.S. }60}$ names U.S. 60 everywhere.)

The answer to (2) is always NO. And your example shows that $\Box(\tau\approx\sigma)$ is not necessary for $\tau=\sigma$. To put it another way $\tau=\sigma$ does not imply $\Box(\tau\approx\sigma)$. For suppose that at the moment The President $=$ The Crook, then the gallery representing the office of the President and the FBI gallery will share a picture: namely that of the current office holder. But if in olden days presidents were more honest, the two galleries will disagree at many points and so $\Box$(The President $\approx$ The Crook) will be false.

To show that $\Box(\tau\approx\sigma)$ is not sufficient, we must of course again assume the possibility of merging and dividing (otherwise $[\Box(\tau\approx\sigma)\rightarrow\tau\approx\sigma]$ and $[\tau\approx\sigma\leftrightarrow\tau=\sigma]$). Did you know that the soon to be built Federal highway 0 will be, for its whole length, a 73 lane road? (It will run through the center of all major cities; the idea was conceived in Los Angeles.) Being built with joint Federal State funds the Highway will coincide at every point with one of the state highways which will of course widen to 73 lanes where it joins Fed. 0 and then narrow down again when it separates. Our points of reference are the major cities, in each city it is true that: *The State highway with 73 lanes* $\approx$ $\overline{\textit{Fed. }0}$, so $\Box$ (*The State highway with 73 lanes* $\approx$ $\overline{\textit{Fed. }0}$) is also true. But it is nowhere true that *The State highway with 73 lanes* $=$ $\overline{\textit{Fed. }0}$, since all State highways veer off somewhere (and besides, don't run beyond the state borders).

To defend the principle on your page 132. From left to right we have $\tau=\sigma\rightarrow\tau\approx\sigma$ is valid so $\Box(\tau=\sigma)\rightarrow\Box(\tau\approx\sigma)$, and as you point out earlier we have $\bar{a}=\bar{b}\rightarrow\Box(\bar{a}=\bar{b})$. From right to left: $\Box(\bar{a}\approx\bar{b})$ says that for all $i\in I$, $(\|\bar{a}\|_i)_i=(\|\bar{b}\|_i)_i$. But by definition for all $i\in I$, $\|\bar{a}\|_i=a$, so we have for all $i\in I$, $a_i=b_i$. Now a and b *are* just functions with domain I, so $a=b$.

Montague makes one of these points in a different way (June 30, 1968):

You raise the question of the intuitive meaning of $\Box(\sigma\approx\tau)$. Your example, of 'the President' & 'the biggest crook', is unfortunate, because both terms denote humans, and hence continuants which (as you noticed) are identical if ever incident. But let's try $\sigma=$'the President' & $\tau=$'the heap of molecules in the President's chair'. Then $\sigma\neq\tau$ is true, because no organism is a heap of molecules; indeed, the equality fails not only at present but at all times: $\Box(\sigma\neq\tau)$ is true. But suppose that the president (and nothing else) is now occupying the president's chair. Then $\sigma\approx\tau$ is true. If we also suppose that the presidency is so engrossing that at every time i, the President at time i (and nothing else) occupies the President's chair (so that each president sits at his desk continuously from the moment he is sworn in till the moment he leaves office), then $\Box(\sigma\approx\tau)$ is true. Thus $\Box(\sigma\approx\tau)$ is compatible not only with $\sigma\neq\tau$, but even with $\Box(\sigma\neq\tau)$. (It's of course easy to construct examples in which we have $[\sigma=\tau\wedge\neg\Box(\sigma\approx\tau)]$; but as you observe we cannot have $[\sigma=\tau\wedge\Box\neg(\sigma\approx\tau)]$ or $[\Box(\sigma=\tau)\wedge\neg\Box(\sigma\approx\tau)]$.)

Further Montague objects strongly to my interpretation of predicate constants and feels that preservation of validity under substitution of formulas for predicates is an 'empty dogma'. And he has good grounds for this view. He goes on to say:

Thus I remain unregenerately convinced that the natural systems are, in order of increasing strength: (1) strict modal logic, which is the *1st* order part, containing only individual terms and the logical operator $\Box$, of the system I have sketched (this is essentially the system of Kripke-Cocchiarella-Thomason); (2) pragmatics, which adds arbitrary nonlogical propositional operators (of 1 or more places, but binding 0 var's), on the interpretation of which we agree; (3) extended pragmatics, which contains arbitrary variable-binding operators but in connection with which there is some latitude as to which kinds of operators shall be taken as basic; (4) the *2nd* order system I have sketched for you; (5) higher order systems built up on that pattern.

He then concludes that my system as amended with regards predicate constants, or where "predicate constants are (as a technical simplification) discarded and general variable-binders (more general than yours) used instead", could be regarded as a variant of his extended pragmatics. I am sorry not to be able to summarize his systems and arguments here (nor the subsequent correspondence between Kaplan and Montague) but can only refer the reader to the several recent papers where Montague explains his approach to intensional logic.